HGG Advances

ARTICLE

Functional genomic analysis reveals *HAVCR1* as the key regulator of 5q33.3 locus linked to hyperlipidemia

Wei-Ting Chen,[1,5] Chia-Ni Hsiung,[2,3] Yuh-Shan Jou,[1] Hui-Chun Wang,[4] Chen-Yang Shen,[1,6] and Wen-Cheng Chou[1,5,7,*]

Summary

Hyperlipidemia is a major risk factor for atherosclerosis and other serious cardiovascular diseases, yet it often presents without obvious symptoms in the absence of comorbidities, complicating early detection and management. Given the high heritability of lipid traits, genome-wide association studies (GWASs) have identified numerous loci associated with serum lipids; however, pinpointing the causal variants remains a challenge. Here, we investigated the 5q33.3 locus and identified *HAVCR1* as a likely effector gene influencing lipid metabolism. By integrating functional genomics and experimental validation, we identified two non-linked variants, rs6882076 and rs17573010, that modulate *HAVCR1* via allele-specific binding by IRF2 and GATA4, respectively. rs6882076 was refined from prior GWAS findings, while rs17573010 emerged from ancestry-specific analysis. These findings highlight the complexity of genetic regulation across populations and establish a mechanistic link between lipid metabolism and inflammation, offering insights into the genetic basis of hyperlipidemia and its potential translational relevance.

Introduction

Genome-wide association studies (GWASs) have dramatically advanced our understanding of the genetic underpinnings of complex traits and diseases by identifying thousands of associated loci.[1,2] However, identifying precise causal variants within these loci remains a significant challenge.[3,4] To address this, researchers have developed a range of strategies, including statistical fine-mapping and functional genomic annotation approaches, to enhance the accuracy of selecting plausible causal variants for functional validation.[3,5] These efforts are critical for translating genomic findings into therapeutic targets and advancing personalized medicine. A prime example is the discovery of proprotein convertase subtilisin/kexin type 9 (PCSK9), which plays a crucial role in lipid metabolism and cardiovascular disease.[6] Insights into PCSK9's function have led to the development of PCSK9 inhibitors, which are now widely used to lower low-density lipoprotein cholesterol (LDL-C) levels in patients with hypercholesterolemia, thereby reducing their risk of cardiovascular events.[6,7] This success highlights how understanding specific genetic variations can guide the development of targeted therapies. Nevertheless, most loci identified by GWASs remain poorly understood, largely because the non-coding nature of many associated regions complicates the identification of their target genes and biological functions.[8]

Dysregulation of plasma lipid levels, including triglycerides (TGs), total cholesterol (TC), LDL-C, and high-density lipoprotein cholesterol (HDL-C), is commonly associated with various metabolic disorders, such as cardiovascular diseases and type 2 diabetes mellitus.[9–11] Although individuals with hyperlipidemia alone may not have symptoms, understanding the genetic risk factors for developing this condition is crucial for early intervention and management of plasma lipids, reducing the likelihood of vascular complications such as atherosclerosis, myocardial infarction, and stroke.[12] Given the high heritability of lipid traits, we previously conducted a GWAS on hyperlipidemia using participants from the Taiwan Biobank (TWB).[13] This analysis identified 50 statistically significant single-nucleotide polymorphisms (SNPs) ($p < 5 \times 10^{-8}$), leading to the discovery of five independent top SNPs, each representing a distinct locus: rs780092 (2p23.3), rs1501908 (5q33.3), rs2954031 (8q24.13), rs662799 (11q23.3), and rs2075650 (19q13.32). In our previous study, we demonstrated a causal role of GATA4 in the regulation of *APOA5* levels via rs651821 at the rs662799 (11q23.3) locus.[13] Additionally, the biological relevance of the loci containing rs780092 (2p23.3), rs2954031 (8q24.13), and rs2075650 (19q13.32) has been established through their associations with nearby genes, highlighting their potential significance in hyperlipidemia pathogenesis.[14–16] Functional assessments derived from GWAS findings provide valuable insights into

[1]Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; [2]Program in Precision Medicine, National Tsing Hua University, Hsinchu, Taiwan; [3]Institute of Molecular Medicine, National Tsing Hua University, Hsinchu, Taiwan; [4]Graduate Institute of Natural Products, Kaohsiung Medical University, Kaohsiung, Taiwan
[5]Present address: Taiwan Biobank, Academia Sinica, Taipei, Taiwan
[6]Deceased
[7]Lead contact
*Correspondence: blence@gate.sinica.edu.tw
https://doi.org/10.1016/j.xhgg.2026.100630.

hyperlipidemia prevention and treatment strategies. However, the role of the rs1501908 locus, a C/G polymorphism with the highest risk (C) allele frequency among the five top SNPs, remains unclear. Notably, the 1000 Genomes Project revealed that the risk allele frequency of rs1501908 in Asian populations is 73%, which is significantly higher than that in African populations (32%), implicating the potential importance of the rs1501908 locus in ethnicity-specific hyperlipidemia susceptibility. This variant is situated between two genes, *T cell immunoglobulin and mucin domain containing 4* (*TIMD4*) and *hepatitis A virus cellular receptor 1* (*HAVCR1*), neither of which has been previously implicated in hyperlipidemia, as cell-based RNA interference screening failed to support their involvement.[17] Alternatively, *apolipoprotein O pseudogene 1* (*APOOP1*) has been proposed as the causal gene in the 5q33.3 region, influencing LDL-C levels.[18] Notably, this association, including the genetic locus influencing *APOOP1* expression, has only been observed in the haplotypes exclusive to the Amish population, with no link to rs1501908 in Taiwanese or other populations. Thus, the precise mechanisms underlying hyperlipidemia at the 5q33.3 locus require further investigation to be fully elucidated.

In this study, we provide compelling evidence supporting *HAVCR1*, rather than *TIMD4* or *APOOP1*, as the gene responsible for hyperlipidemia risk at the 5q33.3 locus, based on integrated multi-source analyses and cell-based experiments. By prioritizing functional variants within this locus, we identified two independent SNPs that regulate *HAVCR1* transcription through distinct transcription factor (TF) binding sites and regulatory mechanisms. These findings highlight the complexity of the relationship between the top genetic loci identified through association studies and their corresponding target gene. More importantly, this discovery offers potential strategies for managing hyperlipidemia in clinical settings, thereby advancing the goal of personalized healthcare.

Material and methods

Study cohort

A full description of the TWB cohort, consisting of 23,988 samples, including genotyping and biochemical data, has been provided in previous publications.[13,19] Hyperlipidemia cases were defined as individuals either self-reporting a diagnosis of hyperlipidemia or presenting with TGs $\geq$200 mg/dL or TC $\geq$240 mg/dL. Control individuals were selected based on TG levels <150 mg/dL and TC <200 mg/dL. Participants with intermediate TG or TC levels were excluded from the analysis. This study was approved by the Institutional Review Board of Academia Sinica (AS-IRB02-108010, AS-IRB03-113010), and written informed consent was obtained from each participant in accordance with the institutional requirements and the Declaration of Helsinki.

Genotyping and association analysis

The procedures for SNP selection, genotype quality control, biochemical measurements, and association analyses have been previously described.[13] We fine-mapped rs1501908 at the 5q33.3 locus by imputing a 2-Mb region centered on this SNP using SHAPEIT and IMPUTE2, with the East Asian population data from phase 3 (version 5) of the 1000 Genomes Project serving as the reference panel. Imputed SNPs were excluded if they had an INFO score <0.9, call rate <1%, minor allele frequency (MAF) <1%, or Hardy-Weinberg equilibrium $p < 1 \times 10^{-6}$. The remaining 10,614 SNPs were plotted using GraphPad Prism to visualize the significance. The rs6831550 genotype was also imputed to assess potential genetic interactions with rs6882076 in association with lipid levels.

Summary statistics for the association of SNPs with log-transformed TG and TC levels were retrieved from the meta-analysis results of the Global Lipids Genetics Consortium (GLGC) cohort, which included more than 1 million individuals from 201 primary studies across five genetic ancestry groups.[20]

SNP prioritization and eQTL analysis

The prioritization of functional variants was conducted through several steps. For gene-expression analysis, transcripts per million (TPM)-normalized RNA-sequencing data for *TIMD4*, *HAVCR1*, *APOOP1*, and *IRF* family genes were downloaded from the UCSC Xena platform and the Genotype-Tissue Expression (GTEx) portal. Statistically significant expression quantitative trait loci (*cis*-eQTLs) for genes at the 5q33.3 locus across various tissue types were also obtained directly from GTEx (Analysis V8), with *p* values provided by the platform's algorithm.[21] These GTEx *cis*-eQTL annotations were used to nominate candidate genes within the 5q33.3 interval for subsequent functional evaluation. For *IRF2*, the top *cis*-eQTL in visceral adipose tissue (omentum) was initially selected from 9,491 SNPs with a MAF >1% located within $\pm$1 Mb of the transcription start site of *IRF2* (Table S1). We excluded candidate SNPs with a MAF <10% because smaller sample sizes limit effective stratification by minor allele status. Consequently, rs6831550 emerged as the top eQTL for *IRF2* in omental tissue, with the lowest *p* value of 5.66×10^{-4}. Genotypic data for rs6882076 in the omentum were accessed via the Database of Genotypes and Phenotypes (dbGaP) under accession number phs000424.v9.p2.[21]

To visualize epigenetic regulation surrounding risk-associated SNPs, we compiled DNase I hypersensitive site (DHS) signals and histone modification marks (H3K4me1, H3K4me3, H3K27ac, and H3K9ac) from liver lineages and adipocytes. This included data from adipose nuclei, adult liver, and HepG2 cells from the Roadmap Epigenomics project as well as embryonic hepatocytes (e59 and e101) from the ENCODE project. These epigenetic signals, along with candidate SNPs reaching significance at $p < 5 \times 10^{-8}$, were presented as tracks in the UCSC Genome Browser using the human hg19/GRCh37 assembly. The corresponding UCSC sessions were Liver_rs1501908, Liver_rs6882076, and Liver_rs17573010. GeneHancer interactions were also visualized through the UCSC Genome Browser in the same sessions using the GeneHancer track last updated at UCSC on November 19, 2025 at 04:38:52. For the GeneHancer interaction involving the rs6882076-containing region and the *HAVCR1* promoter, the interaction score threshold was adjusted to include the interaction with a score of 9, which is below the current default display threshold of 10.

Cell lines and culture conditions

Human embryonic kidney cells 293T (synonyms HEK293T, RRID:CVCL_0063) obtained from Sheau-Yann Shieh (Academia

Sinica), liver cells HepG2 (RRID:CVCL_0027) purchased from the Bioresource Collection and Research Center (Hsinchu, Taiwan), Huh6 (RRID:CVCL_4381) obtained from Dr. Chia-Hung Yen (Kaohsiung Medical University), and Huh7 (RRID:CVCL_0336) and Hep3B (synonyms Hep3B2.1-7, RRID:CVCL_0326), obtained from Dr. Hui-Chun Wang (Kaohsiung Medical University), were cultured in DMEM as previously described.[13] H23 (synonyms NCI-H23, RRID:CVCL_1547) was obtained from Dr. Fu-Fei Hsu (Academia Sinica, Taiwan), and ZR75-30 (RRID:CVCL_1661) was purchased from the Bioresource Collection and Research Center. Both of these cells were cultured in RPMI 1640 medium. THLE-2 (RRID:CVCL_3803) and THLE-3 (RRID:CVCL_3804) were purchased from the American Type Culture Collection and maintained in BEGM medium (#CC-3170, Lonza), supplemented with 5 ng/mL epidermal growth factor (EGF; #AF-100-15, PeproTech) and 70 ng/mL phosphoethanolamine (#P0503, Sigma-Aldrich). Additionally, these cells were grown on plates pre-coated with BEBM medium (#CC-3171, Lonza) containing a mixture of 10 μg/mL fibronectin (#33010018, Thermo Fisher Scientific), 30 μg/mL bovine collagen type I (#A1064401, Thermo Fisher Scientific), and 10 μg/mL BSA. All cell lines were maintained in complete medium containing 10% fetal bovine serum (#26140, Thermo Fisher Scientific), confirmed to be mycoplasma free, and authenticated using short-tandem repeat profiling within the last 3 years.

Plasmid construction and transfection

Full-length human cDNAs for *IRF2*, *IRF3*, *IRF4*, *IRF8*, *GATA3*, *GATA4*, and *GATA6* were amplified from a cDNA pool derived from more than 20 human cell lines and cloned into Myc-tagged pXJ (pXJ-Myc) vectors. Similarly, *HAVCR1* and *TIMD4* were cloned into a FLAG-tagged pcDNA6 vector. The pseudogene *APOOP1* (NG_029196.3, nucleotides 31–621) was amplified from Huh7 genomic DNA and cloned into the pcDNA6 vector. *HAVCR1* haplotype 1 and haplotype 2 variants were amplified from THLE-3 and H23 genomic DNA, respectively, whereas the mucin domain truncation was generated using the site-directed mutagenesis method with PfuUltra II Fusion HS DNA polymerase (#600671, Agilent). For transient expression, cells were transfected with plasmids using the TransIT-X2 (#MIR6000, Mirus) according to the manufacturer's instructions. For CRISPR activation and CRISPR inhibition experiments, we utilized the pU6-sgRNA vector and dCas9-VPR or dCas9-KRAB plasmids, which had been previously described.[22] Duplexes of single-guide RNA (sgRNA) were annealed and cloned into the pU6-sgRNA vector using the BsmBI restriction site, followed by verification through Sanger sequencing. We transfected HepG2 cells with 1.2 μg of the dCas9-VPR (CRISPR activation) or dCas9-KRAB (CRISPR inhibition) plasmids and co-transfected 0.4 μg of pU6-sgRNA plasmids encoding either control sgRNA (sgEGFP) or sgRNAs targeting indicated SNP loci. Relative gene expression was quantified using SYBR quantitative PCR (qPCR) (#A25741, Thermo Fisher Scientific), and details of the primer pairs used for plasmid construction are listed in Table S2. For stable expression, *HAVCR1*, *TIMD4*, and *APOOP1* cDNAs were subcloned into the pLAS2w.Phyg vector, and lentiviral particles were generated via co-transfection of 293T cells with pCMVΔR8.91 and pMD.G plasmids, both obtained from the National RNAi Core Facility at Academia Sinica. The harvested lentivirus was used to infect Huh6 and Huh7 cells, following previously established protocols.[23] Two days post transduction, cells were selected with 200 μg/mL hygromycin for 2 weeks, after which stable expression was validated by immunoblotting or reverse-transcription qPCR.

For gene knockdown, small interfering RNAs (siRNAs) were transfected using Lipofectamine RNAiMAX (#13778075, Thermo Fisher Scientific). Control siRNA has been previously described,[13] while *HAVCR1* siRNA (#s25631) and *TIMD4* siRNA (#s40813) were obtained from Thermo Fisher Scientific.

RNA-sequencing and pathway analysis

Total RNA was extracted from Huh6 cells (stably expressing *APOOP1* or RFP control) and Huh7 cells (stably expressing *HAVCR1*, *TIMD4*, *APOOP1*, or RFP control) and submitted to Biotools Company (New Taipei City, Taiwan) for transcriptome sequencing on the Illumina NovaSeq 6000 platform. The resulting raw data were processed, and differentially expressed genes (DEGs) were identified using DEGseq by comparing gene-overexpressing cells to the respective RFP-expressing controls. These DEGs were subsequently subjected to pathway analysis (Biotools Cloud), including Gene Ontology (GO) biological process analysis and gene set enrichment analysis, to determine the involvement of lipid metabolism pathways and infer the potential role of the overexpressed genes in lipid metabolism. The processed RNA expression profiles are shown in dataset 2 of Table S3.

Prediction of TF motif and structural modeling

We first predicted TFs with differential binding affinities to both alleles of the target SNP using the sTRAP program. This analysis identified interferon regulatory factor (IRF) motifs, which were further examined for specificity across the nine human IRF family members using the Find Individual Motif Occurrences (FIMO) tool. The respective TF motifs were sourced from the JASPAR 2024 database.

For structural modeling, we employed the AlphaFold Server.[24] The model included a segment of double-stranded DNA (forward strand GCTCTAAAATCCCAGXTATGACCTAATTAAC, where "X" represents either the G or A allele of rs17573010) and the DNA-binding domain of GATA4 protein (amino acids 216–321, NP_001295022.1). Additionally, a zinc ion was incorporated as a co-factor to facilitate proper structural modeling of the GATA4-DNA complex.

Luciferase reporter assay

The primer pairs used for variant-containing fragment (VCF) construction are listed in Table S2. The length of the VCFs (VCF1–VCF14) ranged from 302 to 673 bp. These VCFs, featuring distinct genotypes or haplotypes, were amplified from H23 genomic DNA and cloned into the pNL3.1 vector (#N1031, Promega), except for VCF14 and VCF4, which were generated using the site-directed mutagenesis method. For the luciferase reporter assays, we transfected 20 ng of the pNL3.1 vector containing the VCFs into Huh6 (15,000 cells/well), Huh7 (10,000 cells/well), HepG2 (16,000 cells/well), or THLE-2 (6,000 cells/well) cells seeded in 48-well plates. Additionally, 20 ng of either an empty vector or vectors expressing the indicated TFs were co-transfected. pGL4.54 (#N6051, Promega), encoding firefly luciferase, was co-transfected as an internal control. Following 48 h of incubation, cells were lysed using passive lysis buffer (#E1941, Promega), and luciferase activity was measured using the Nano-Glo Dual-Luciferase Reporter Assay System (#E1610, Promega). Luminescence signals were normalized to the background activity of cells transfected with the empty pNL3.1 vector. The mean and standard deviation (SD) for each group were calculated from three independent wells

(with four or more wells for certain experiments as indicated), and two independent experiments were performed.

qPCR for gene expression and ChIP

The primer pairs used for SYBR qPCR (#A25741, Thermo Fisher Scientific) are listed in Table S2. The relative expression levels of *TIMD4*, *HAVCR1*, and *APOOP1* were quantified using reverse-transcription qPCR and normalized to the expression of *TBP* using the comparative CT method, as previously described.[13,22]

Chromatin immunoprecipitation (ChIP) was performed using the Magna ChIP G kit (#17-611, Merck-Millipore) according to the manufacturer's instructions. For endogenous ChIP, THLE-2 (1.5×10^6 cells/dish), H23 (4×10^6 cells/dish), ZR75-30 (2×10^6 cells/dish), or THLE-3 (1.2×10^6 cells/dish) cells were seeded in 100-mm dishes and fixed with 1% formaldehyde (#F8775, Sigma-Aldrich). The chromatin was then sonicated using the Bioruptor Pico (Diagenode) and immunoprecipitated with 1 μg of ChIP-grade anti-IRF2 antibody (#700226, Thermo Fisher Scientific), with normal rabbit immunoglobulin G (IgG) serving as a control. For cells overexpressing Myc-tagged IRF family proteins, the sheared DNA was immunoprecipitated with anti-Myc antibody (#M4439, Sigma-Aldrich). The immunoprecipitated DNA was eluted and quantified by SYBR qPCR, with results normalized to the input genomic DNA, as previously described.[13,22] The mean and SD values from qPCR analysis were calculated from three technical replicates unless otherwise specified, and two independent assays were performed.

Western blotting

Cell lysates were prepared in RIPA buffer (50 mM Tris-HCl [pH 7.4], 1 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, 0.1% SDS, and 10% glycerol) containing 1 mM dithiothreitol and 1× protease inhibitor cocktail (#S8830, Sigma-Aldrich). Immunoblotting for protein expression of full-length cDNA clones was performed as described earlier,[22] using anti-Myc (#M4439) or anti-FLAG (#F3165, Sigma-Aldrich) antibodies. Additional antibodies used included anti-Actin (#A2066) and anti-α-tubulin (#T6199), both from Sigma-Aldrich.

In vitro steatosis model

For the *in vitro* model of hepatic steatosis, oleic acid (OA) sodium salt (#O7501, Sigma-Aldrich) was dissolved in 99% methanol as a stock solution. Prior to treatment, HepG2 cells were washed twice with Hank's balanced salt solution (#H1387, Sigma-Aldrich) and incubated in 0.5 mM OA-containing medium. This medium was prepared by adding OA to phenol red-free DMEM (#D2902, Sigma-Aldrich), supplemented with 1% fatty-acid-free BSA (#A7030, Sigma-Aldrich) and 10% charcoal-stripped fetal bovine serum (#12676, Thermo Fisher Scientific). The approximate ratio of fatty acids bound to BSA was maintained at 3.3:1.

Intracellular TG and TC measurement

TG and TC content in steatotic HepG2 cells were measured using the Adipogenesis Detection Assay (#ab102513, Abcam) and the Cholesterol/Cholesterol Ester-Glo Assay (#J3190, Promega), respectively. In brief, HepG2 cells, either transfected with expression vectors or subjected to siRNA knockdown as indicated, were seeded into a 96-well plate at a density of 10,000 cells per well and treated with 0.5 mM OA for 24 h. After the treatment, the medium was removed, and the cells were washed twice with Hank's balanced salt solution before lysis using the buffers provided in the assay kits. The concentrations of TGs and TC were determined based on standard curves generated from known concentrations of each lipid.

Statistical analyses

All plotted data are presented as mean ± SD, unless otherwise specified, and described in the corresponding section of the method details and figure legends. Statistical differences between the two groups were compared and determined by Student's *t* test, and *p* values were calculated using GraphPad Prism 10.0.3, indicated in figures as $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$, and ns (not significant).

Results

Identification of the candidate gene linked to the hyperlipidemia-associated locus at 5q33.3

To investigate possible mechanistic evidence linking rs1501908 to hyperlipidemia, we imputed SNPs within a 2-Mb window surrounding the index SNP. This analysis identified 24 additional significant variants ($p < 5 \times 10^{-8}$) in strong linkage disequilibrium (LD) ($r^2 > 0.8$) with rs1501908 (Figure 1A and Table S4). These variants are located in non-coding regions, suggesting that they may influence gene expression through transcriptional regulation. To prioritize candidate genes within this interval, we examined *cis*-eQTLs across GTEx tissues and found that the LD region overlapped with eQTL signals linked to *TIMD4* and *HAVCR1* (Figure 1B). Because this locus had not been clearly resolved to a functionally relevant gene for hyperlipidemia or lipid metabolism, we next used functional assays to prioritize the candidate gene(s) most closely linked to lipid regulation. Initially, we hypothesized that the pseudogene *APOOP1* might be regulated by these variants, as a previous study proposed *APOOP1* as the causal gene in this region based on a functional haplotype associated with LDL-C levels.[18] However, *APOOP1* is predominantly expressed in the testis, with negligible expression in the other tissues (Figure S1A). Overexpression of *APOOP1* in liver cells indicated its role in quality control of protein synthesis (Figures S1B–S1D), suggesting that *APOOP1* is unlikely to be the causal gene in lipid metabolism. In contrast, both TIMD4 and HAVCR1, homologous proteins containing immunoglobulin and mucin domains, are expressed in hematopoietic lineage cells and modulate immune responses. Additionally, the expression of these genes was detected in hepatocyte-lineage cells (Figure S1E), warranting further investigation into their potential roles in lipid regulation. To explore this, we treated HepG2 cells with OA, a common inducer of steatosis, to induce lipid accumulation and monitor the changes in TG and TC levels.[25] Knockdown of *HAVCR1* led to significant TG and TC accumulation (Figures 1C and 1D), whereas overexpression of HAVCR1 alleviated these effects (Figures S1F and S1G). No such effects were observed for either TIMD4 or *APOOP1*. Further experiments involving deletion of the mucin domain of HAVCR1 confirmed its

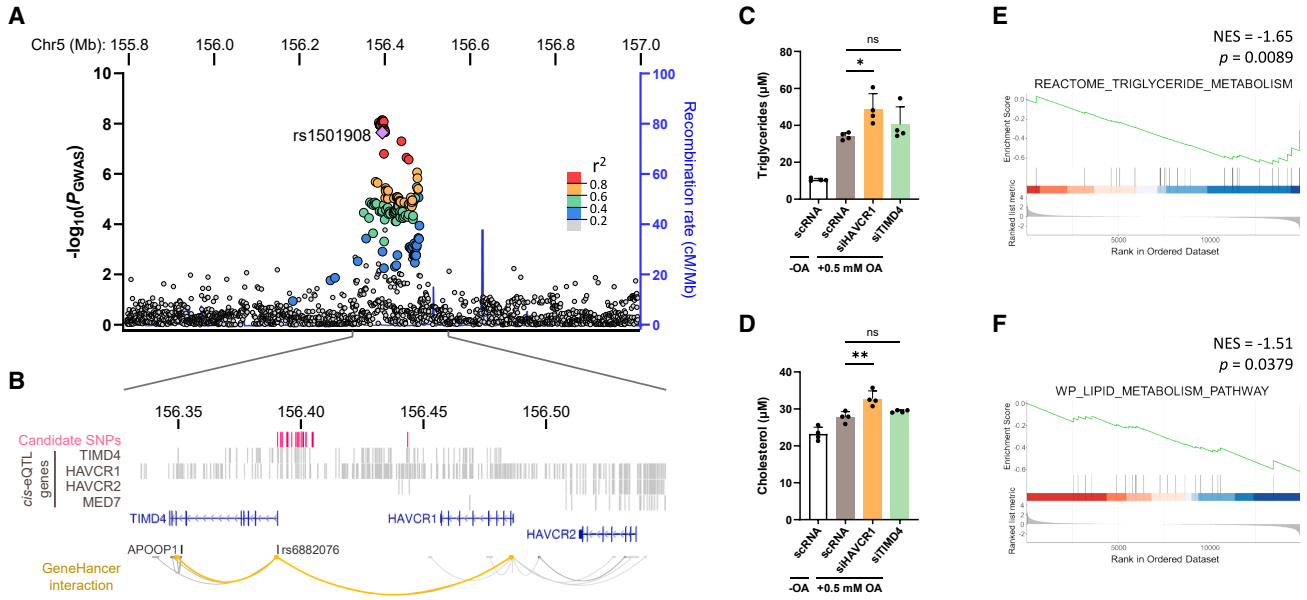


Figure 1. *HAVCR1* is the candidate gene underlying hyperlipidemia risk at 5q33.3

(A) Regional association plot for hyperlipidemia showing the SNP rs1501908 (purple diamond). Other GWAS-identified or imputed SNPs are color coded based on their linkage disequilibrium (r^2) with the index SNP.

(B) Visualization of common-tissue eQTLs within the LD cluster around the rs1501908 locus (chr5:156325001–156550000). These annotations were used to nominate candidate genes for downstream functional evaluation. Common-tissue eQTLs are defined as eQTLs associated with indicated gene expression in at least two GTEx tissue types. GeneHancer track (bottom) illustrates potential interactions between regulatory elements and the promoters of target genes. Hyperlipidemia-associated SNPs and eQTLs linked to the indicated genes can be visualized as tracks in the UCSC Genome Browser (human hg19 assembly).

(C and D) Quantification of (C) TG or (D) TC level in HepG2 cells. Cells transfected with control siRNA or siRNA targeting the indicated genes were treated with or without oleic acid (OA). The *p* and SD values are indicated, and two independent measurements of $n = 4$ are performed. Statistical differences between the two groups are compared and determined by Student's *t* test. $*p < 0.05$, $**p < 0.01$; ns, not significant.

(E and F) Gene set enrichment analysis comparing RNA-sequencing data from parental Huh7 cells and Huh7 cells overexpressing *HAVCR1*. Genes related to (E) TG and (F) lipid metabolism were enriched in cells overexpressing *HAVCR1*. NES, normalized enrichment score.

critical role in TG regulation (Figures S1H and S1I). Transcriptome sequencing of *APOOP1*, *TIMD4*, and *HAVCR1* overexpression in Huh7 cells, followed by gene set enrichment analysis, revealed that only *HAVCR1* overexpression significantly influenced TG and lipid metabolism (Figures 1E, 1F, and S1J). These findings strongly indicate that *HAVCR1* is the candidate gene responsible for modulating TG homeostasis at the 5q33.3 locus.

Prioritization of functional variants regulating *HAVCR1* expression at the 5q33.3 locus

Identifying the causal variant harboring transcriptional activity is critical to establishing the involvement of HAVCR1 in lipid metabolism. To achieve this, we initially screened the luciferase activity of DNA fragments encompassing all 25 significant variants within the LD interval (Table S4), thereby systematically prioritizing candidate functional variants for downstream analysis. We constructed 13 VCFs, which together covered all 25 significant variants, with each fragment containing at least one major allele. Of these, VCF4, VCF10, and VCF11 consistently induced luciferase activity across all three liver cell lines (Figure 2A), indicating the potential involvement of multiple causal variants. Notably, only VCF4 exhibited haplotype-specific activity (Figures S2A and S2B), with the minor alleles of rs9715911 and rs11745603, both within VCF4, significantly reducing luciferase activity in one or both liver cell lines (Figures S2C and S2D). Despite this, CRISPR activation using dCas9-VPR and proximal sgRNAs targeting the rs9715911 and rs11745603 loci failed to enhance *HAVCR1* expression (Figure 2B). In parallel, rs6882076 was included in the CRISPR activation experiment because it was supported by chromatin interaction with the *HAVCR1* promoter (Figure 1B). A different pattern was observed for the rs6882076-containing fragment (VCF1), which did not increase luciferase activity as anticipated (Figure 2A). However, subsequent epigenetic analysis revealed that rs6882076 resides within a region marked by DHS and active histone modifications (H3K4me1, H3K4me3, H3K27ac, and H3K9ac) in both liver and adipocyte cells (Figure 2C). Moreover, sgRNA targeting rs6882076 significantly upregulated *HAVCR1* expression (Figure 2B). Interestingly, *TIMD4* expression was also notably increased, as rs6882076 is located near the *TIMD4* transcription start site. Genetic imputation in the TWB cohort confirmed a significant association between

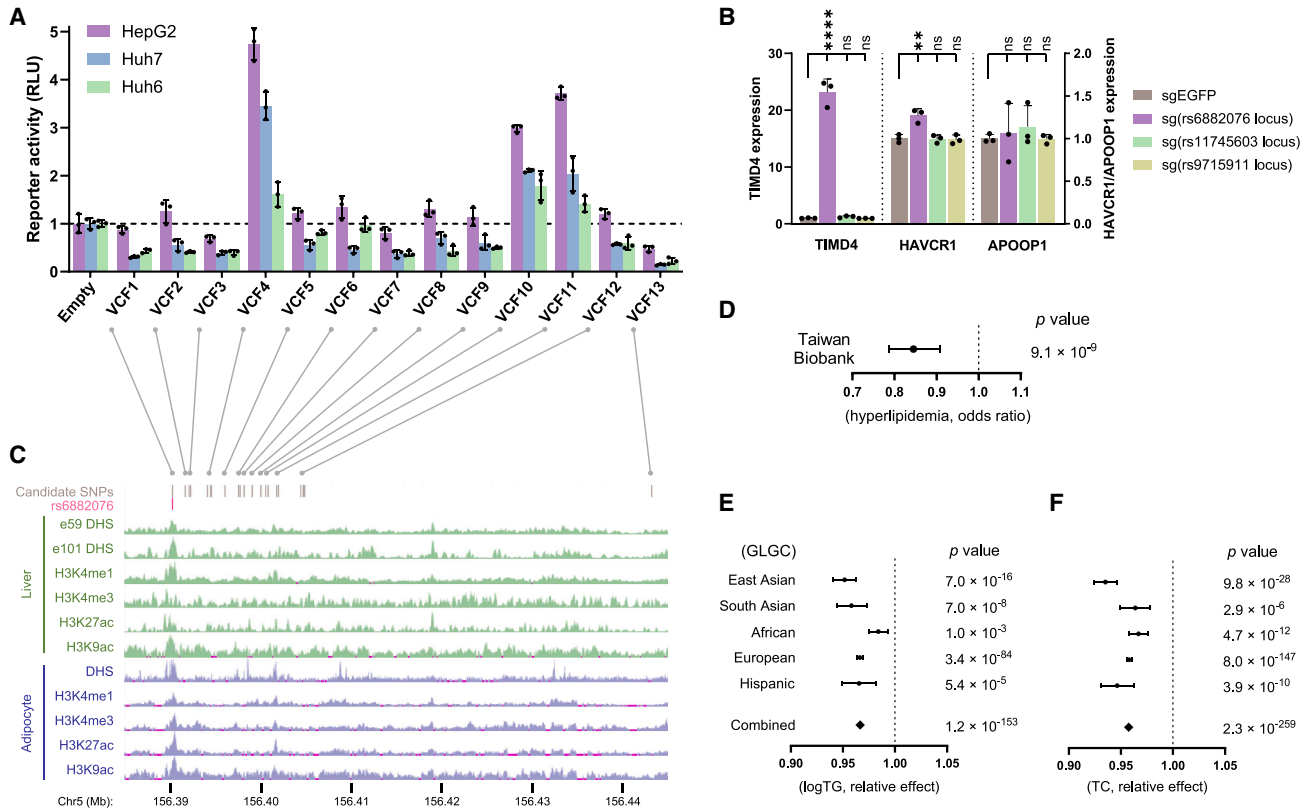


Figure 2. rs6882076 is the most likely causal variant regulating *HAVCR1* expression

(A) Bar chart showing reporter activity for the 13 variant-containing fragments (VCFs). Liver cells were transfected with empty vector or pNL3.1 containing VCFs, and luciferase activity was measured and normalized to that of cells transfected with empty pNL3.1. RLU, relative luminescence units.

(B) Bar chart showing relative gene expression of *TIMD4*, *HAVCR1*, and *APOOP1*, as measured via qPCR. HepG2 cells transfected with control sgRNA (sgEGFP) or sgRNAs targeting the rs6882076, rs9715911, and rs11745603 loci were co-transfected with dCas9-VPR for CRISPR activation analysis. The p and SD values for the results within the figure are indicated, and two independent measurements of $n = 3$ were performed. Significance between two groups is determined by Student's t test. $^{**}p < 0.01$, $^{****}p < 0.0001$; ns, not significant.

(C) Functional assessment for 25 candidate SNPs in liver (top) and adipocyte (bottom) cells. DNase I hypersensitive (DHS) regions and histone modifications (H3K4me1, H3K4me3, H3K27ac, and H3K9ac) indicate potential epigenetic regulation at this locus. Detailed cell types linked to these marks are described in the material and methods section. Each histone mark can be visualized as a track online using the UCSC Genome Browser (human hg19 assembly). Connected lines denote the approximate regions of 13 VCFs used in the reporter assays, covering all of 25 candidate SNPs.

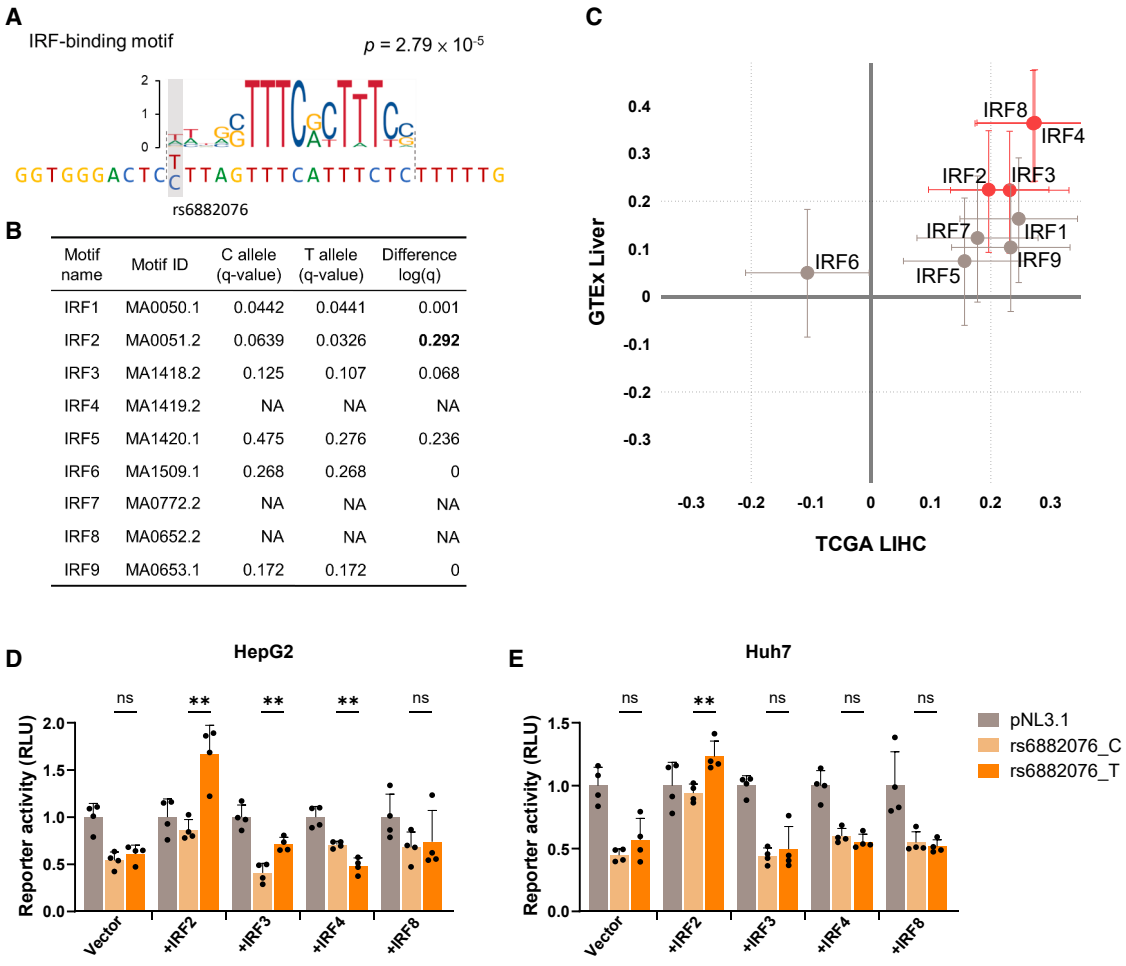
(D) Hyperlipidemia risk associated with the rs6882076 polymorphism. The odds ratio with the 95% confidence interval (CI) and p value were obtained from imputation analysis described in the material and methods section.

(E and F) Forest plot showing the association between the rs6882076 polymorphism and (E) log-transformed triglycerides (logTG) or (F) total cholesterol (TC), levels across different ancestries within the Global Lipids Genetics Consortium (GLGC) cohort. The relative effect, 95% CI, and p value are indicated.

rs6882076 and hyperlipidemia (odds ratio = 0.85, 95% confidence interval [CI] = 0.79–0.91, $p = 9.1 \times 10^{-9}$) (Figure 2D). Furthermore, rs6882076 was significantly associated with TG and TC levels across ancestries in the GLGC cohort (Figures 2E and 2F). These results prioritize rs6882076 over other LD variants as a causal variant affecting *HAVCR1* expression.

Allele-specific reporter activation at rs6882076 mediated by IRF2

Having established that both the T allele of rs6882076 (the effect allele associated with lower TG levels in GWASs) and higher *HAVCR1* expression level are linked to lower TG levels, we aimed to determine whether this allele is responsible for enhancing *HAVCR1* expression. Using reporter assays performed in liver cells, we found that the basal luciferase activity of VCF1 showed modest and context-dependent differences between the C and T alleles (Figure S3A). To explore whether this polymorphic change affects the binding of specific TFs and consequently alters *HAVCR1* expression, we performed *in silico* motif analysis. This analysis predicted that the T allele of rs6882076 creates a binding site for IRF2 (Figure 3A). Among the nine human IRF family members that share similar DNA-binding motifs, IRF2 displayed the strongest potential for allele-specific binding (Figure 3B). To confirm the relevance of IRF2, we analyzed gene-expression profiles in liver tissues from the GTEx and The Cancer Genome Atlas (TCGA). We found a significant positive correlation between *HAVCR1* expression and four *IRF*



Motif name	Motif ID	C allele (q-value)	T allele (q-value)	Difference log(q)
IRF1	MA0050.1	0.0442	0.0441	0.001
IRF2	MA0051.2	0.0639	0.0326	**0.292**
IRF3	MA1418.2	0.125	0.107	0.068
IRF4	MA1419.2	NA	NA	NA
IRF5	MA1420.1	0.475	0.276	0.236
IRF6	MA1509.1	0.268	0.268	0
IRF7	MA0772.2	NA	NA	NA
IRF8	MA0652.2	NA	NA	NA
IRF9	MA0653.1	0.172	0.172	0

Figure 3. IRF2 enables allele-specific reporter activation at rs6882076

(A) Predicted DNA-binding motif of IRF family over the sequence surrounding the C and T alleles of rs6882076. Sequence logos were retrieved from the JASPAR 2024 database (IRF2:MA0051.2) and analyzed using the FIMO program to determine *p* value.

(B) FIMO program results showing the match *q* value (adjusted *p* value) and differences between IRF DNA-binding motifs and the C or T alleles of rs6882076.

(C) Scatterplot showing the concordance of the correlation between the expression of *HAVCR1* and nine *IRF* genes in TCGA liver cancer (*x* axis) or GTEx liver tissues (*y* axis). The 95% CIs for the correlation coefficients *r*, determined by the Spearman test, are shown.

(D and E) Bar charts showing reporter assay for a DNA fragment containing the C or T allele of rs6882076 in (D) HepG2 and (E) Huh7 cells transfected with control vector or plasmids overexpressing various IRF proteins. RLU, relative luminescence units. The *p* and SD values are indicated, and two independent transfections of $n = 4$ were performed. Statistical differences between the two groups are compared and determined by Student's *t* test. $^{**}p < 0.01$.

family members: *IRF2*, *IRF3*, *IRF4*, and *IRF8* (Figure 3C). To assess their functional involvement, we overexpressed these TFs in liver cells (Figure S3B). Among them, only IRF2 consistently and significantly increased the luciferase activity of VCF1 harboring the T allele of rs6882076 in both HepG2 and Huh7 cells compared to the control plasmid (Figures 3D and 3E). These findings indicated that IRF2 can modulate *HAVCR1* expression in an rs6882076 T-allele-preferential manner.

Genotype-dependent enhancement of *HAVCR1* expression by IRF2

Our analysis of the available liver cancer cell lines revealed that all of these cells possessed the rs6882076-CC genotype (Figure S4A). These cells exhibited detectable IRF2 enrichment at the rs6882076 site (Figure S4B), but this was not accompanied by clear *HAVCR1* induction (Figure S4C). In contrast, the THLE-2 cell line, an immortalized human liver epithelial cell line, is heterozygous for the rs6882076 locus (rs6882076-CT) (Figure S4A). We performed ChIP assays followed by qPCR in THLE-2 cells to confirm IRF2's preferential binding to the T allele of rs6882076. The results showed a significant enrichment of IRF2 at the rs6882076 locus (Figure 4A), accompanied by an increase in *HAVCR1* expression upon IRF2 overexpression (Figure 4B). To extend these findings to other genetic backgrounds, we validated IRF2 binding at the rs6882076 site in additional cell lines with the rs6882076-CT (H23 and ZR75-30) and rs6882076-TT (THLE-3) genotypes (Figure 4C). Consistently, IRF2 overexpression in H23 and THLE-3 cells

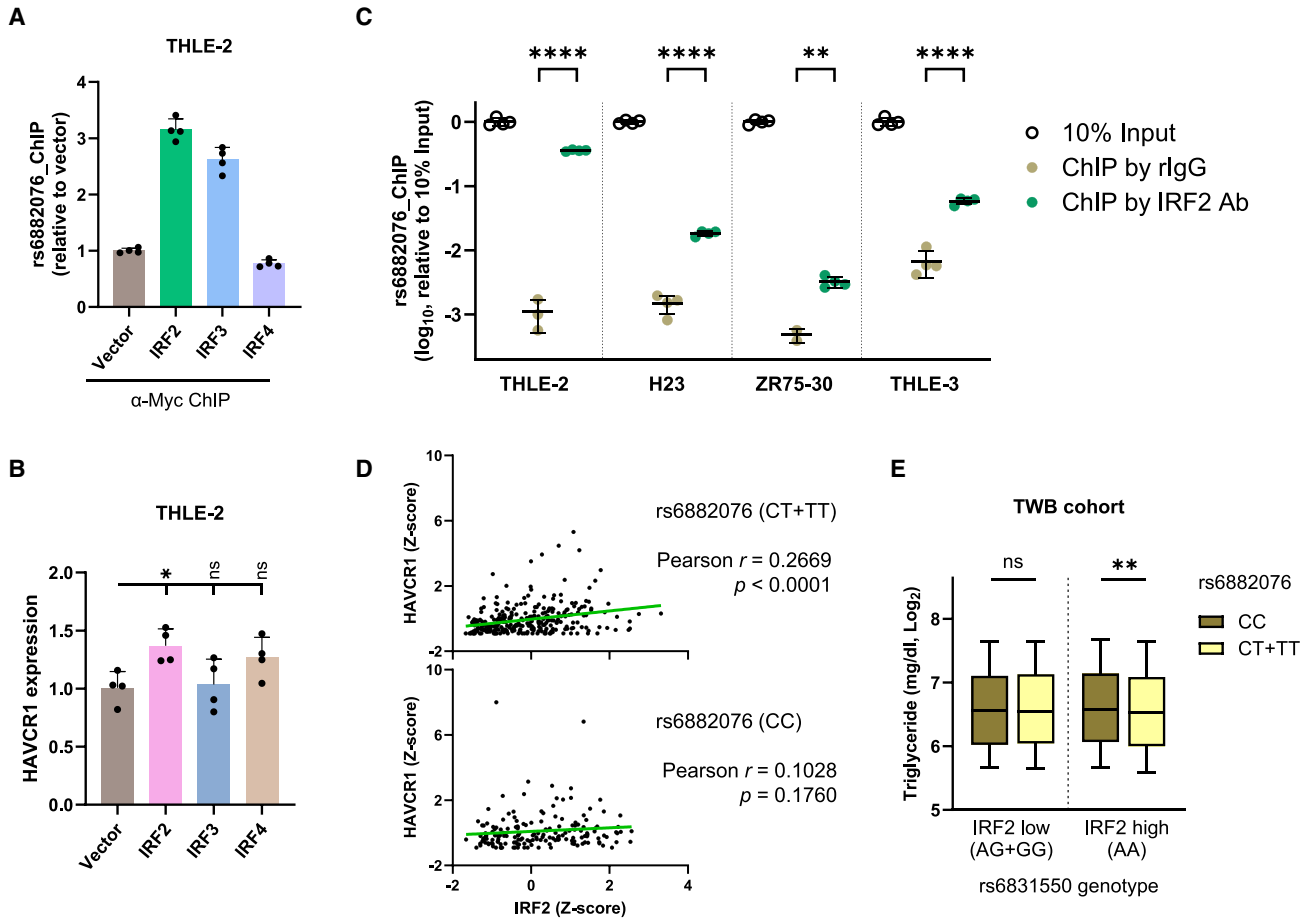


Figure 4. IRF2 enhances *HAVCR1* expression depending on rs6882076 genotype

(A) Bar chart displaying relative enrichment of the rs6882076 site following ChIP with anti-Myc antibody, as quantified by qPCR. DNA fragments were extracted and sheared from THLE-2 cells transfected with either empty vector or pXJ-Myc expressing IRF proteins.

(B) Bar chart showing relative *HAVCR1* expression in THLE-2 cells transfected with empty vector or pXJ-Myc expressing IRF proteins, with RNA expression quantified by qPCR.

(C) Bar chart displaying relative enrichment of the rs6882076 site after ChIP with normal rabbit IgG (rIgG) or anti-IRF2 antibody, using sheared DNA from THLE-2, H23, ZR75-30, and THLE-3 cells, quantified by SYBR-based qPCR. The *p* and SD values for the results within the figure are indicated, and two independent transfections of $n = 4$ were performed. In (B) and (C), statistical differences between the two groups are compared and determined by Student's *t* test. $*p < 0.05$, $**p < 0.01$, $****p < 0.0001$; ns, not significant.

(D) Scatterplots showing the relationship between *IRF2* and *HAVCR1* expression in GTEx visceral adipose tissue (omentum), stratified by rs6882076 CT + TT (upper) and CC (lower) genotype. Pearson correlation *r* value and *p* value are indicated.

(E) Log-transformed TG levels retrieved from the TWB cohort, sequentially stratified by rs6831550 and rs6882076 genotypes. Median values and the 10th–90th percentile range are indicated. Statistical differences between the two groups are compared and determined by Student's *t* test. $**p < 0.01$; ns, not significant.

increased *HAVCR1* expression (Figure S4D), although ZR75-30 cells failed to express *HAVCR1*. To further explore this in a broader biological context, we analyzed population-level data from the GTEx dataset. Given the relatively small sample size of liver tissues, we focused on omental adipose tissues. Among individuals with the rs6882076-CT/TT genotype, where at least one allele harbors the IRF2 binding motif, *HAVCR1* expression was positively correlated with *IRF2* expression (Figure 4D). In contrast, this correlation was not significant in individuals with the rs6882076-CC genotype, which lacks the IRF2 binding motif (Figure 4D). Given the genotype-dependent relationship between *IRF2* and *HAVCR1* expression, we next examined whether this regulatory context was also reflected in lipid traits. We identified rs6831550 as the most significant *cis*-eQTL for *IRF2* expression among the variants retained for analysis in omental adipose tissue in the GTEx dataset (Figure S4E). As expected, rs6831550 was not directly associated with lipid traits such as TGs and TC (Figure S4F), consistent with its use here as a *cis*-eQTL proxy for *IRF2* expression rather than as a lipid-associated variant. Among the TWB participants, those with the rs6831550-GG/AG genotypes, which corresponded to lower *IRF2* expression levels, showed no significant changes in lipid levels between the rs6882076 genotypes (Figures 4E and S4G), whereas individuals carrying the rs6831550-AA genotype (associated with elevated *IRF2* expression) and rs6882076-CT/TT genotypes exhibited notable reductions in plasma

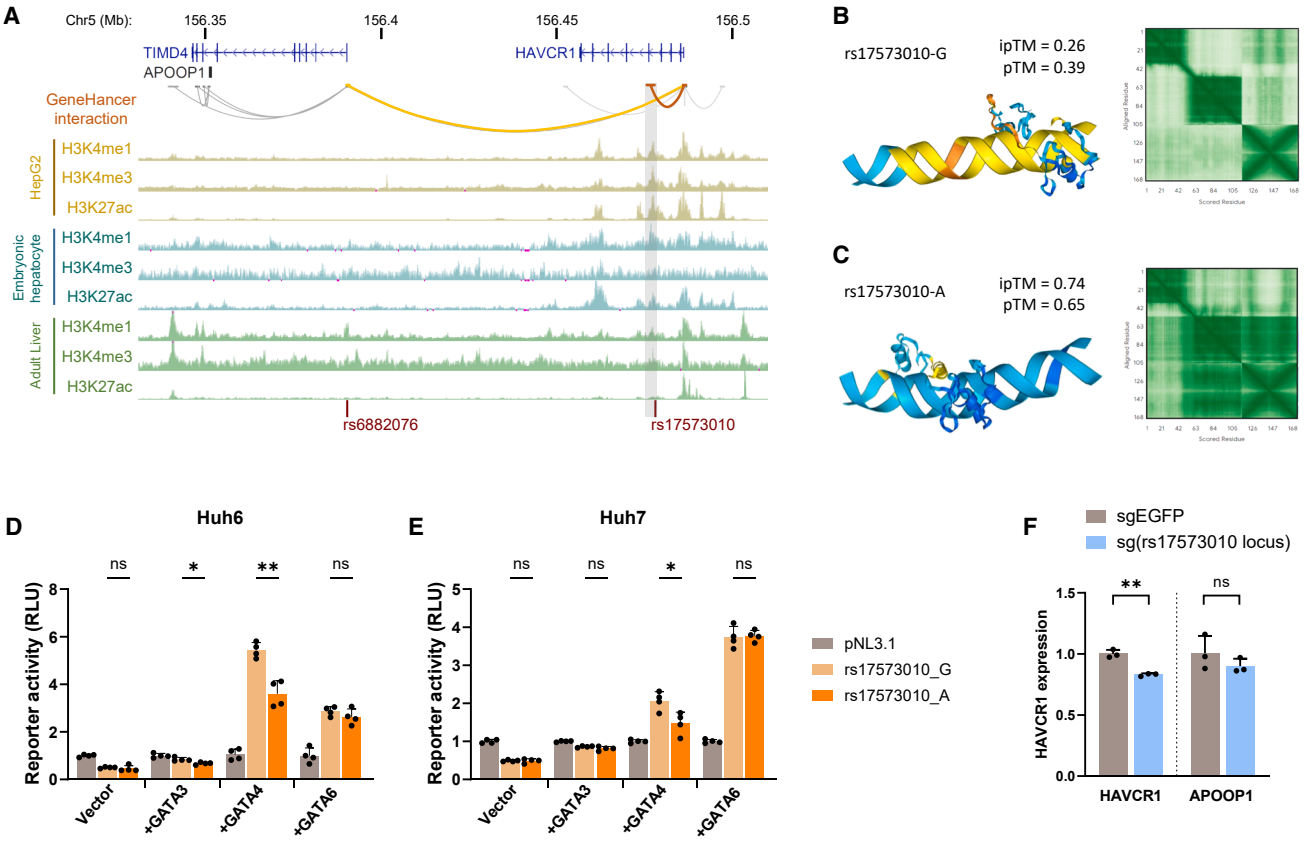


Figure 5. The ancestry-specific variant rs17573010 downregulates *HAVCR1* expression

(A) Visualization of epigenetic regulation at the rs17573010 locus. Genes neighboring rs17573010 are shown, and GeneHancer track illustrates probable interactions between regulatory elements and promoter of target genes. Histone modifications (H3K4me1, H3K4me3, and H3K27ac) suggest putative epigenetic regulation in HepG2, embryonic, and adult liver tissues. Detailed information on the cell types associated with these marks is provided in the material and methods section. The marks can also be visualized online via the UCSC Genome Browser (human hg19 assembly).

(B and C) AlphaFold molecular modeling (left) and expected aligned error (right) predicting the binding of the DNA-binding domain of GATA4 to DNA oligonucleotides containing either the (B) G allele or (C) A allele of rs17573010. The predicted template modeling (pTM) score and the interface predicted template modeling (ipTM) score are indicated.

(D and E) Bar charts showing reporter analysis of DNA fragment containing either the G or A allele of rs17573010 in (D) Huh6 or (E) Huh7 cells transfected with control vector or plasmids overexpressing various GATA proteins. Relative luminescence units (RLU) were measured. The p and SD values for the results are indicated, and two independent transfections of $n = 4$ were performed.

(F) Bar chart showing relative gene expression of *HAVCR1* and *APOOP1*, as measured by qPCR analysis. HepG2 cells were transfected with control sgRNA (sgEGFP) or sgRNAs targeting the rs17573010 locus, along with co-transfection of vectors expressing dCas9-KRAB for CRISPR inhibition analysis. The p and SD values for the results are indicated, and two independent measurements of $n = 3$ are performed.

In (D)–(F), statistical differences between the two groups within the figure are determined by Student's t test. *$p < 0.05$, **$p < 0.01$; ns, not significant.

TG and TC levels (Figures 4E and S4G). It does not seem to modulate all traits universally but is specific to certain lipid levels (Figures S4H and S4I). These findings suggest that the T allele of rs6882076 provides a more favorable context for IRF2-dependent activation of *HAVCR1* and that the rs6882076 locus exerts its regulatory effects in an allele-dependent manner.

HAVCR1 downregulation by an ancestry-specific locus previously associated with elevated lipid levels

A previous association study in the Amish population identified rs17573010 at the 5q33.3 locus, where the minor allele A is associated with a 13.1 mg/dL increase in LDL-C levels.[18] Following our finding that *HAVCR1* is likely the true causal gene at this locus, we extended our investigation to explore whether the risk allele of rs17573010 or its linked variants is involved in *HAVCR1* downregulation. Among the 14 most likely causal variants highlighted in the original study,[18] rs17573010 emerged as a strong candidate as a functional variant within a regulatory region characterized by the GeneHancer regulatory element linked to the *HAVCR1* promoter and active histone marks (H3K4me1, H3K4me3, and H3K27ac) in liver cells (Figure 5A). rs17573010 is a G-to-A polymorphism, and *in silico* analysis suggests that the A allele may create a binding site for GATA TFs (Figure S5A). Of the six GATA family genes, we previously demonstrated

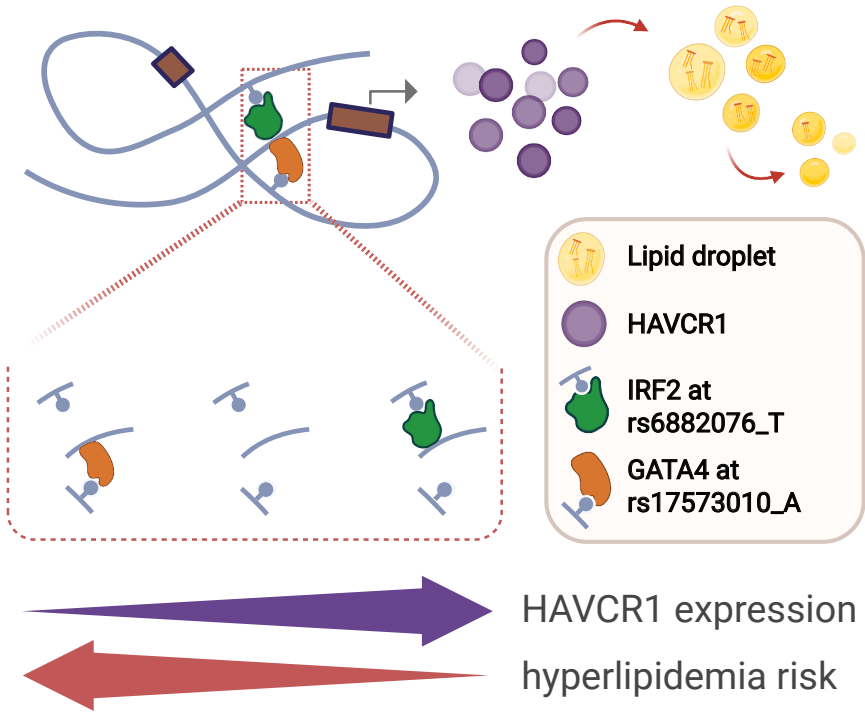


Figure 6. Proposed model of allele-dependent regulation of *HAVCR1* at the 5q33.3 locus

Two independent variants at 5q33.3 are proposed to modulate *HAVCR1* expression through distinct allele-dependent TF-binding events. The A allele of rs17573010 favors GATA4 binding and is associated with reduced HAVCR1 regulatory activity, whereas the T allele of rs6882076 favors IRF2 binding and is associated with increased HAVCR1 expression. The inset is arranged schematically to reflect the proposed direction from lower to higher HAVCR1 regulation, while the opposing arrows summarize the inverse relationship between HAVCR1 expression and hyperlipidemia risk.

that GATA4 is highly expressed in the liver,[13] while GATA3 and GATA6 show lower levels of expression, and the remaining GATA proteins are nearly undetectable. Structural modeling by AlphaFold predicted that the A allele of rs17573010 showed a higher affinity for GATA4 binding than the G allele (Figures 5B and 5C),[24] supporting the notion of allele-specific binding at this locus. To experimentally validate this, we performed reporter assays using constructs containing the rs17573010 region (VCF14) with either the G or A alleles in liver cell lines. The results demonstrated that, although both allelic fragments retained reporter activity above the vector control, the A allele showed significantly lower luciferase activity than the G allele in the presence of GATA4 expression, indicating a repressive effect on *HAVCR1* transcription (Figures 5D, 5E, and S5B). This finding aligns with the hypothesis that the A allele of rs17573010 is associated with *HAVCR1* downregulation. Since rs17573010 appears to be an ancestry-specific variant,[18] we lacked cell lines carrying the A allele for further study. To circumvent this, we transfected HepG2 cells with dCas9-KRAB (CRISPR inhibition) and targeted the rs17573010 locus using proximal sgRNAs. These experiments showed a significant repression of *HAVCR1* expression but not *APOOP1* expression (Figure 5F). Collectively, these findings provide a plausible explanation for the genetic association observed in the Amish study, in which rs17573010 may exert its effects on lipid levels through the regulation of *HAVCR1* expression, similar to the rs6882076 variant. Altogether, our data support a model in which diverse genetic variants at the 5q33.3 locus, acting through distinct TFs, regulate *HAVCR1* expression and impact lipid metabolism across populations (Figure 6). This highlights the complexity of the genetic regulation underlying lipid levels and the value of integrating multi-ancestry data to understand these processes.

Discussion

In this study, we explored a potential candidate locus for hyperlipidemia and TG and TC regulation at 5q33.3 and identified *HAVCR1* as the candidate gene linked to this locus, a finding that is different from previous assumptions about the role of *APOOP1*.[18] Our results provide compelling evidence that *HAVCR1*, rather than *TIMD4* or *APOOP1*, is responsible for hyperlipidemia risk at this locus. The association between the rs6882076 and rs17573010 variants and *HAVCR1* expression underscores the importance of this gene in lipid metabolism, an observation that has not been previously reported in GWASs.

Regarding the plausibility of IRF2 and GATA4 in modulating *HAVCR1* expression, it is not surprising that our findings align with established roles of GATA4 in the liver,[26] where it can regulate *APOA5* expression via the rs651821 locus.[13] Likewise, IRF2, which is basally expressed in the liver,[27,28] can be further induced by interferon signaling and plays a critical role in modulating T cell exhaustion during chronic inflammation.[29] IRF2 induction helps to restrain excessive interferon signaling, thereby preventing immune overactivation.[28,30] We propose that the interaction between IRF2 and HAVCR1 may be part of a broader biological response to excessive lipid levels, which induce inflammatory responses and contribute to the progression of atherosclerosis,[31] attracting more immune cells and

leading to plaque instability.[32] The upregulation of *HAVCR1* by IRF2 may function as a feedback mechanism to mitigate inflammation.[33] Our findings provide genetic insights that may serve as a foundation for future studies investigating signal transduction of lipid disorders in inflammatory contexts.

Previous studies in the Amish population have identified a founder haplotype, including the minor allele of rs17573010, which may influence the expression of the pseudogene *APOOP1*.[18] It has been hypothesized that increased *APOOP1* expression could suppress its homologous *APOO*, which is localized to the inner mitochondrial membrane and plays a role in lipid regulation.[34,35] However, this hypothesis has yielded inconsistent results across different studies.[36,37] Although the exact mechanism by which APOO regulates LDL-C, but not TGs or TC, remains unclear, our experiments showed that *APOO* expression was not significantly affected by *APOOP1* overexpression in Huh6 and Huh7 cells. This discrepancy, along with the effects of rs17573010, further supports our conclusion that HAVCR1, rather than *APOOP1*, is more likely to contribute to the regulation of TGs and TC at the 5q33.3 locus.

Our study extends beyond the traditional approach of focusing on a single variant as the sole regulator of gene expression. Although post-GWAS statistical analyses have suggested that a locus may harbor multiple potential regulatory variants,[38,39] this hypothesis has typically been confirmed only through allele-specific transcriptional regulation,[40,41] often using reporter assays. Regarding this, previous examples have demonstrated that a pair of LD SNPs in enhancer regions function together to drive prostate cancer[42] or that distinct variants, a regulatory enhancer and a protein-coding variant, confer risk for type 2 diabetes through separate mechanisms.[43] In this study, distinct top variants that are not in LD can independently regulate the same gene. These studies presented rare examples of multiple causal variants cooperating in the regulation of a single gene, shedding light on the complex regulatory architecture underlying the genetic effects on disease traits. On the other hand, our study shows that variants exhibiting allele-specific activity do not necessarily indicate functional relevance, aligned with the more in-depth analyses in a previous study.[44] This underscores the need for further validation of variants identified through a single genome-wide functional screen.

Our interpretation of GTEx eQTL signals at the 5q33.3 locus was used primarily to define candidate genes for downstream functional evaluation. Formal co-localization analysis and cell-type-specific eQTL refinement may further improve statistical resolution of this locus but were beyond the scope of the present functional study. It is noteworthy that our GWAS findings relied on imputed SNPs with MAF <1% to investigate the underlying biological mechanisms. While this approach provides valuable insights, it may overlook the role of rare variants. For instance, we identified rs17573010 as a potential causal rare variant, specific to the Amish population. Furthermore, our conclusions regarding the involvement of HAVCR1 in TG and TC levels were based on assays in HepG2 cells, which might not fully capture the complexity of *in vivo* lipid regulation. Nonetheless, our analyses successfully linked two distinct causal variants, rs6882076 and rs17573010, to *HAVCR1* regulation. Lack of evidence supports the involvement of other proteins in hyperlipidemia risk at this locus. Our data support a role for IRF2 in driving *HAVCR1* expression. Although the role of HAVCR1 in lipid metabolism has not yet been well explored, as it is beyond the scope of this study, HAVCR1 deletion has been shown to exacerbate hepatic steatosis.[45] A previous study indicated that the mucin domain of HAVCR1 facilitates fatty acid import into cells,[46] further supporting the idea that *HAVCR1* is the causal gene at the hyperlipidemia-associated locus on 5q33.3.

Our findings provide compelling evidence that *HAVCR1* is the causal gene at the 5q33.3 locus, with both rs6882076 and rs17573010 contributing to its regulation through distinct, non-LD mechanisms. In this context, the regulatory role of IRF2 links immune modulation to lipid metabolism, offering a new perspective on the interplay between inflammation and lipid homeostasis. These insights pave the way for future studies to further elucidate the molecular pathways involved in hyperlipidemia and to explore the translational relevance of HAVCR1-mediated regulation in lipid-related disorders.

Data and code availability

Individual-level genotyping data from the TWB were used under license for the current study and cannot be redistributed. Access to TWB data can be requested directly from the Taiwan Biobank (biobank@gate.sinica.edu.tw) with permission from the Ministry of Health and Welfare, Taiwan. GWAS summary statistics for hyperlipidemia risk are available from the GWAS Catalog under accession number GCST90090994.[13] Summary statistics for the association of SNPs with log-transformed TG and TC levels were retrieved from the meta-analysis results of the GLGC cohort.[20]

GTEx-based expression and eQTL analyses were performed using publicly available GTEx Portal queries. Genotypic data for rs6882076 in visceral adipose tissue (omentum) were accessed through dbGaP under accession number phs000424.v9.p2.[21] Epigenetic and regulatory annotations were examined using the UCSC Genome Browser, and the corresponding UCSC sessions provided in the paper are Liver_rs1501908, Liver_rs6882076, and Liver_rs17573010.

RNA-sequencing experiments generated in this study have been deposited in the Gene Expression Omnibus under accession number GEO: GSE280379. The deposited materials include raw sequencing data and processed expression results. RNA-sequencing preprocessing and quantification were performed by Biotools Company using the following workflow: Trimmomatic v.0.38 (parameters: LEADING:3; TRAILING:3; SLIDINGWINDOW:4:20; MINLEN:36), HISAT2 v.2.1.0 (default), featureCounts v.2.0.0 (default), and the GRCh38 reference assembly. Differential expression and pathway analyses were also performed by Biotools as described in material and methods. No custom or original code was generated in this study. The full-length cDNA expression

clones described in this study have been deposited at Addgene under plasmid IDs 232642, 232643, 232667, 232668, 232646, 232647, 232648, 232669, and 232670.

Acknowledgments

This study was supported by grants from the National Science and Technology Council, Taiwan (NSTC 112-2314-B-001-002 to C.-Y.S., NSTC 113-2314-B-001-014 and NSTC 114-2314-B-001-004 to Y.-S.J.). W.-C.C. was supported by a research fellowship from NSTC, Taiwan (NSTC 113-2811-B-001-094). We thank the investigators of the Roadmap Epigenomics, ENCODE, TCGA, and GTEx projects, who have made their invaluable data available to the scientific community. We thank the National RNAi Core Facility in Academia Sinica, Taiwan, for providing pLAS2w. Phyg, pCMVΔR8.91, and pMD.G plasmids. We also thank the Data Science Statistical Cooperation Center of Academia Sinica (AS-CFII-111-215) for statistical support and the DNA Sequencing Core Facility at Academia Sinica (AS-CFII-113-A12) for sequencing service.

In memory of the co-author, Dr. Chen-Yang Shen, who was the supervising author for this study.

Author contributions

W.-C.C. and C.-Y.S. conceived the study; W.-T.C. and W.-C.C. designed and performed cell-based experiments; C.-N.H. and W.-C.C. carried out the statistical and public database analyses; Y.-S.J. and H.-C.W. provided scientific input; and W.-C.C. wrote the manuscript with help and input from W.-T.C. and Y.-S.J. All authors read and approved the manuscript.

Declaration of interests

The authors declare no competing interests.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2026.100630.

Web resources

AlphaFold Server, https://alphafoldserver.com/
Biotools Cloud, https://cloud.toolsbiotech.com/
FIMO, https://meme-suite.org/meme/tools/fimo
Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo/
GLGC cohort, https://csg.sph.umich.edu/willer/public/glgc-lipids2021/
GTEx, www.gtexportal.org
GWAS Catalog, https://www.ebi.ac.uk/gwas/
JASPAR, https://jaspar.elixir.no/
sTRAP program, http://trap.molgen.mpg.de/cgi-bin/trap_two_seq_form.cgi
UCSC Genome Browser, https://genome.ucsc.edu/cgi-bin/hgPublicSessions
UCSC Xena, https://xena.ucsc.edu/

Received: January 15, 2026
Accepted: May 21, 2026

References

1. Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., and Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. Am. J. Hum. Genet. *101*, 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005.
2. Lappalainen, T., Li, Y.I., Ramachandran, S., and Gusev, A. (2024). Genetic and molecular architecture of complex traits. Cell *187*, 1059–1075. https://doi.org/10.1016/j.cell.2024.01.023.
3. Fabo, T., and Khavari, P. (2023). Functional characterization of human genomic variation linked to polygenic diseases. Trends Genet. *39*, 462–490. https://doi.org/10.1016/j.tig.2023.02.014.
4. Brandes, N., Weissbrod, O., and Linial, M. (2022). Open problems in human trait genetics. Genome Biol. *23*, 131. https://doi.org/10.1186/s13059-022-02697-9.
5. IGVF Consortium (2024). Deciphering the impact of genomic variation on function. Nature *633*, 47–57. https://doi.org/10.1038/s41586-024-07510-0.
6. Roberts, R. (2015). A genetic basis for coronary artery disease. Trends Cardiovasc. Med. *25*, 171–178. https://doi.org/10.1016/j.tcm.2014.10.008.
7. Sabatine, M.S. (2019). PCSK9 inhibitors: clinical evidence and implementation. Nat. Rev. Cardiol. *16*, 155–165. https://doi.org/10.1038/s41569-018-0107-8.
8. Pang, B., van Weerd, J.H., Hamoen, F.L., and Snyder, M.P. (2023). Identification of non-coding silencer elements and their regulation of gene expression. Nat. Rev. Mol. Cell Biol. *24*, 383–395. https://doi.org/10.1038/s41580-022-00549-9.
9. Cadby, G., Giles, C., Melton, P.E., Huynh, K., Mellett, N.A., Duong, T., Nguyen, A., Cinel, M., Smith, A., Olshansky, G., et al. (2022). Comprehensive genetic analysis of the human lipidome identifies loci associated with lipid homeostasis with links to coronary artery disease. Nat. Commun. *13*, 3124. https://doi.org/10.1038/s41467-022-30875-7.
10. Hsiung, C.N., Chang, Y.C., Lin, C.W., Chang, C.W., Chou, W.C., Chu, H.W., Su, M.W., Wu, P.E., and Shen, C.Y. (2020). The Causal Relationship of Circulating Triglyceride and Glycated Hemoglobin: A Mendelian Randomization Study. J. Clin. Endocrinol. Metab. *105*, 908–919. https://doi.org/10.1210/clinem/dgz243.
11. Semple, R.K., Sleigh, A., Murgatroyd, P.R., Adams, C.A., Bluck, L., Jackson, S., Vottero, A., Kanabar, D., Charlton-Menys, V., Durrington, P., et al. (2009). Postreceptor insulin resistance contributes to human dyslipidemia and hepatic steatosis. J. Clin. Investig. *119*, 315–322. https://doi.org/10.1172/JCI37432.
12. Borén, J., Taskinen, M.R., Björnson, E., and Packard, C.J. (2022). Metabolism of triglyceride-rich lipoproteins in health and dyslipidaemia. Nat. Rev. Cardiol. *19*, 577–592. https://doi.org/10.1038/s41569-022-00676-y.
13. Chou, W.C., Chen, W.T., and Shen, C.Y. (2022). A common variant in 11q23.3 associated with hyperlipidemia is mediated by the binding and regulation of GATA4. npj Genom. Med. *7*, 4. https://doi.org/10.1038/s41525-021-00279-5.
14. Douvris, A., Soubeyrand, S., Naing, T., Martinuk, A., Nikpay, M., Williams, A., Buick, J., Yauk, C., and McPherson, R. (2014). Functional analysis of the TRIB1 associated locus linked to plasma triglycerides and coronary artery disease. J. Am. Heart Assoc. *3*, e000884. https://doi.org/10.1161/JAHA.114.000884.

15. López Rodríguez, M., Kaminska, D., Lappalainen, K., Pihlajamäki, J., Kaikkonen, M.U., and Laakso, M. (2017). Identification and characterization of a FOXA2-regulated transcriptional enhancer at a type 2 diabetes intronic locus that controls GCKR expression in liver cells. Genome Med. *9*, 63. https://doi.org/10.1186/s13073-017-0453-x.
16. Zimoń, M., Huang, Y., Trasta, A., Halavatyi, A., Liu, J.Z., Chen, C.Y., Blattmann, P., Klaus, B., Whelan, C.D., Sexton, D., et al. (2021). Pairwise effects between lipid GWAS genes modulate lipid plasma levels and cellular uptake. Nat. Commun. *12*, 6411. https://doi.org/10.1038/s41467-021-26761-3.
17. Blattmann, P., Schuberth, C., Pepperkok, R., and Runz, H. (2013). RNAi-based functional profiling of loci from blood lipid genome-wide association studies identifies genes with cholesterol-regulatory function. PLoS Genet. *9*, e1003338. https://doi.org/10.1371/journal.pgen.1003338.
18. Montasser, M.E., O'Hare, E.A., Wang, X., Howard, A.D., McFarland, R., Perry, J.A., Ryan, K.A., Rice, K., Jaquish, C.E., Shuldiner, A.R., et al. (2018). An APOO Pseudogene on Chromosome 5q Is Associated With Low-Density Lipoprotein Cholesterol Levels. Circulation *138*, 1343–1355. https://doi.org/10.1161/CIRCULATIONAHA.118.034016.
19. Feng, Y.C.A., Chen, C.Y., Chen, T.T., Kuo, P.H., Hsu, Y.H., Yang, H.I., Chen, W.J., Su, M.W., Chu, H.W., Shen, C.Y., et al. (2022). Taiwan Biobank: A rich biomedical research database of the Taiwanese population. Cell Genom. *2*, 100197. https://doi.org/10.1016/j.xgen.2022.100197.
20. Graham, S.E., Clarke, S.L., Wu, K.H.H., Kanoni, S., Zajac, G.J.M., Ramdas, S., Surakka, I., Ntalla, I., Vedantam, S., Winkler, T.W., et al. (2021). The power of genetic diversity in genome-wide association studies of lipids. Nature *600*, 675–679. https://doi.org/10.1038/s41586-021-04064-3.
21. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776.
22. Chou, W.C., Chen, W.T., Kuo, C.T., Chang, Y.M., Lu, Y.S., Li, C.W., Hung, M.C., and Shen, C.Y. (2024). Genetic insights into carbohydrate sulfotransferase 8 and its impact on the immunotherapy efficacy of cancer. Cell Rep. *43*, 113641. https://doi.org/10.1016/j.celrep.2023.113641.
23. Wang, S.H., Chou, W.C., Huang, H.C., Lee, T.A., Hsiao, T.C., Wang, L.H., Huang, K.B., Kuo, C.T., Chao, C.H., Chang, S.J., et al. (2022). Deglycosylation of SLAMF7 in breast cancers enhances phagocytosis. Am. J. Cancer Res. *12*, 4721–4736.
24. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493–500. https://doi.org/10.1038/s41586-024-07487-w.
25. Eynaudi, A., Díaz-Castro, F., Bórquez, J.C., Bravo-Sagua, R., Parra, V., and Troncoso, R. (2021). Differential Effects of Oleic and Palmitic Acids on Lipid Droplet-Mitochondria Interaction in the Hepatic Cell Line HepG2. Front. Nutr. *8*, 775382. https://doi.org/10.3389/fnut.2021.775382.
26. Bideyan, L., López Rodríguez, M., Priest, C., Kennelly, J.P., Gao, Y., Ferrari, A., Rajbhandari, P., Feng, A.C., Tevosian, S.G., Smale, S.T., and Tontonoz, P. (2022). Hepatic GATA4 regulates cholesterol and triglyceride homeostasis in collaboration with LXRs. Genes Dev. *36*, 1129–1144. https://doi.org/10.1101/gad.350145.122.
27. Zeng, C., Zhu, X., Li, H., Huang, Z., and Chen, M. (2024). The Role of Interferon Regulatory Factors in Liver Diseases. Int. J. Mol. Sci. *25*, 6874. https://doi.org/10.3390/ijms25136874.
28. Klune, J.R., Dhupar, R., Kimura, S., Ueki, S., Cardinal, J., Nakao, A., Nace, G., Evankovich, J., Murase, N., Tsung, A., and Geller, D.A. (2012). Interferon regulatory factor-2 is protective against hepatic ischemia-reperfusion injury. Am. J. Physiol. Gastrointest. Liver Physiol. *303*, G666–G673. https://doi.org/10.1152/ajpgi.00050.2012.
29. Sheikh, A.A., and Utzschneider, D.T. (2022). IRF2 integrates inflammatory signals to balance T cell exhaustion. Immunity *55*, 2225–2227. https://doi.org/10.1016/j.immuni.2022.11.010.
30. Lukhele, S., Rabbo, D.A., Guo, M., Shen, J., Elsaesser, H.J., Quevedo, R., Carew, M., Gadalla, R., Snell, L.M., Mahesh, L., et al. (2022). The transcription factor IRF2 drives interferon-mediated CD8(+) T cell exhaustion to restrict anti-tumor immunity. Immunity *55*, 2369–2385.e10. https://doi.org/10.1016/j.immuni.2022.10.020.
31. Douna, H., Smit, V., van Puijvelde, G.H.M., Kiss, M.G., Binder, C.J., Bot, L., Kuchroo, V.K., Lichtman, A.H., Kuiper, J., and Foks, A.C. (2022). Tim-1 mucin domain-mutant mice display exacerbated atherosclerosis. Atherosclerosis *352*, 1–9. https://doi.org/10.1016/j.atherosclerosis.2022.05.017.
32. Kobiyama, K., and Ley, K. (2018). Atherosclerosis. Circ. Res. *123*, 1118–1120. https://doi.org/10.1161/CIRCRESAHA.118.313816.
33. Xiao, S., Brooks, C.R., Sobel, R.A., and Kuchroo, V.K. (2015). Tim-1 is essential for induction and maintenance of IL-10 in regulatory B cells and their regulation of tissue inflammation. J. Immunol. *194*, 1602–1608. https://doi.org/10.4049/jimmunol.1402632.
34. Koob, S., Barrera, M., Anand, R., and Reichert, A.S. (2015). The non-glycosylated isoform of MIC26 is a constituent of the mammalian MICOS complex and promotes formation of crista junctions. Biochim. Biophys. Acta *1853*, 1551–1563. https://doi.org/10.1016/j.bbamcr.2015.03.004.
35. Turkieh, A., Caubère, C., Barutaut, M., Desmoulin, F., Harmancey, R., Galinier, M., Berry, M., Dambrin, C., Polidori, C., Casteilla, L., et al. (2014). Apolipoprotein O is mitochondrial and promotes lipotoxicity in heart. J. Clin. Investig. *124*, 2277–2286. https://doi.org/10.1172/JCI74668.
36. Chen, J., Hu, J., Guo, X., Yang, Y., Qin, D., Tang, X., Huang, Z., Wang, F., Hu, D., Peng, D., and Yu, B. (2024). Apolipoprotein O modulates cholesterol metabolism via NRF2/CYB5R3 independent of LDL receptor. Cell Death Dis. *15*, 389. https://doi.org/10.1038/s41419-024-06778-4.
37. Tian, F., Wu, C.L., Yu, B.L., Liu, L., and Hu, J.R. (2017). Apolipoprotein O expression in mouse liver enhances hepatic lipid accumulation by impairing mitochondrial function. Biochem. Biophys. Res. Commun. *491*, 8–14. https://doi.org/10.1016/j.bbrc.2017.06.128.
38. Kichaev, G., Yang, W.Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A.L., Kraft, P., and Pasaniuc, B. (2014). Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet. *10*, e1004722. https://doi.org/10.1371/journal.pgen.1004722.
39. Hormozdiari, F., Kostem, E., Kang, E.Y., Pasaniuc, B., and Eskin, E. (2014). Identifying causal variants at loci with multiple signals of association. Genetics *198*, 497–508. https://doi.org/10.1534/genetics.114.167908.
40. McAfee, J.C., Lee, S., Lee, J., Bell, J.L., Krupa, O., Davis, J., Insigne, K., Bond, M.L., Zhao, N., Boyle, A.P., et al. (2023).

Systematic investigation of allelic regulatory activity of schizophrenia-associated common variants. Cell Genom. *3*, 100404. https://doi.org/10.1016/j.xgen.2023.100404.

41. Abell, N.S., DeGorter, M.K., Gloudemans, M.J., Greenwald, E., Smith, K.S., He, Z., and Montgomery, S.B. (2022). Multiple causal variants underlie genetic associations in humans. Science *375*, 1247–1254. https://doi.org/10.1126/science.abj5117.
42. Hua, J.T., Ahmed, M., Guo, H., Zhang, Y., Chen, S., Soares, F., Lu, J., Zhou, S., Wang, M., Li, H., et al. (2018). Risk SNP-Mediated Promoter-Enhancer Switching Drives Prostate Cancer through lncRNA PCAT19. Cell *174*, 564–575.e18. https://doi.org/10.1016/j.cell.2018.06.014.
43. Rusu, V., Hoch, E., Mercader, J.M., Tenen, D.E., Gymrek, M., Hartigan, C.R., DeRan, M., von Grotthuss, M., Fontanillas, P., Spooner, A., et al. (2017). Type 2 Diabetes Variants Disrupt Function of SLC16A11 through Two Distinct Mechanisms. Cell *170*, 199–212.e20. https://doi.org/10.1016/j.cell.2017.06.011.
44. Long, E., Patel, H., Golden, A., Antony, M., Yin, J., Funderburk, K., Feng, J., Song, L., Hoskins, J.W., Amundadottir, L.T., et al. (2024). High-throughput characterization of functional variants highlights heterogeneity and polygenicity underlying lung cancer susceptibility. Am. J. Hum. Genet. *111*, 1405–1419. https://doi.org/10.1016/j.ajhg.2024.05.021.
45. George, J., Zhang, Y., Sloan, J., Sims, J.M., Imig, J.D., and Zhao, X. (2021). Tim-1 Deficiency Aggravates High-Fat Diet-Induced Steatohepatitis in Mice. Front. Immunol. *12*, 747794. https://doi.org/10.3389/fimmu.2021.747794.
46. Mori, Y., Ajay, A.K., Chang, J.H., Mou, S., Zhao, H., Kishi, S., Li, J., Brooks, C.R., Xiao, S., Woo, H.M., et al. (2021). KIM-1 mediates fatty acid uptake by renal tubular cells to promote progressive diabetic kidney disease. Cell Metab. *33*, 1042–1061.e7. https://doi.org/10.1016/j.cmet.2021.04.004.