Supplementary Materials

MEMBERS OF THE NONALCOHOLIC STEATOHEPATITIS CLINICAL RESEARCH NETWORK

Adult Clinical Centers

Cleveland Clinic Foundation, Cleveland, OH: Daniela Allende, MD; Srinivasan Dasarathy, MD; Arthur J. McCullough, MD; Revathi Penumatsa, MPH; Jaividhya Dasarathy, MD

Columbia University, New York, NY: Joel E. Lavine, MD, PhD

Duke University Medical Center, Durham, NC: Manal F. Abdelmalek, MD, MPH; Mustafa Bashir, MD; Stephanie Buie; Anna Mae Diehl, MD; Cynthia Guy, MD; Christopher Kigongo, MB, CHB; Mariko Kopping, MS, RD; David Malik; Dawn Piercy, MS, FNP

Indiana University School of Medicine, Indianapolis, IN: Naga Chalasani, MD; Oscar W. Cummings, MD; Samer Gawrieh, MD; Linda Ragozzino, RN; Kumar Sandrasegaran, MD; Raj Vuppalanchi, MD

Saint Louis University, St Louis, MO: Elizabeth M. Brunt, MD (2002-2008); Theresa Cattoor, RN; Danielle Carpenter, MD; Janet Freebersyser, RN; Debra King, RN (2004-2015); Jinping Lai, MD (2015-2016); Brent A. Neuschwander-Tetri, MD; Joan Siegner, RN (2004-2015); Susan Stewart, RN (2004-2015); Susan Torretta; Kristina Wriston, RN (2015)

Swedish Medical Center, Seattle, WA: Maria Cardona Gonzalez; Jodie Davila; Manan Jhaveri, MD; Kris V. Kowdley, MD; Nizar Mukhtar, MD; Erik Ness, MD; Michelle Poitevin; Brook Quist; Sherilynn Soo

University of California San Diego, San Diego, CA: Brandon Ang; Cynthia Behling, MD, PhD; Archana Bhatt; Rohit Loomba, MD, MHSc; Michael S. Middleton, MD, PhD; Claude Sirlin, MD

University of California San Francisco, San Francisco, CA: Maheen F. Akhter, BS; Nathan M. Bass, MD, PhD (2002-2011); Danielle Brandman, MD, MAS; Ryan Gill, MD, PhD; Bilal Hameed, MD; Jacqueline Maher, MD; Norah Terrault, MD, MPH; Ashley Ungermann, MS

University of Washington Medical Center, Seattle, WA: Matthew Yeh, MD, PhD

Virginia Commonwealth University, Richmond, VA: Sherry Boyett, RN, BSN; Melissa J. Contos, MD; Sherri Kirwin; Velimir AC Luketic, MD; Puneet Puri, MD (2009-2017); Arun J. Sanyal, MD; Jolene Schlosser, RN, BSN; Mohammad S. Siddiqui, MD; Leslie Yost-Schomer, RN

Washington University, St. Louis, MO: Elizabeth M. Brunt, MD (2008-2015); Kathryn Fowler, MD (2012-2015)

Resource Centers

National Cancer Institute, Bethesda, MD: David E. Kleiner, MD, PhD

National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD: Edward C. Doo, MD; Sherry Hall, MS; Jay H. Hoofnagle, MD; Jessica J. Lee, MD, MMSc; Patricia R. Robuck, PhD, MPH (2002-2011); Averell H. Sherker, MD; Rebecca Torrance, RN, MS

Data Coordinating Center, Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD: Patricia Belt, BS; Jeanne M. Clark, MD, MPH; John Dodge; Michele Donithan, MHS; Erin Hallinan, MHS; Milana Isaacson, BS; Mariana Lazo, MD, PhD, ScM; Jill Meinert; Laura Miriel, BS; Jacqueline Smith, AA; Michael Smith, BS; Alice Sternberg, ScM; James Tonascia, PhD; Mark L. Van Natta, MHS; Annette Wagoner; Laura A. Wilson, ScM; Goro Yamada, PhD, MHS, MHS, MMS; Katherine Yates, ScM

SUPPLEMENTARY METHODS

Hepatitis C Cohort

For validation of associations with liver fat content, we analyzed data from a cohort of adult Caucasian patients with chronic hepatitis C who were seen at NIH Clinical Center Liver Clinic. These patients were eligible for inclusion if they consented to genetic testing and a liver biopsy was available for analysis. Subjects infected with genotype 3 were excluded to avoid confounding by virally-mediated steatosis (1-3). Liver histology in the hepatitis C samples was scored semi-quantitatively by a single pathologist; inflammation was scored according to the HAI scoring system (4), fibrosis using the Ishak score (5), and steatosis on a scale of 0-4.

Michigan Genomics Initiative (MGI) Cohort

Individuals in the University of Michigan Health System provided informed consent for broad long-term use of their electronic health information and genetic data.

MGI Phenotypes

Outpatient lab values and diagnoses (using ICD-9) were extracted from the electronic health record from individuals seen between January 2012 and December 31, 2015. Alanine aminotransferase and aspartate aminotransferase were measured in blood from outpatient visits only. We calculated the mean and standard deviation (SD) for each trait and each individual and excluded any measures that were more than 1 SD from the mean to decrease entry errors. We then inverse normally-transformed continuous traits across all individuals measured for that trait before analysis. We defined individuals as cirrhotic if they had a diagnosis of cirrhosis using the ICD-9 code of 571.2, 571.5, and 571.6; individuals without this diagnosis were used as controls.

MGI Genotypes

Genotypes were obtained as previously described (6). Briefly, samples were genotyped using the Illumina HumanCoreExome v12.1 array with custom supplementary genotypes. Only individuals with call rate >99% were used for analyses. Variants with HWE > 0.0001 and call rate > 99 were used for imputation (n=493,463). Phasing was carried out using SHAPEIT2 (v2.r837) (7) on autosomal chromosomes and genotypes imputed using Haplotype Reference Consortium (chromosome 1-22: HRC release 1; chromosome X: HRC release 1.1; http://www.haplotype-reference-consortium.org/) using Minimac3 (v1.0.13) (8). Variants with low imputation quality (R2 <0.3) were removed for a final panel 39,127,678 million SNPs.

Genotypes from 35888 individuals were used to calculate principal components in LASER(9, 10) using the worldwide(HGDP) panel. Using the KNN classification method (11) and these principal components we identified 32505 individuals with European ancestry. We recalculated the principal components of these 32505 individuals using the Europe(POPRES) panel in LASER and these latter principal components were used in analyses. 31226 unrelated individuals (no 1st- or 2nd-degree relatives) were then selected using KING (12); 31221 European ancestry individuals with no missing data were used for final analyses.

MGI Association analyses

Association analyses between phenotypes and genotypes were performed using mixed model regression controlling for age, age^2, sex, and the first 10 PCs, limited to unrelated individuals of European ancestry using RARE METAL WORKER

(https://genome.sph.umich.edu/wiki/RAREMETALWORKER) (13). Linear effects on prevalence were converted to odds ratios as described (14).

UK Biobank Cohort

The UK Biobank is a prospective population based epidemiological study of over 450,000 deeply phenotyped individuals from the United Kingdom where participants have given broad consent for genetic and epidemiological analyses. The design is described extensively elsewhere (15), but diagnosis codes are primarily from in-patient hospital data.

UK Biobank Phenotypes

We analyzed associations with diagnostic codes K74 (Fibrosis and cirrhosis of liver) and K70 (alcoholi liver disease), K70-K77 (Diseases of liver), and K75 (other inflammatory liver diseases), as well as H54 (Blindness and low vision), where individuals with these diagnoses were used as cases and all others used as controls. Both primary and secondary diagnosis codes are included as cases.

UK Biobank Genotypes

Genotypes of the UK Biobank participants were assayed using either of two genotyping arrays, the Affymetrix UK BiLEVE Axiom or Affymetrix UKBiobank Axiom array. SNPs were Quality controlled and imputed to ~96 million genetic variants from the Haplotype Reference Consortium, the thousand genomes and the UK 10K projects (16).

UK Biobank Association analyses

Individuals were excluded if they were designated by the UK Biobank as outliers based on either genotyping missingness rate or heterogeneity, whose sex inferred from the genotypes did not match their self-reported sex and who were not of White-British ancestry. Finally, individuals were removed if they had missingness >5% across variants which passed QC procedures. Analyses were limited to the White-British cohort. A genetic relatedness matrix was constructed using 93,511 high quality genotyped variants. Association analyses were carried out in UK Biobank using SAIGE (17) (controlling for sex, age, age^2, and the first 10 genomic principal components as computed by the UK Biobank).

SNP Selection

We genotyped the SNPs in Table 1 of Chambers et al, omitting SNPs located near the genes for alkaline phosphatase and GGT, as well as those near *PNPLA3* and *CPN1*, which we have previously genotyped in this study population (18). If a genotyping assay was predicted to fail, an alternative SNP in linkage equilibrium (LD) with the GWAS marker SNP was selected for genotyping. For loci that showed suggestive association with NAFLD, we then performed in-depth genotyping of variants in or near select genes suspected to be related with the tag SNPs. We aimed to genotype known coding and splice site SNPs, as well as non-coding SNPs that capture most of the variability in the gene of interest. Tagger (http://www.broadinstitute.org/mpg/tagger/) was used to select non-coding SNPs. LD calculations between SNPs were performed using LDLink (http://analysistools.nci.nih.gov/LDlink/) (19), SNAP (http://www.broadinstitute.org/mpg/snap/), and GLIDERS (http://www.sanger.ac.uk/resources/software/gliders/), limited to populations of European ancestry.

SNP Genotyping
SNPs in the NAFLD, HCV and control cohorts were genotyped using the Sequenom MassARRAY iPLEX Gold platform (Sequenom, Inc; San Diego, CA), with PCR primers purchased from Invitrogen (Carlsbad, CA). Genotyping concordance was 100% as determined by inclusion of 16-32 replicate NAFLD patient samples and 100% using 96 replicate control population samples. The average genotyping call rate across all SNP assays was 99.5% and all call rates exceeded 95%. Genotyping of rs6834314 in additional samples, and genotyping of rs58542926 (TM6SF2), and rs626283 (MBOAT7) was performed using TaqMan® Pre-Designed SNP Genotyping Assays purchased from Invitrogen (Carlsbad, CA).

Hepatic Gene Expression
Samples were homogenized in TRIzol Reagent (Life Technologies, Carlsbad, CA), and total RNA was isolated according to manufacturer instructions; cDNA was synthesized using 1st Strand cDNA Synthesis System for Quantitative RT-PCR (OriGene, Rockville, MD) from 1 μg of total RNA in 20 μl reaction volume and diluted 5-fold in nuclease-free water before PCR. Gene expression was quantified using TaqMan real-time PCR assay and primers (Integrated DNA Technology, Coralville, IA). Primer sequences can be found in Supplementary Table 12.

The associations between genotype and tissue gene expression were queried from the Genotype-Tissue Expression (GTEx) Project, which was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from the GTEx Portal on September 21, 2017.

Primary Human Hepatocytes
Primary human hepatocytes (PHH) were obtained from Dr. Stephen Strom at the University of Pittsburgh, through the NIH-funded Liver Tissue and Cell Distribution System (LTCDS). Genomic DNA and total RNA were extracted. The cells were genotyped for select SNPs using TaqMan SNP Genotyping Assay (Life Technologies).

Cell Culture
HepG2 and HEK293 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Sigma, St. Louis, MO) supplemented with MEM Non-Essential Amino Acid Solution, 100× (MEM NEAA, Life Technologies, Carlsbad, CA) and 10% fetal bovine serum (Sigma, St. Louis, MO) in a humidified atmosphere of 5% CO_2 at 37°C. HSD17B13-GFP plasmid was obtained from OriGene (Rockville, MD). *HSD17B13*-Flag plasmid was obtained from VectorBuilder.Com (Cyagen Biosciences, Santa Clara, CA) and the sequence was confirmed by sequencing. Mutant plasmids were generated using Q5 Site-Directed Mutagenesis Kit (New England Biolabs Inc., Ipswich, MA); custom primer sequences can be found in Supplementary Table 12. Plasmids were transfected into cells using Lipofectamine 3000 (Life Technologies, Carlsbad, CA) according to the manufacturer's instruction. Lipid droplet formation was induced with 200 μM of oleate and 200 μM of palmitate (Sigma Aldrich, St. Louis, MO).

Immunofluorescence Staining

Cells seeded in Lab-Tek II chamber slides were fixed by 4% paraformaldehyde and blocked and permeabilized with 3% BSA (Sigma Aldrich, St. Louis, MO), 10% normal goat serum (Vector Laboratories, Burlingame, CA), and 0.3% Triton X-100 (Sigma Aldrich, St. Louis, MO) before incubation with primary antibodies (Cell Signaling Technology, Danvers, MA and Abcam, Cambridge, MA). Corresponding Alexa Fluor secondary antibodies (Life Technologies, Carlsbad, CA) were used following the wash steps. Hoechst 33342 and LipidTox Red (Life Technologies, Carlsbad, CA) were added prior to confocal microscope observation to stain nuclei and lipid droplets, respectively.

Lipid Droplet Isolation

Lipid droplets were isolated from HepG2 cells stably transfected with HSD17B13-GFP plasmid as described in previous reports (20, 21) with modifications. Briefly, cells seeded in four 10 cm dishes were collected in Tris-HCl buffer (20 mM, pH 7.4) with 1 mM EDTA, 250 mM Sucrose, and protease inhibitor (Roche, Indianapolis, IN), homogenized on ice, and centrifuged for 10 min at 1000 g, at 4°C. 70% Sucrose buffer was added to the supernatant to make a final concentration of 26% sucrose buffer. 51%, 43%, 35%, 26% (supernatant from cell lysis), 18%, 10%, and 2% sucrose buffer were sequentially added to 12 ml centrifuge tube, followed by centrifugation at 28,000 rpm (SW41 rotor, ~190,000 g) for 2 hours, at 4°C. Lipid droplets were collected from the top layer, stained with LipidTox Red (Life Technologies, Carlsbad, CA), and placed on a glass slide for confocal microscope observation.

Lipid Content Measurement in HSD17B13-Tet-on 3G Inducible System

HepG2 cells cultured with Tet System Approved FBS (Clontech, Mountain View, CA) were transfected with pCMV-Tet3G (kind gift from Dr. Haiqing Fu, NCI) and TRE-HSD17B13-GFP (Vector builder, TRE3G>hHSD17B13:EGFP:IRES:Puro) in 96 well plates. Doxycycline (0-50 ng/ml, Sigma Aldrich, St. Louis, MO) was used to induce HSD17B13 expression in transfected groups and used to test if Doxycycline itself would change lipid content in non-transfected groups. The following day lipid droplet formation was induced with oleate (200 μM) supplementation. Lipid content was measured by fluorescence intensity of Nile Red staining after 48 hours of transfection.

Stable HepG2 Cell Lines

HepG2 cells were transfected with plasmid expressing HSD17B13 (A) (OriGene, Rockville, MD) or CRISPR-Cas 9 knock-down system plasmids (Santa Cruz Biotechnology, Dallas, Texas) to overexpress or knock down *HSD17B13* gene, respectively. Cells were cultured for 2-4 weeks under selection of G418 (Life Technologies, Carlsbad, CA) for overexpression cell lines or Puromycin (Life Technologies, Carlsbad, CA) for knock down cells lines. Stable cell lines from single colonies were checked under fluorescence microscope and confirmed by qPCR for overexpression or knock down efficiency.

Retinol Dehydrogenase (RDH) Activity Assay

HEK293 cells are commonly used for ectopic protein expression, their endogenous RDH activity is well-defined, and they have been previously established as a mature platform for RDH enzymatic activity measurements (22). Cells were plated in 35-mm dishes and transiently transfected in triplicate with 1.5 μg/well of expression vectors for HSD17B13, RDH10 as positive control, or empty vector. One day after transfection, growth medium was replaced with 2 ml of

fresh medium containing 2 or 5 μM all-trans-retinol (Toronto Research Chemicals, Toronto, Canada) in ethanol, with a final ethanol concentration ≤ 0.5% (v/v). After 8 hours of incubation, media and cells were collected separately under reduced light and aliquots of cell suspensions were taken for protein quantification and western blot analysis. Medium samples were mixed with an equal volume of ethanol and extracted with a double volume of hexane. Hexane phase was transferred into siliconized glass tubes, aqueous phase was acidified by the addition of 150 μl of 4 N hydrochloric acid, and hexane extraction was repeated. Extractions were pooled, dried under N_2 gas, reconstituted into hexane and separated by normal phase HPLC with Spherisorb S3W column (4.6 mm×100 mm) (Waters Corp., Milford, MA) in hexane:ethyl acetate:acetic acid (95:4.975:0.025) mobile phase at the flow rate of 0.6 ml/min using Waters Alliance 2695 HPLC system equipped with a Waters 2996 photodiode array. Peaks were identified by comparison to retention times of retinoid standards and evaluation of wavelength maxima and quantified as described previously (23).

For Western blot analysis, 50 μg of total cellular protein were separated in 12% SDS-PAGE, transferred to the PVDF membrane and probed with anti-FLAG antibody (Sigma Aldrich, St. Louis, MO) at 1:2,500 dilution.

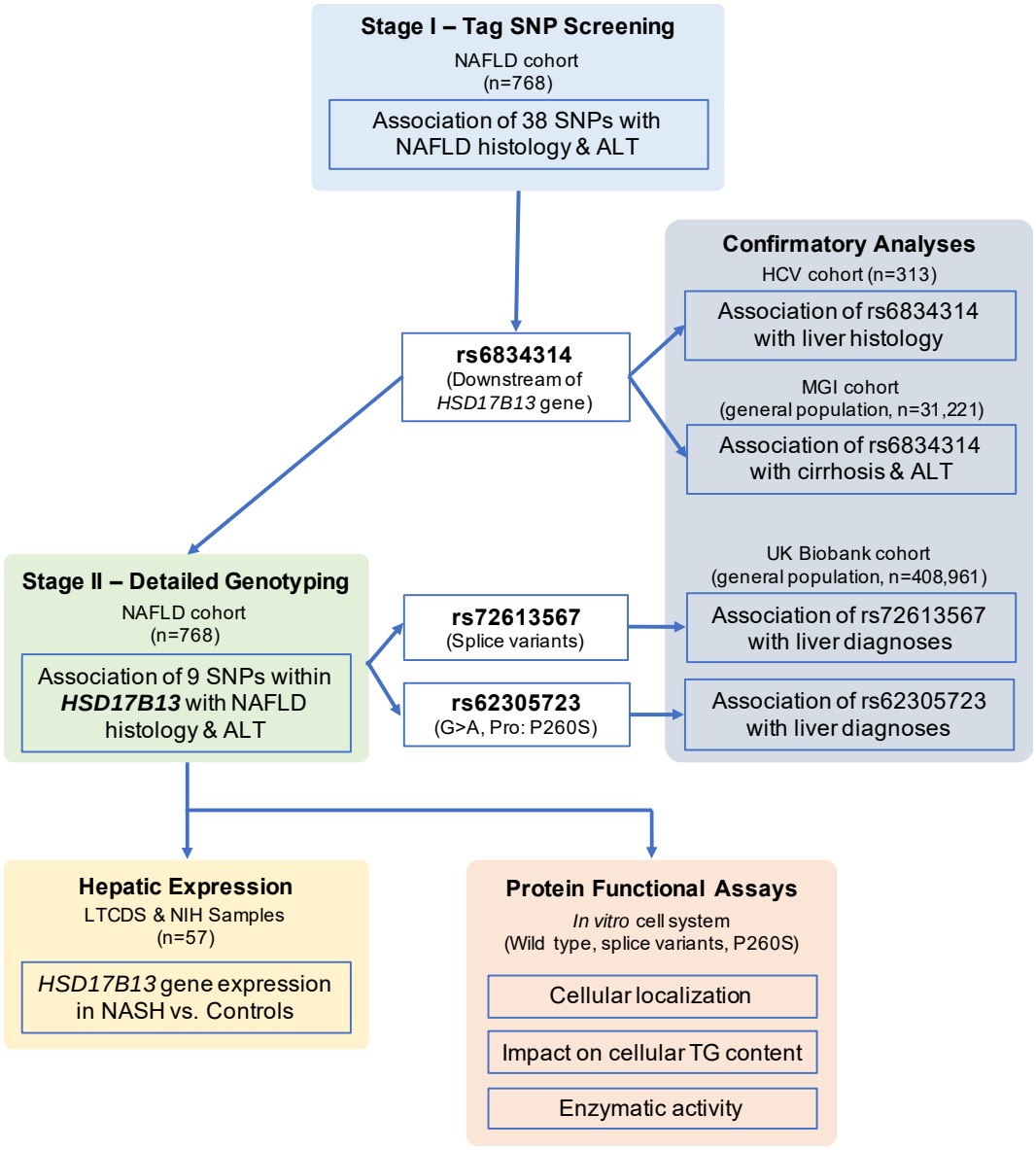
Splice Variant Detection and Quantitation

The delta 6 and G-insertion variants were detected in liver samples from LTCDS by amplifying select regions of the HSD17B13 gene with primers designed using the NCBI-Primer-BLAST primer designing tool (Supplementary Table 10) and Invitrogen SuperFi Master Mix (thermo scientific cat. 12358010), followed by gel purification of the PCR product on a 1.5% agarose gel, purification of the product band with the QIAquick Gel Extraction Kit (Qiagen cat. 28704/28706), and sequencing of the purified product with the forward primer of the pair used for amplification (Eurofins Genomic Sequencing, Louisville, KY).

Custom primers, designed using the NCBI Primer-BLAST primer designing tool, were used to quantify the amount of delta 6 and exon-6 containing variants in human liver samples from LTCDS. Multiple primer pairs were tested and optimized for sensitivity and specificity. Quantitation assays were performed in duplicate by qRT-PCR, using PowerUp SYBR Green Master Mix (thermos scientific Cat. A25776) with 0.05μM of forward and reverse primers and 1 μL of cDNA. Transcript copy number was estimated by interpolating Ct values to standard curves generated using delta 6 and HSD17B13(A) plasmid standards.

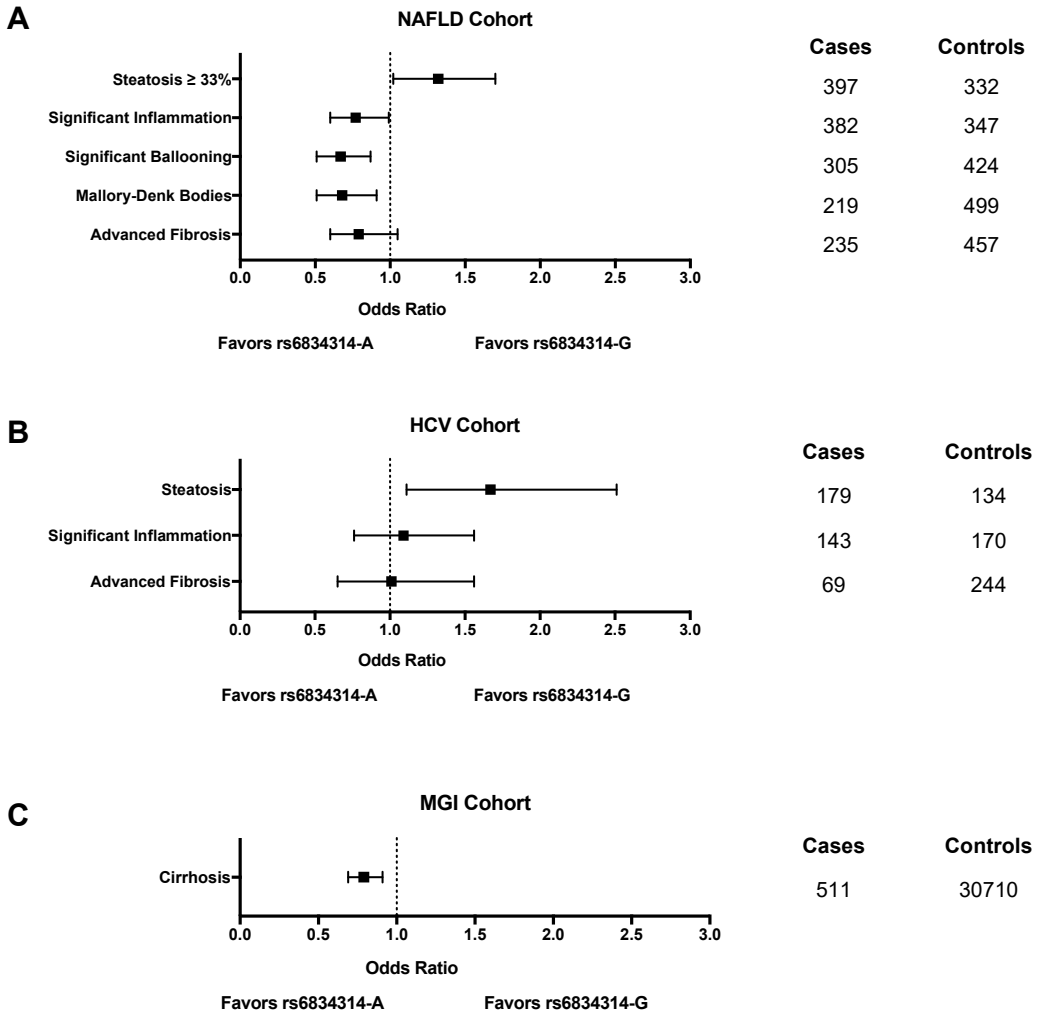
SUPPLEMENTARY FIGURES

Supplementary Figure 1



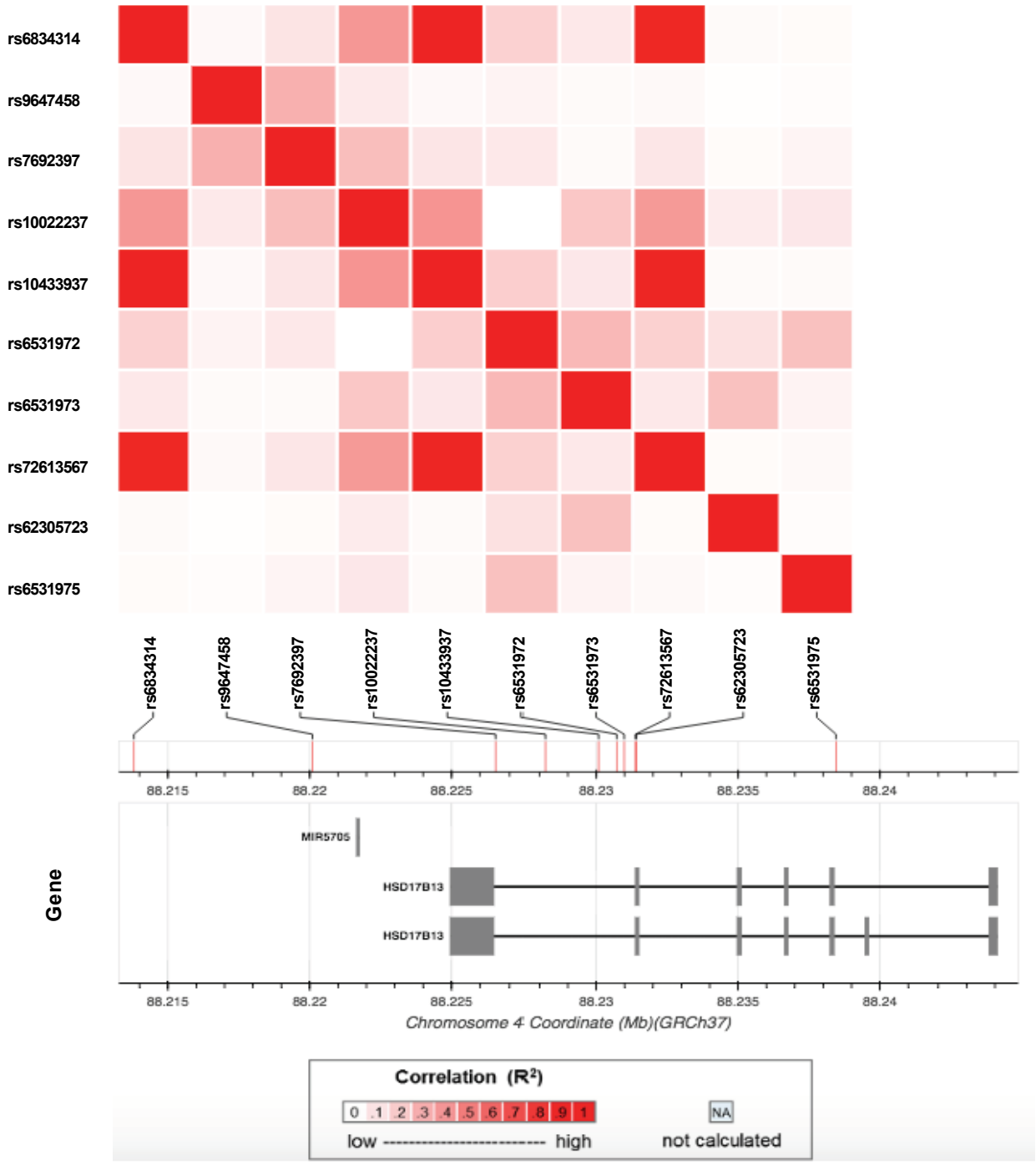
Supplementary Figure 1. Study Design. Initial screening included 38 SNPs described by Chambers et al (24) in a cohort of subjects with biopsy-proven NAFLD to identify loci associated with histological features of the disease. The association of the most interested SNP (rs6834314) with liver enzyme, cirrhosis diagnosis, and liver histological features were further validated in Michigan Genomic Initiative (MGI) cohort and Hepatitis C cohort. After identifying the locus of interest, additional SNPs in the candidate gene locus (*HSD17B13*) in NAFLD cohort were genotyped and confirmed the associations in additional cohort (UK Biobank cohort). Finally, functional analyses were performed to characterize the gene product.

Supplementary Figure 2



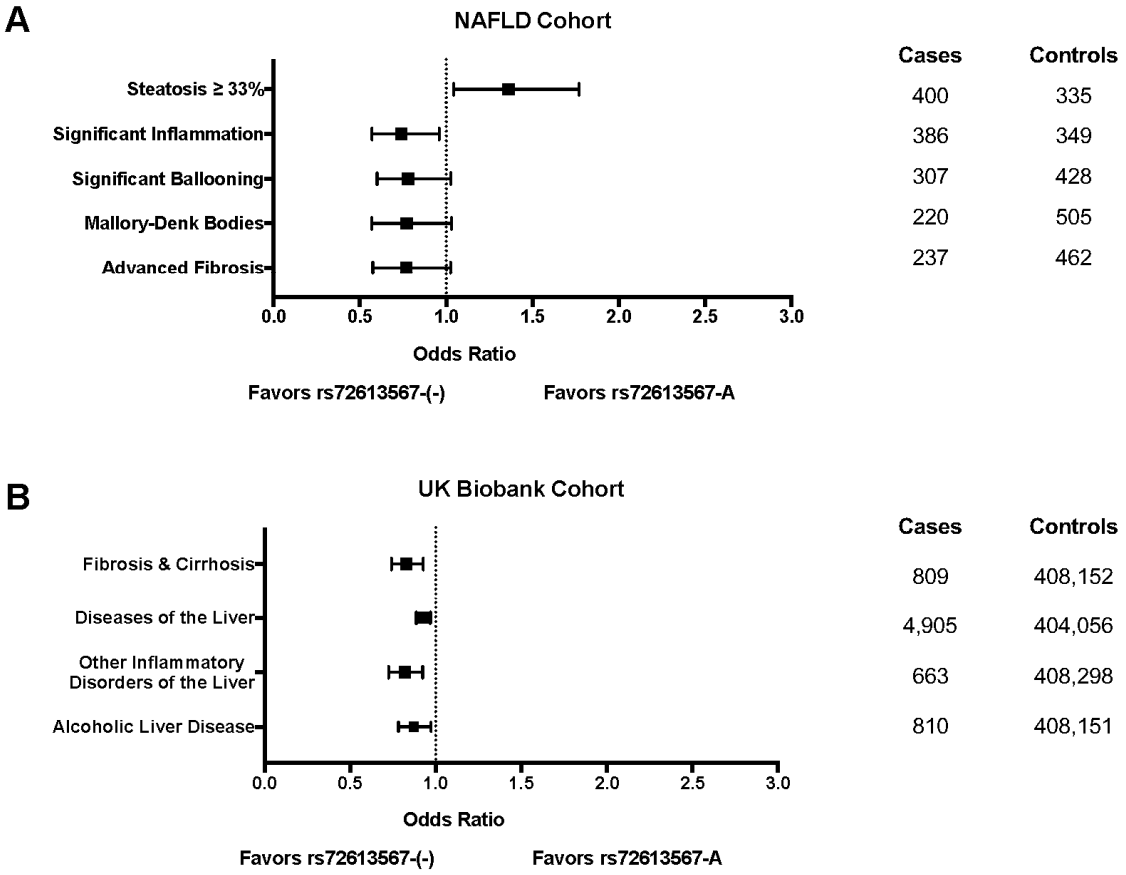
Supplementary Figure 2. Odds ratios for associations with rs6834314. (A) Odds ratios for association of rs6834314 genotype in the NAFLD cohort significant steatosis (≥33% vs <33%), significant inflammation (combined portal and lobular inflammatory score ≥3 vs. <3), significant ballooning (many cells vs. none or few), Mallory-Denk bodies (many vs. rare or none) and significant fibrosis (≥ bridging fibrosis vs. less). Odds ratios calculated using binary logistic regression with an additive inheritance model, controlled for age, gender and BMI. (B) Odds ratios for association of rs6834314 genotype in the chronic HCV cohort with histological steatosis (present vs. absent), significant inflammation (Knodell HAI score ≥8 vs. <8), and significant fibrosis (Ishak fibrosis score ≥4 vs. <4). Odds ratios calculated using binary logistic regression with an additive inheritance model, controlled for age, gender, and BMI. (C) Odds ratio for allelic association of rs6834314 with ICD-9 diagnosis of cirrhosis in the MGI cohort. Association assessed using linear mixed model regression controlling for age, age^2, sex, and the first 10 genomic principal components.

Supplementary Figure 3



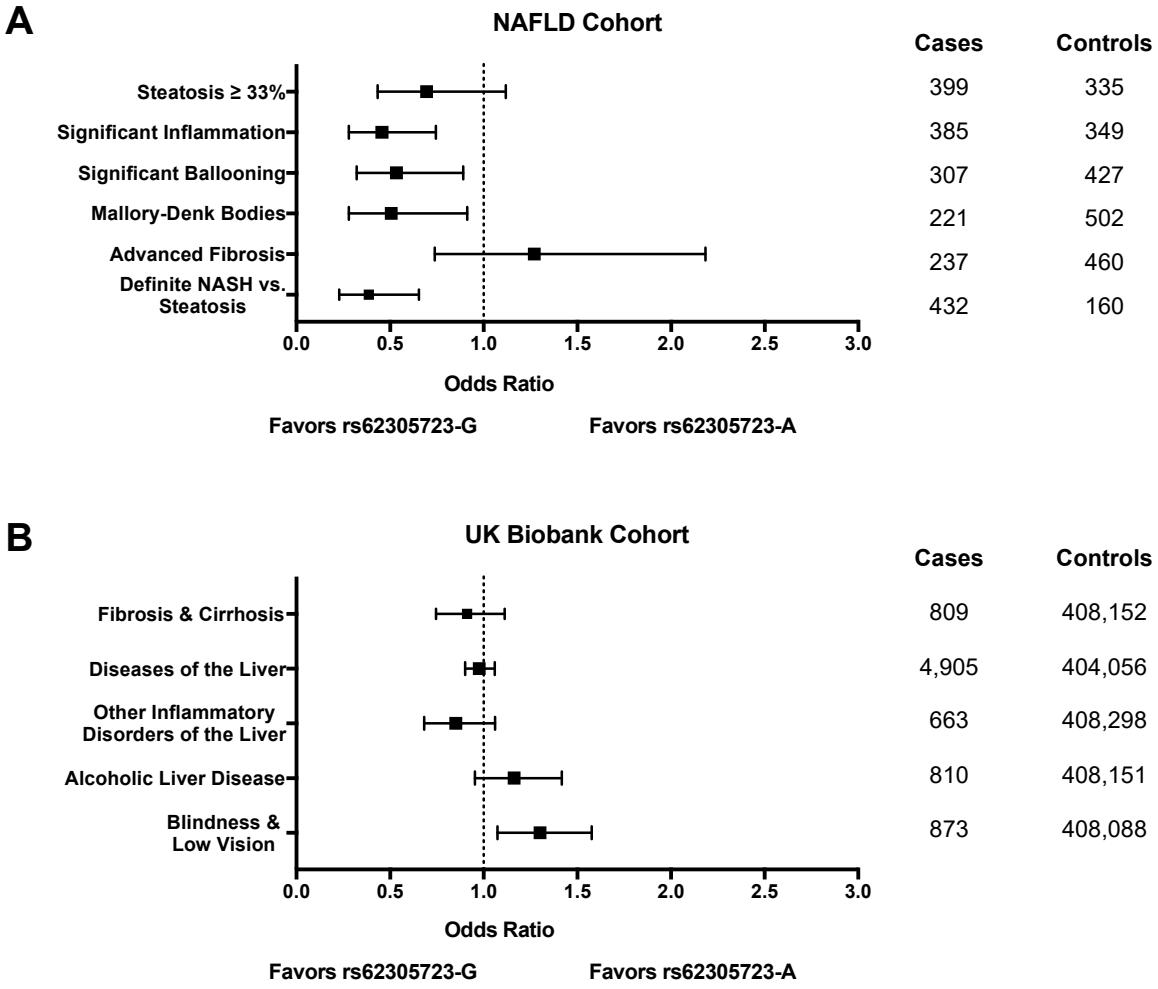
Supplementary Figure 3. Linkage disequilibrium (LD) map of genotyped SNPs in or around *HSD17B13*. Figure generated using LDLink (http://analysistools.nci.nih.gov/LDlink/) with linkage data limited to subjects of European descent (EUR).

Supplementary Figure 4



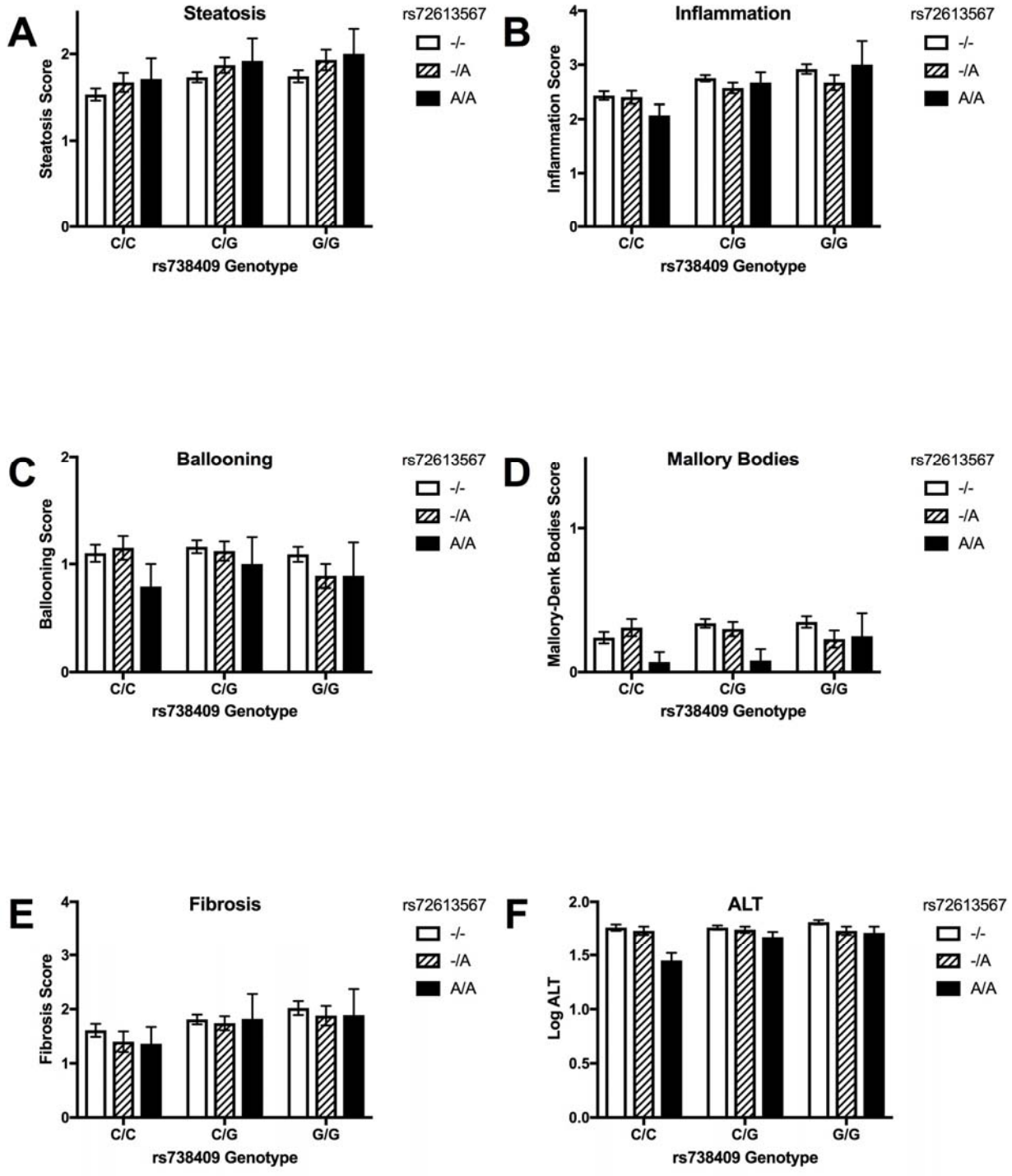
Supplementary Figure 4. Odds ratios for associations with rs72613567. (A) Odds ratios for association of rs72613567 genotype in the NAFLD cohort with at least moderate steatosis (≥33% vs <33%), significant inflammation (combined portal and lobular inflammatory score ≥3 vs. <3), significant ballooning (many cells vs. none or few), Mallory-Denk bodies (many vs. rare or none) and significant fibrosis (≥ bridging fibrosis vs. less). Odds ratios calculated using binary logistic regression with an additive inheritance model, controlled for age, gender, and BMI. (B) Odds ratios for allelic association of rs72613567 with liver-related diagnoses in the UK Biobank cohort. Association assessed using logistic mixed model controlling for sex, age, age^2, and the first 10 genomic principal components. Odds ratios were estimated from the β-coefficient of the regression.

Supplementary Figure 5



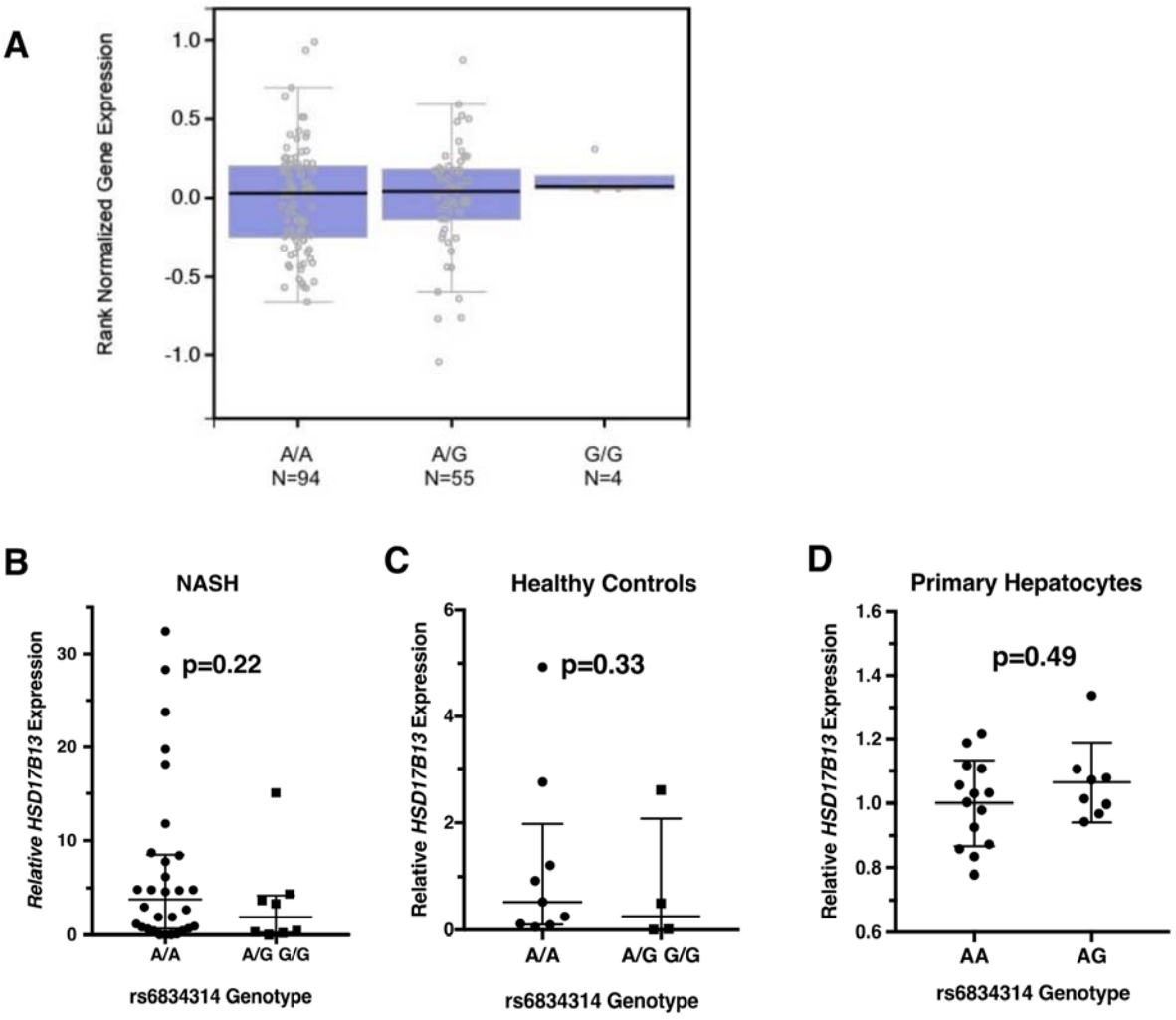
Supplementary Figure 5. Odds ratios for associations with rs62305723. (A) Odds ratios for association of rs62305723 genotype in the NAFLD cohort with at least moderate steatosis (≥33% vs <33%), significant inflammation (combined portal and lobular inflammatory score ≥3 vs. <3), significant ballooning (many cells vs. none or few), Mallory-Denk bodies (many vs. rare or none), significant fibrosis (≥ bridging fibrosis vs. less) and definite NASH (vs. steatosis only). Odds ratios calculated using binary logistic regression with an additive inheritance model, controlled for age, gender and BMI. (B) Odds ratios for allelic association of rs62305723 with diagnoses in the UK Biobank cohort. Association assessed using logistic mixed model controlling for sex, age, age^2, and the first 10 genomic principal components. Odds ratios were estimated from the β-coefficient of the regression.

Supplementary Figure 6



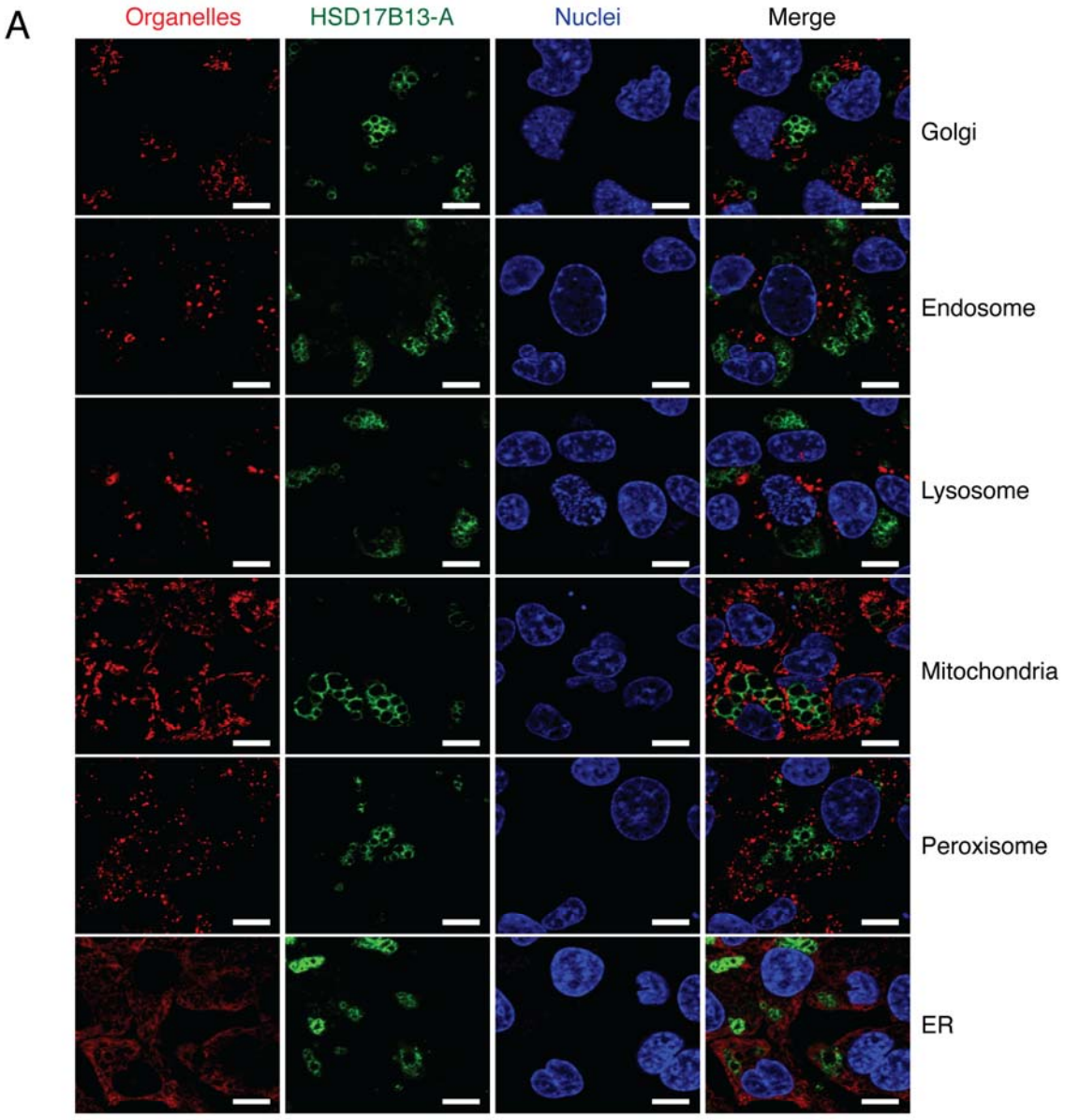
Supplementary Figure 6. *HSD17B13* rs72613567 genotype and features of NAFLD by *PNPLA3* rs738409 genotype in the NAFLD cohort (n=732). (A) Histological steatosis score (B) inflammatory score, (C) ballooning degeneration score, (D) Mallory-Denk bodies, and (E) fibrosis scores from liver biopsies. (F) Log-transformed ALT activity. Mean±SEM.

Supplementary Figure 7



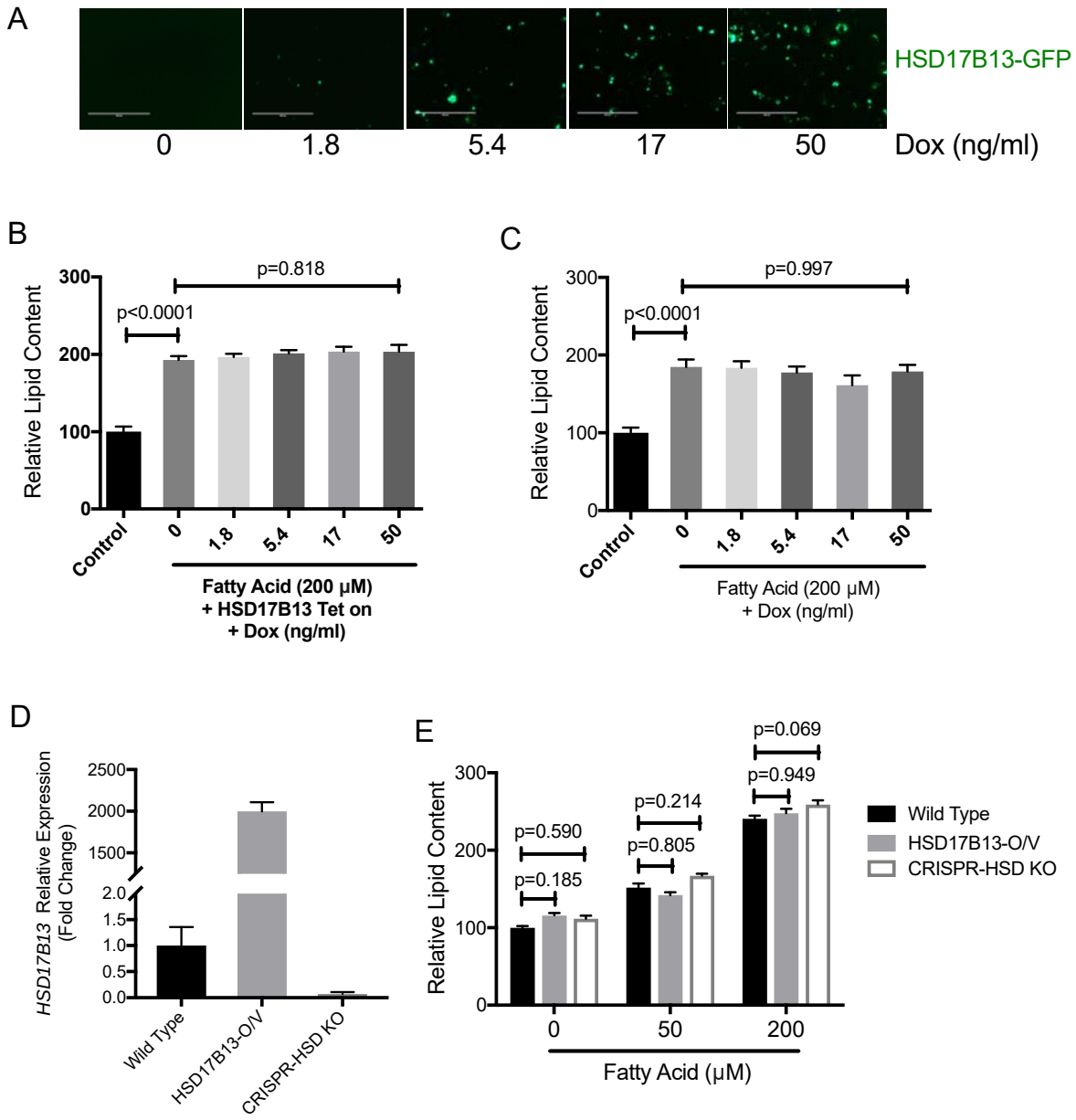
Supplementary Figure 7. rs6834314 and hepatic expression of *HSD17B13*. (A) Hepatic expression by genotype in the GTEx database (accessed September 21, 2017). P=0.60. (B-C) Expression level by genotype in liver samples from subjects with (B) NASH (AA: n=30, GA/GG: n=8) or (C) healthy controls (AA: n=9, GA/GG: n=4). rs6834314A/G and G/G were pooled together due to lower frequency. Lines denote median and interquartile range and data is normalized to controls. Mann-Whitney test was used for significance testing. (D) Expression level by genotype in primary human hepatocytes from 22 donors (AA: n=14, GA: n=8). Mann-Whitney test was used for significance testing.

Supplementary Figure 8



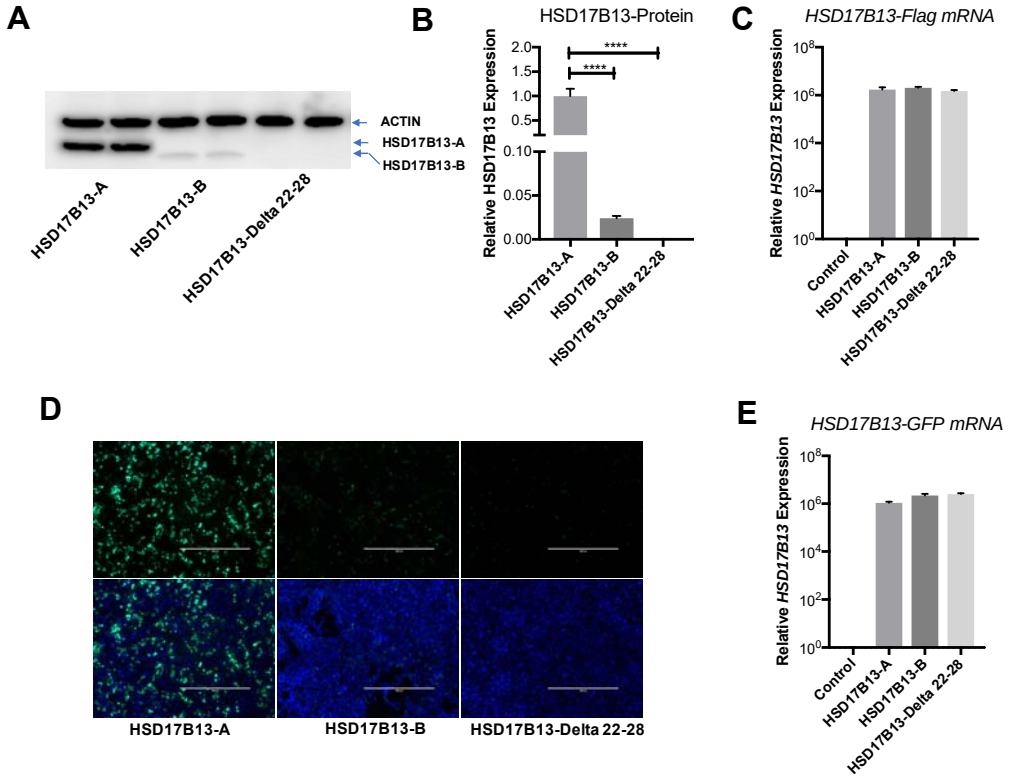
Supplementary Figure 8. HSD17B13 localization. (A) HepG2 cells were transfected with HSD17B13(A)-GFP and stained with Organelle Localization Kit, containing markers for Golgi, endosomes, lysosomes, mitochondria, peroxisomes and endoplasmic reticulum. (B) Amino acids 22-28 in HSD17B13 are conserved within other lipid droplet targeting proteins, ADRP and TIP47.

Supplementary Figure 9



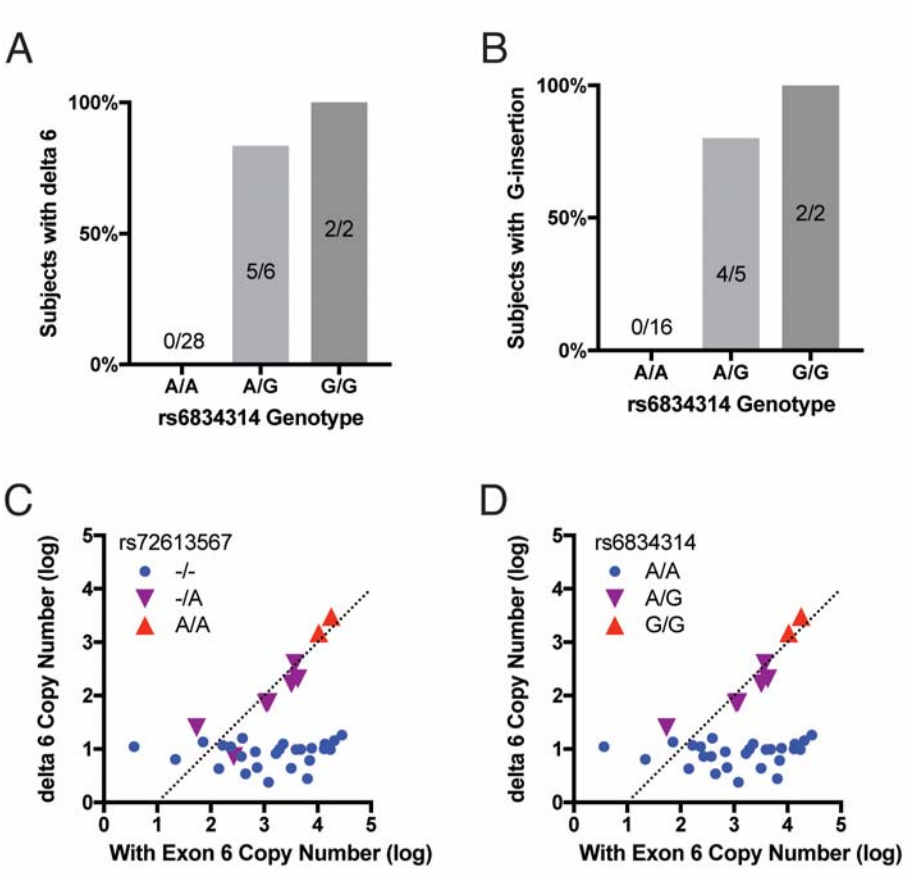
Supplementary Figure 9. HSD17B13 knock-out or over-expression does not affect hepatocyte lipid content. (A) HepG2 cells were transiently transfected with HSD17B13(A)-GFP Tet-on expression system and treated with varying doses of doxycycline (0-50 ng/ml) to induce HSD17B13 expression. Bar indicates 200 μm. (B-C) Oleate and palmitate were added to cells to induce lipid droplet synthesis. Lipid content were determined by the fluorescence intensity of Nile Red staining. (D) Gene expression of *HSD17B13*. and (E) lipid content were measured in wild type HepG2 cells, HepG2 cells with stable overexpression of HSD17B13 (HSD17B13-O/V), and stable knockdown HepG2 cell line (CRISPR-HSD KO). Data are presented as Mean±SEM.

Supplementary Figure 10



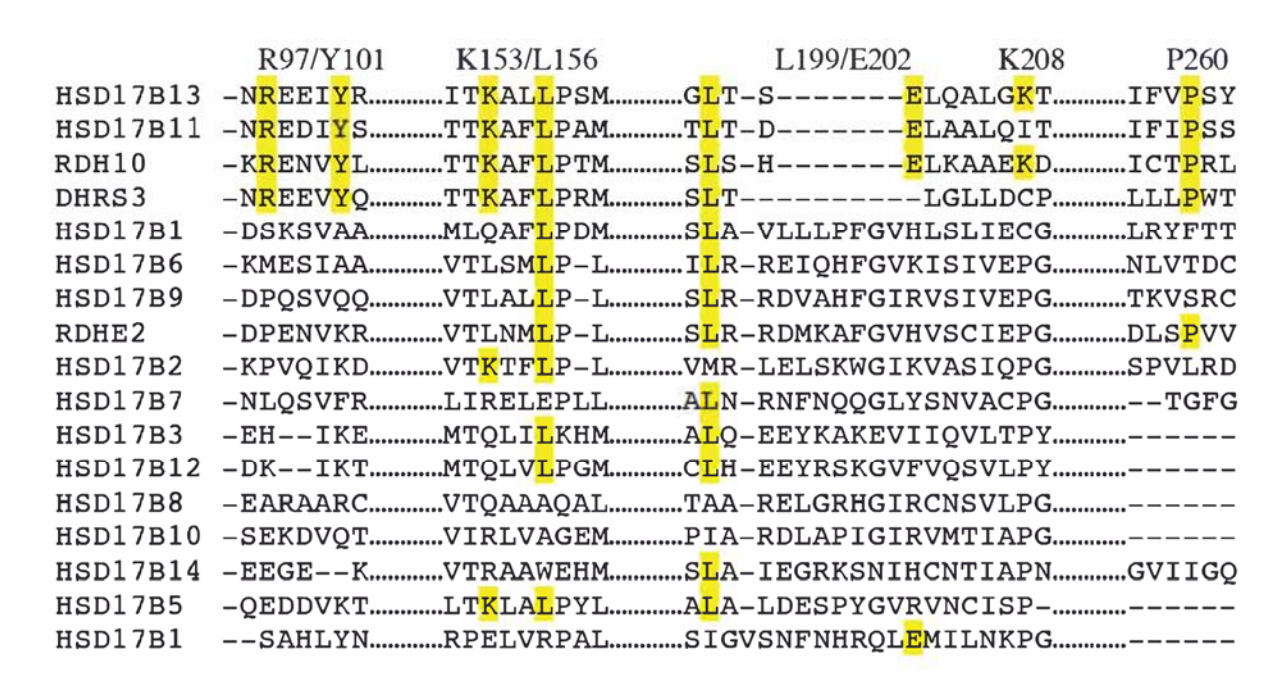
Supplementary Figure 10. Protein stability of HSD17B13variants . (A-C) HSD17B13-A-Flag, HSD17B13-B-Flag, and HSD17B13-Delta 22-28-Flag plasmids were transfected into HEK293 cells. (A) Anti-Flag Western blot was used to determine protein stability; Anti-ACTIN Western blot was used as loading control; (B) Relative protein quantification of three protein variants; data are presented as Mean±SEM, **** $p<0.0001$. (C) Ectopic gene expression level was determined by qPCR, data are presented as Mean±SEM. (D-E) HSD17B13-A-GFP, HSD17B13-B-GFP, and HSD17B13-Delta 22-28-GFP plasmids were transfected into HepG2 cells. (D) GFP was used to visualize protein. Nuclei were counterstained with Hoechst 33342. Bars indicate 400 μm; (E) Ectopic gene expression level was determined by qPCR, data are presented as Mean±SEM.

Supplementary Figure 11



Supplementary Figure 11. Genetic determinants of HSD17B13 splice variants. rs6834314 genotype predicts the presence of the delta-6 (A) or G-insertion (B) splice variants. Subjects with detectable quantities of the delta-6 variant express it at a fixed ratio to transcripts that do not skip exon 6 (C-D). Human liver samples from LTCDS. Variants detected by sequencing and quantified by variant-specific qPCR.

Supplementary Figure 12



Supplementary Figure 12. Alignment of HSD17B13 with HSD family proteins and known RDH proteins. Conserved sites at putative homodimer interaction site (R97/Y101), substrate binding sites (K153/L156, L199/E202, and K208), and P260 were highlighted in yellow.

SUPPLEMENTARY TABLES

Supplementary Table 1 – NAFLD Cohort Characteristics

N	768
Male gender	286 (37.2%)
Data Source	
NASH CRN database	554 (72.1%)
PIVENS	180 (23.4%)
NIH Clinical Center	34 (4.4%)
Age	48.7±11
BMI	34.6±6.4
ALT (median, range)	57 (10-303)
AST	42 (10-180)
GGT	46 (5-793)
Alcoholic drinks <= 2-4/month	95.6%
DM Type 1	4 (0.5%)
DM Type 2	206 (28%)
NAS (median, range)	4 (0-8)
Steatosis grade	
0 (< 5%)	49 (6.4%)
1 (5-33%)	296 (38.6%)
2 (33-66%)	241 (31.4%)
3 (> 66%)	181 (23.6%)
Fibrosis stage	
0 (no fibrosis)	166 (22.8%)
1 (zone 3 or periportal only)	169 (23.3%)
2 (zone 3 and periportal)	143 (19.7%)
3 (bridging)	162 (22.3%)
4 (cirrhosis)	88 (12.1%)
NASH diagnosis	
No NASH	166 (21.7%)
Borderline	144 (18.7%)
Definite NASH	457 (59.5%)

Supplementary Table 2 – Genotyped SNPs

SNP	Alleles	Allele Frequency[A]	Locus	Nearest Gene	Enzyme Affected[B, C]
rs1497406	A/G	0.40/0.60	1p36.13	*EPHA2*	GGT
rs12145922	C/A	0.58/0.42	1p22	*PKN2*	GGT
rs1335645	T/C	0.88/0.12	1p13	*CEPT1*	ALT, GGT
rs10157801	A/G	0.58/0.42	1q21	*DPM3*	(GGT)
rs1260326	C/T	0.52/0.48	2p23	*GCKR*	GGT
rs16856332	T/G	0.96/0.04	2q24	*ABCB11*	AlkP
rs13030978	C/T	0.73/0.27	2q12	*MYO1B*	GGT
rs2140773	G/T	0.59/0.41	2q37	*EFHD1*	GGT
rs10513686	G/A	0.86/0.14	3q26	*SLC2A2*	GGT
rs6834314	A/G	0.79/0.21	4q22	*HSD17B13*	ALT
rs1027841	C/T	0.81/0.19	4q31	*ZNF827*	(GGT)
rs6888304	A/G	0.72/0.28	5p15	*CDH6*	GGT
rs4074793	T/C	0.91/0.09	5q11	*ITGA1*	GGT
rs1126617	A/G	0.25/0.75	6p22	*GPLD1*	AlkP
rs4715438	G/A	0.68/0.32	6p12	*MLIP*	(GGT)
rs17145750	C/T	0.86/0.14	7q11	*MLXIPL*	GGT
rs2954021	A/G	0.58/0.42	8q24	*TRIB1*	ALT, AlkP
rs6984305	A/T	0.14/0.87	8p23	*PPP1R3B*	AlkP
rs10819937	C/G	0.22/0.78	9q21	*C9orf125*	AlkP
rs579459	C/T	0.21/0.80	9q34	*ABO*	AlkP
rs7082470	G/A	0.52/0.48	10q21	*REEP3, JMJD1C*	(AlkP)
rs754466	T/A	0.76/0.24	10q23	*DLG5*	GGT
rs2236653	C/T	0.58/0.42	11q24	*ST3GAL4*	AlkP
rs174601	C/T	0.64/0.36	11q12	*FADS2*	AlkP
rs7310409	A/G	0.42/0.58	12q24	*HNF1A*	GGT
rs10131298	A/T	0.24/0.75	14q32	*EXOC3L4*	(GGT)
rs339969	A/C	0.65/0.35	15q21	*RORA*	GGT
rs8038465	C/T	0.58/0.43	15q23	*CD276*	GGT
rs7186908	G/C	0.80/0.20	16q22	*PMFBP1*	AlkP
rs4581712	C/A	0.73/0.27	16q23	*DYNLRB2*	GGT
rs314253	A/G	0.64/0.36	17p13	*DLG4*	AlkP
rs9913711	C/G	0.65/0.35	17q24	*SOX9*	GGT
rs12968116	C/T	0.89/0.11	18q21.31	*ATP8B1*	GGT
rs4503880	C/T	0.80/0.20	18q21.32	*NEDD4L*	GGT

rs281377	C/T	0.54/0.46	19q13	*FUT2*	AlkP, (GGT)
rs7267979	A/G	0.45/0.55	20p11	*ABHD12*	AlkP
rs1076540	G/A	0.76/0.24	22q11.21	*MICAL3*	GGT
rs2739330	C/T	0.56/0.44	22q11.23	*GSTT2B*	GGT

[A] Alllele frequency in 768 Caucasian with biopsy-proven NAFLD (NAFLD cohort)
[B] Liver enzyme affected in GWAS by Chambers et al (24)
[C] Liver enzyme in parentheses denotes association found by Chambers et al for a SNP in LD with the genotyped SNP

Supplementary Table 3 – p-Values for Associations of SNPs with Histological Features in the NAFLD cohort[A]

SNP	Locus	Steatosis Grade			Inflammation			Ballooning			Mallory-Denk Bodies		
		Univariate[B]	Multivariate I[C]	Multivariate II[D]	Univariate	Multivariate I	Multivariate II	Univariate	Multivariate I	Multivariate II	Univariate	Multivariate I	Multivariate II
rs10157801	1q21	0.096	0.042*	0.037	0.025	0.021	0.015	NS			NS		
rs13030978	2q12	NS[E]			NS			0.021	0.009*	0.018	NS		
rs2140773	2q37	NS			NS			NS			0.076	0.082	0.18
rs10513686	3q26	NS			0.083	0.039	0.0024	0.080	0.096	0.12	NS		
rs6834314	4q22	0.019	0.013*	0.015	0.048	0.102	0.172	0.017	0.034*	0.043	0.007	0.009*	0.022
rs6888304	5p15	0.096	0.031*	0.023	NS			0.092	0.17	0.199	0.034	0.11	0.14
rs1126617	6p22	0.006	0.010*	0.013	NS			NS			0.064	0.15	0.21
rs17145750	7q11	0.031	0.027*	0.048	NS			NS			NS		
rs2954021	8q24	NS			NS			NS			0.072	0.029	0.055
rs6984305	8p23	0.005	0.007*	0.009	NS			0.002	0.015*	0.012	0.001	0.012*	0.008
rs174601	11q12	NS			NS			NS			0.056	0.16	0.12
rs7310409	12q24	0.069	0.18	0.113	0.041	0.144	0.199	NS			NS		
rs10131298	14q32	0.069	0.037*	0.018	0.055	0.27	0.319	NS			NS		
rs339969	15q21	0.014	0.008*	0.007	NS			NS			NS		
rs8038465	15q23	0.002	0.002*	0.002	NS			NS			NS		
rs281377	19q13	NS			NS			0.081	0.034*	0.038	NS		

[A]Results reported only for SNPs with univariate p-values <0.1. The SNPs rs1497406, rs12145922, rs1335645, rs1260326. rs16856332, rs1027841, rs4074793, rs4715438, rs10819937, rs579459, rs7082470, rs754466, rs2236653, rs7186908, rs4581712, rs314253, rs9913711, rs12968116, rs4503880, rs7267979, rs1076540, rs2739330 (see Supplementary Table 2) were genotyped but showed no significant association with NAFLD histology

[B]Jonkheere-Terpstra test

[C]Multivariate ordinal logistic regression adjusted for age, BMI, gender and diabetes

[D]Multivariate ordinal logistic regression adjusted for age, BMI, gender, diabetes and rs738409 genotype

[E]NS – Non-significant (p>0.1)

*FDR q-value < 0.05 (assessed for model I only)

Supplementary Table 4 –*P*-values for Association of SNPs with Liver Enzymes in the NAFLD Cohort

SNP	ALT	AST	AlkP	GGT
rs6834314	$5X10^{-5}$	$1.5X10^{-4}$	NS	0.014
rs17145750	NS	NS	0.008	NS
rs10819937	0.019	0.02	NS	NS
rs579459	NS	NS	$2X10^{-5}$	NS
rs174601	NS	NS	0.015	NS
rs7310409	NS	NS	NS	0.007
rs7186908	NS	NS	NS	0.036
rs314253	NS	NS	0.013	NS
rs4503880	NS	NS	NS	0.033

Linear regression of log-transformed enzyme activity with SNP, adjusted for age, gender, diabetes and BMI. NS – not significant

Supplementary Table 5 – *P*-Values for the Association of rs6834314 With Histological Features in the NAFLD Cohort Using Two Different Models

	Model 1[A]	Model 2[B]
Fat	0.013	N/A
Inflammation	0.102	0.045
Ballooning	0.034	0.009
Mallory-Denk Bodies	0.009	0.008
Fibrosis	0.31	0.13

[A]Model 1 - Multivariate ordinal logistic regression adjusted for age, BMI, diabetes and gender (see Supplementary Table 3)
[B]Model 2 - Multivariate ordinal logistic regression adjusted for age, BMI, diabetes, gender and degree of steatosis

Supplementary Table 6 – *P*-Values for the Association of rs6834314 With Histological Features in the NAFLD Cohort after Adjustment for Other Genetic Variants

	Model 1[A]	Model 2[B]
Fat	0.013	0.013
Inflammation	0.102	0.067
Ballooning	0.034	0.031
Mallory-Denk Bodies	0.009	0.030
Fibrosis	0.31	0.34

[A]Model 1 - Multivariate ordinal logistic regression adjusted for age, BMI, diabetes and gender (see Supplementary Table 3)
[B]Model 2 - Multivariate ordinal logistic regression adjusted for age, BMI, diabetes, gender, rs738409 (PNPLA3), rs58542926 (TM6SF2), rs626283 (MBOAT7), and rs1260326 (GCKR) genotypes

Supplementary Table 7 – HCV Cohort Characteristics

N	313[A]
Male gender	188 (60%)
Age	46.8±9.6
BMI	27.7±5.8
HCV Genotype	
1	233 (79.8%)
2	53 (18.2%
4	5 (1.7%)
5	1 (0.3%)
ALT (median, interquartile range)	73 (45-126)
Steatosis on Biopsy	
None	134 (42.8%)
Trace (<5%)	60 (19.2%)
5-25%	97 (31%)
26-50%	18 (5.8%)
>50%	4 (1.3%)
Inflammatory Score (HAI, median, IQR)	8 (6-10)
Cirrhosis	32 (10.2%)
rs6834314 Genotype	
AA	165 (52.7%)
AG	125 (40%)
GG	23 (7.4%)
Allele Frequency (A/G)	72.7%/27.3%

[A] rs6834314 genotyped successfully in 313 of 317 (99%) subjects

Supplementary Table 8 – Michigan Genomic Initiative (MGI) Cohort Characteristics

N	31,221
Male gender	14,631(46.9%)
European Ancestry	100%
Age	52.1±16.1
BMI	29.8±7.0
ALT (median, interquartile range)[A]	24 (18-32.8)
rs6834314 Genotype	
AA	16,506 (52.9%)
AG	12,326 (39.5%)
GG	2,389 (7.7%)
Allele Frequency (A/G)	72.6%/27.4%

[A] ALT available for 19,598 subjects

Supplementary Table 9 – UK Biobank Cohort Characteristics

N	408,961
Male gender	187,909 (46%)
Age	66.9±8
BMI	27.4±4.8
rs72613567 Genotype[A, B]	
-/-	211,662 (52%)
-/A	164,132 (40.3%)
A/A	31,476 (7.8%)
Allele Frequency (-/A)	72.1%/27.9%
rs62305723 Genotype[C]	
G/G	357,433 (87.6%)
G/A	48,827 (12%)
A/A	1,634 (0.4%)
Allele Frequency (G/A)	93.6%/6.4%

[A] rs6834314 not genotyped in this cohort
[B] rs72613567 genotyped successfully in 407,720 subjects (99.6%)
[B] rs62305723 genotyped successfully in 407,894 subjects (99.7%)

Supplementary Table 10 – *P*-Values from Conditional analysis for the Association of rs72613567 and rs62305723 With Histological Features in the NAFLD Cohort

Model		Fat	Inflammation	Ballooning	Mallory-Denk Bodies	Fibrosis
Model 1[A]	rs72613567	0.012	0.016	0.85	0.024	0.23
	rs62305723	0.51	$2X10^{-5}$	0.001	0.014	0.048
Model 2[B]	rs72613567	0.022[A]	0.03	0.73	0.044	0.36
	rs62305723	0.47	$3X10^{-5}$	0.001	0.027	0.10

[A] Model 1 - Multivariate ordinal regressions (for each histological feature separately), with both rs72613567 and rs62305723 genotypes included in the model, adjusted for age, gender, BMI and diabetes.
[A] Model 2 - Multivariate ordinal regressions (for each histological feature separately), with both rs72613567 and rs62305723 genotypes included in the model, adjusted for age, gender, BMI, diabetes and rs738409 (PNPLA3) genotype.

Supplementary Table 11 – *P*-Values from Conditional analysis for the Association of HSD17B13 SNPs rs72613567, rs62305723 and PNPLA3 SNP rs738409 With Histological Features in the NAFLD Cohort

	Fat	Inflammation	Ballooning	Mallory-Denk Bodies	Fibrosis
rs72613567	0.022[A]	0.03	0.73	0.044	0.36
rs62305723	0.47	$3X10^{-5}$	0.001	0.027	0.10
rs738409	0.083	$1X10^{-6}$	0.53	0.159	$2X10^{-5}$

[A] p-values obtained from multivariate ordinal regressions (for each histological feature separately), with both rs72613567, rs62305723 and rs738409 genotypes included in the model, adjusted for age, gender, BMI and diabetes.

Supplementary Table 12 – Primer Sequences

Name	Sequence
HSD17B13 (total) Forward	ATC ACA AAA GCA CTT CTT CCA TC
HSD17B13 (total) Reverse	AGA CCT CTG TGA AAG CCA AC
HSD17B13 (total) Probe	/56-FAM/TA CCT CAT C/Zen/C CAT ATT GTT CCA GCA AAT TTG /3IABkFQ/
HSD17B13-Delta22-28 Forward	CAGAGGAGAAAATCTGTGG
HSD17B13-Delta22-28 Reverse	CGACTCCAAGTAGGAGTAG
HSD17B13-Delta71-106 Forward	GTGAAGAAAGAAGTGGGTG
HSD17B13-Delta71-106 Reverse	CTTATTAATATCCCACAGAACC
HSD17B13-P260S Forward	GATTTTTGTTtCATCGTATATCAATATC
HSD17B13-P260S Reverse	ATTTTCTTATTGGTAAGTATTCCATC
HSD17B13-G47/G49A-Forward	AGcCAGGCAGACTACTTATGAATTTG
HSD17B13-G47/G49A-Reserse	ATTgCATGCCCAGCTCCAGTAA
HSD17B13-Delta-exon 6 Forward	GTTTCTTCCTGAACGCGC
HSD17B13-Delta-exon 6 Reverse	CTTGTGCTTGGATTTTTGGTG
HSD17B13-G insertion Forward	gGTTTCTTCCTGAACGCGC
HSD17B13-G insertion Reverse	TTCTGTAGTCTCAGAAAGATATTG
HSD17B13-G insertion step 2 Forward	GACTACAAAGACGATGACG
HSD17B13-G insertion step 2 Reverse	GGAAGAAACCTTCTGTAGTC
rs72613567 genotyping Forward	CATCTAATGCCTCGCCACCA
rs72613567 genotyping Reverse	CCCCAGGGATGGAGAGTTTC
Splice Variant Quantitation-with exon 6 Forward	TCAGAACTTCAGGCCTTGGG
Splice Variant Quantitation-with exon 6 Reverse	ACGACTTCATCTGTCTCCAATACA
Splice Variant Quantitation-Delta exon 6 Forward	AATTTGCCGCTGTTGGCTTT
Splice Variant Quantitation-Delta exon 6 Reverse	GCGTTCAGGAAGAAACCTTGT
Splice Variant Detection and Sequence Forward	AAATTTGCCGCTGTTGGCTTT
Splice Variant Detection and Sequence Reverse	AGCTGTGCACTCATTCTGTGT

SUPPLEMENTARY REFERENCES

1. Poynard T, Ratziu V, McHutchison J, Manns M, Goodman Z, Zeuzem S, Younossi Z, et al. Effect of treatment with peginterferon or interferon alfa-2b and ribavirin on steatosis in patients infected with hepatitis C. Hepatology 2003;38:75-85.
2. Rubbia-Brandt L, Quadri R, Abid K, Giostra E, Male PJ, Mentha G, Spahr L, et al. Hepatocyte steatosis is a cytopathic effect of hepatitis C virus genotype 3. J Hepatol 2000;33:106-115.
3. Rembeck K, Maglio C, Lagging M, Christensen PB, Farkkila M, Langeland N, Buhl MR, et al. PNPLA 3 I148M genetic variant associates with insulin resistance and baseline viral load in HCV genotype 2 but not in genotype 3 infection. BMC Med Genet 2012;13:82.
4. Knodell RG, Ishak KG, Black WC, Chen TS, Craig R, Kaplowitz N, Kiernan TW, et al. Formulation and application of a numerical scoring system for assessing histological activity in asymptomatic chronic active hepatitis. Hepatology 1981;1:431-435.
5. Ishak K, Baptista A, Bianchi L, Callea F, De Groote J, Gudat F, Denk H, et al. Histological grading and staging of chronic hepatitis. J Hepatol 1995;22:696-699.
6. Dey R, Schmidt EM, Abecasis GR, Lee S. A Fast and Accurate Algorithm to Test for Binary Phenotypes and Its Application to PheWAS. Am J Hum Genet 2017;101:37-49.
7. O'Connell J, Gurdasani D, Delaneau O, Pirastu N, Ulivi S, Cocca M, Traglia M, et al. A general approach for haplotype phasing across the full spectrum of relatedness. PLoS Genet 2014;10:e1004234.
8. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284-1287.
9. Wang C, Zhan X, Liang L, Abecasis GR, Lin X. Improved ancestry estimation for both genotyping and sequencing data using projection procrustes analysis and genotype imputation. Am J Hum Genet 2015;96:926-937.
10. Wang C, Zhan X, Bragg-Gresham J, Kang HM, Stambolian D, Chew EY, Branham KE, et al. Ancestry estimation and control of population stratification for sequence-based association studies. Nat Genet 2014;46:409-415.
11. Venables W, Ripley B. Modern Applied Statistics with S. Springer, New York, NY. 2002.
12. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinformatics 2010;26:2867-2873.
13. Feng S, Liu D, Zhan X, Wing MK, Abecasis GR. RAREMETAL: fast and powerful meta-analysis for rare variants. Bioinformatics 2014;30:2828-2829.
14. Pirinen M, Donnelly P, Spencer CC. Efficient computation with a linear mixed model on large-scale data sets with applications to genetic studies. The Annals of Applied Statistics 2013:369-390.
15. Allen N, Sudlow C, Downey P, Peakman T, Danesh J, Elliott P, Gallacher J, et al. UK Biobank: Current status and what it means for epidemiology. Health Policy and Technology 2012;1:123-126.

16. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, et al. Genome-wide genetic data on~ 500,000 UK Biobank participants. bioRxiv 2017:166298.
17. Zhou W, Nielsen JB, Fritsche LG, Dey R, Elvestad MB, Wolford BN, LeFaive J, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. bioRxiv 2018:212357.
18. Rotman Y, Koh C, Zmuda JM, Kleiner DE, Liang TJ. The association of genetic variability in patatin-like phospholipase domain-containing protein 3 (PNPLA3) with histological severity of nonalcoholic fatty liver disease. Hepatology 2010;52:894-903.
19. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 2015;31:3555-3557.
20. Ding Y, Zhang S, Yang L, Na H, Zhang P, Zhang H, Wang Y, et al. Isolating lipid droplets from multiple species. Nat Protoc 2013;8:43-51.
21. Fujimoto Y, Itabe H, Sakai J, Makita M, Noda J, Mori M, Higashi Y, et al. Identification of major proteins in the lipid droplet-enriched fraction isolated from the human hepatocyte cell line HuH7. Biochim Biophys Acta 2004;1644:47-59.
22. Belyaeva OV, Johnson MP, Kedishvili NY. Kinetic analysis of human enzyme RDH10 defines the characteristics of a physiologically relevant retinol dehydrogenase. J Biol Chem 2008;283:20299-20308.
23. Belyaeva OV, Korkina OV, Stetsenko AV, Kim T, Nelson PS, Kedishvili NY. Biochemical properties of purified human retinol dehydrogenase 12 (RDH12): catalytic efficiency toward retinoids and C9 aldehydes and effects of cellular retinol-binding protein type I (CRBPI) and cellular retinaldehyde-binding protein (CRALBP) on the oxidation and reduction of retinoids. Biochemistry 2005;44:7035-7047.
24. Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, Van der Harst P, Holm H, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. Nat Genet 2011;43:1131-1138.