Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease

Julia Kozlitina1,8, Eriks Smagris1,8, Stefan Stender2, Børge G Nordestgaard3–5, Heather H Zhou6, Anne Tybjærg-Hansen2,3,5, Thomas F Vogt6, Helen H Hobbs1,7 & Jonathan C Cohen1

Nonalcoholic fatty liver disease (NAFLD) is the most common form of liver disease. To elucidate the molecular basis of NAFLD, we performed an exome-wide association study of liver fat content. Three variants were associated with higher liver fat levels at the exome-wide significance level of 3.6 × 10−7: two in PNPLA3, an established locus for NAFLD, and one (encoding p.Glu167Lys) in TM6SF2, a gene of unknown function. The TM6SF2 variant encoding p.Glu167Lys was also associated with higher circulating levels of alanine transaminase, a marker of liver injury, and with lower levels of low-density lipoprotein–cholesterol (LDL-C), triglycerides and alkaline phosphatase in 3 independent populations (n > 80,000). When recombinant protein was expressed in cultured hepatocytes, 50% less Glu167Lys TM6SF2 protein was produced relative to wild-type TM6SF2. Adeno-associated virus–mediated short hairpin RNA knockdown of Tm6sf2 in mice increased liver triglyceride content by threefold and decreased very-low-density lipoprotein (VLDL) secretion by 50%. Taken together, these data indicate that TM6SF2 activity is required for normal VLDL secretion and that impaired TM6SF2 function causally contributes to NAFLD.

NAFLD comprises a spectrum of related disorders that stem from the aberrant accumulation of fat in the liver1. Approximately 30% of adults2 have excess liver fat (steatosis), which is stored in the form of triglycerides in cytoplasmic lipid droplets in hepatocytes. Simple steatosis is usually benign but can progress to a chronic inflammatory condition called nonalcoholic steatohepatitis (NASH) and ultimately to cirrhosis, in which the functional cells of the liver are replaced by fibrosis3. The propensity to accumulate hepatic triglycerides varies markedly among individuals2, but the factors underlying this variation have not been fully elucidated.

Previously, we found a missense variant (encoding p.Ile148Met) in patatin-like phospholipase domain–containing 3 (PNPLA3) that is strongly associated with hepatic triglyceride content (HTGC) and with serum levels of alanine transaminase (ALT)4. Subsequent genome-wide association studies (GWAS) found other common SNPs associated with liver fat content5 and levels of circulating liver enzymes6,7. To identify the functional variants at these loci, we used genotyping arrays (HumanExome BeadChip, Illumina) to perform an exome-wide association study in a multi-ancestry, population-based study, the Dallas Heart Study (DHS)8. A total of 138,374 sequence variants that were polymorphic and passed our quality control criteria were tested for association with HTGC in 2,736 DHS participants (1,324 non-Hispanic African Americans, 882 non-Hispanic, European Americans, 467 Hispanics and 63 other ancestry groups), with adjustment for age, sex, ancestry and body mass index (BMI) (Online Methods). Two sequence variants in PNPLA3 (rs738409 and rs2281135) had the lowest association P values (4.0 × 10−10 and 6.9 × 10−12, respectively), and the next most significant variant (rs58542926) was in TM6SF2 (P = 5.7 × 10−9) (Fig. 1a). No other variants exceeded the exome-wide significance threshold of 3.6 × 10−7. After excluding these three SNPs, the quantile-quantile plot of P values showed no systematic deviation from the expected null distribution (Fig. 1b). The TM6SF2 variant was not associated with other risk factors for hepatic steatosis, including BMI, homeostatic model assessment of insulin resistance (HOMA-IR) or alcohol intake (Supplementary Table 1).

The TM6SF2 variant associated with HTGC is an adenine-to-guanine substitution in coding nucleotide 499, which replaces glutamate at residue 167 with lysine (c.499A>G; p.Glu167Lys). Glu167 is highly conserved among mammals and is also an acidic residue (aspartate) in birds (Fig. 1c). The frequency of the TM6SF2 variant encoding p.Glu167Lys was higher in individuals of European ancestry (7.2%) than in African Americans (3.4%) or Hispanics (4.7%). Carriers of the TM6SF2 variant encoding p.Glu167Lys had elevated mean and median HTGC in all three ancestry groups, although the difference in triglyceride content relative to non-carriers did not reach statistical significance in Hispanics, likely owing to the lower number

1McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, Texas, USA. 2Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital and Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 3Copenhagen General Population Study, Herlev Hospital, Copenhagen University Hospital and Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 4Department of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital and Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 5Copenhagen City Heart Study, Frederiksberg Hospital, Copenhagen University Hospital and Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. 6Merck Research Laboratories, Kenilworth, New Jersey, USA. 7Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas, USA. 8These authors contributed equally to this work. Correspondence should be addressed to H.H.H. (helen.hobbs@utsouthwestern.edu) or J.C.C. (jonathan.cohen@utsouthwestern.edu).

Received 26 November 2013; accepted 27 January 2014; published online 16 February 2014; doi:10.1038/ng.2901
of Hispanic participants and the lower frequency of the variant in this group (Fig. 1d). The association remained significant after adjusting for ethanol intake and HOMA-IR \((P = 5.6 \times 10^{-7})\). The effect of the TM6SF2 variant encoding p.Glu167Lys on HTGC was independent of the effect mediated by the PNPLA3 rs738409 polymorphism; we found no evidence for statistical interaction between the two risk alleles \((P = 0.62)\).

Previously, SNPs near TM6SF2 were found to be associated with NAFLD\(^5\)\(^,\)\(^6\). The variant at the locus that was most strongly associated with HTGC in the largest GWAS\(^5\) (total of 2.4 million imputed or assayed SNPs) was in NCAN \((rs2228603)\). The TM6SF2 variant encoding p.Glu167Lys remained robustly associated with HTGC \((P = 1.3 \times 10^{-5})\) after conditioning on rs2228603, as well as on other SNPs from the region on the array (Supplementary Table 2). Conversely, conditioning on the TM6SF2 variant encoding p.Glu167Lys abolished the association between the NCAN variant \((rs2228603)\) and HTGC \((conditional P = 0.8)\).

RT-PCR analysis of cDNA prepared from a panel of human tissues indicated that TM6SF2 is most highly expressed in the small intestine, liver and kidney and is present at lower levels in other tissues (Fig. 1e).

The TM6SF2 variant encoding p.Glu167Lys was also associated with a significant increase in serum ALT activity, consistent with increased hepatic injury (Table 1 and Supplementary Table 3). This finding is similar to what was previously observed with the PNPLA3 variant encoding p.Ile148Met, which is strongly associated with both HTGC and elevated ALT activity\(^5\)\(^,\)\(^6\). To confirm the association with NAFLD, we performed association studies in two

**Table 1** Association between the nonsynonymous TM6SF2 variant \((p.Glu167Lys)\), liver enzymes and plasma lipid levels in DHS, the Dallas Biobank and the Copenhagen Study

| Study                  | Trait | \(EE^a\) | \(EK^a\) | \(KK^a\) | \(P\) value |
|------------------------|-------|---------|---------|---------|-------------|
| DHS                    |       |         |         |         |             |
| \(n\)                  | 4,151 | 423     | 13      | –       |
| ALT (U)                | 23.5 ± 19.9 | 25.8 ± 18.0 | 29.6 ± 26.8 | 0.014 |
| AST (U)                | 24.3 ± 20.6 | 25.1 ± 19.6 | 25.6 ± 16.5 | 0.17 |
| ALP (U)                | 71.6 ± 26.4 | 68.0 ± 20.9 | 63.6 ± 16.4 | 0.031 |
| LDL-C (mg/dl)          | 109 ± 36 | 105 ± 34 | 94 ± 39 | 0.005 |
| Triglycerides (mg/dl)  | 123 ± 102 | 118 ± 91 | 130 ± 66 | 0.037 |
| Dallas Biobank (Europeans Americans) | | | | | |
| \(n\)                  | 7,416 | 1,112 | 57 | – | |
| ALT (U)                | 35.9 ± 16.3 | 36.5 ± 15 | 44.8 ± 23.8 | 0.003 |
| AST (U)                | 27 ± 13.8 | 26.9 ± 10.7 | 30.2 ± 13.5 | 0.22 |
| ALP (U)                | 69 ± 18.9 | 66.9 ± 17.3 | 65.9 ± 19.1 | 1.2 \times 10^{-4} |
| LDL-C (mg/dl)          | 107 ± 34 | 105 ± 32 | 97 ± 32 | 0.029 |
| Triglycerides (mg/dl)  | 112 ± 79 | 105.3 ± 67 | 97.5 ± 54 | 6.7 \times 10^{-5} |
| Copenhagen Study       |       |         |         |         |             |
| \(n\)                  | 61,279 | 11,700 | 553 | – | |
| ALT (U)                | 22.7 ± 17.2 | 23.8 ± 16.0 | 26.7 ± 20.9 | 7.6 \times 10^{-14} |
| AST (U) \((n = 8,487)\) | 22.8 ± 13.2 | 23.9 ± 18.6 | 30.0 ± 36.6 | 1.8 \times 10^{-4} |
| ALP (U)                | 91.1 ± 37.5 | 89.7 ± 39.1 | 85.0 ± 34.2 | 4.3 \times 10^{-7} |
| LDL-C (mg/dl)          | 128 ± 39 | 127 ± 38 | 112 ± 36 | 4.7 \times 10^{-14} |
| Triglycerides (mg/dl)  | 152 ± 106 | 148 ± 96 | 133 ± 90 | 3.7 \times 10^{-9} |

Values are means ± s.d. Association was tested using linear regression with adjustment for age, sex, BMI and ancestry (where appropriate). A logarithm transformation was applied to traits with non-normal distributions. AST was only measured in a subset of the Copenhagen Study \((n = 8,487)\). Ancestry breakdown for DHS is provided in Supplementary Table 3. ALT, alanine aminotransferase; AST, aspartate transaminase; ALP, alkaline phosphatase; LDL-C, low-density lipoprotein–cholesterol.

\(^a\)EE, homozygotes for the allele encoding Glu167; EK, heterozygotes; KK, homozygotes for the allele encoding Lys167.
additional cohorts: the Dallas Biobank (n = 8,585 European Americans) and a cohort from Copenhagen (referred to as the Copenhagen Study in this manuscript) that included both the Copenhagen City Heart Study (CCHS) and the Copenhagen General Population Study (CGPS) (n = 73,532)11 (Supplementary Tables 1 and 5). As in DHS, the TM6SF2 variant encoding p.Glu167Lys was associated with significantly higher serum activity of ALT in both cohorts (Table 1). Mean serum aspartate transaminase (AST) activity was also higher in individuals homozygous for the TM6SF2 variant encoding p.Glu167Lys in both of these larger cohorts, but the increase only reached significance in the Copenhagen Study. These findings further support the hypothesis that the TM6SF2 variant encoding p.Glu167Lys is associated with NAFLD and are consistent with the notion that the variant compromises hepatic integrity. No genotype-specific differences were observed in plasma levels of bilirubin or γ-glutamyltransferase (Supplementary Tables 1, 4 and 5), but, in all three populations, the TM6SF2 variant encoding p.Glu167Lys was associated with a significant reduction in the serum activity of alkaline phosphatase (ALP).

The TM6SF2 variant encoding p.Glu167Lys was associated with highly significant reductions in plasma levels of triglycerides and LDL-C in the DHS, Dallas Biobank and Copenhagen Study cohorts (Table 1). Other SNPs in the region (rs10409196, rs16996148 and rs17216525)12–14 were associated with differences in plasma levels of triglycerides and LDL-C, but TM6SF2 remained associated with HTGC when we controlled for the genotypes at these loci. Conversely, controlling for the TM6SF2 variant encoding p.Glu167Lys abolished the signals for these SNPs (Supplementary Table 2). Plasma levels of high-density lipoprotein–cholesterol (HDL–C) were not associated with the TM6SF2 variant encoding p.Glu167Lys (Supplementary Table 1). Thus, the p.Glu167Lys substitution in TM6SF2 results in an increase in HTGC and a decrease in plasma levels of liver-derived triglyceride-rich lipoproteins.

The specific biological role of TM6SF2 is not known. The protein is predicted to have seven transmembrane domains (TMHMM 2.0)15, but it does not contain any known functional domains16. To assess the effect of the p.Glu167Lys substitution on the expression and localization of TM6SF2, we expressed human wild-type and mutant proteins in a human hepatoma cell line, HuH-7. Levels of TM6SF2 mRNA were comparable in cells expressing wild-type and mutant TM6SF2 (Fig. 2a, left), but levels of the mutant protein were reduced by 46%
(Fig. 2a, right). Similar results were obtained in cells solubilized in 3% SDS and 8 M urea. When lysates from Hepa1c1c7 cells expressing human TM6SF2-V5 were fractionated on sucrose density gradients, TM6SF2 was recovered exclusively in the membrane fraction (Fig. 2b). These findings suggest that TM6SF2 is a polytopic membrane protein and that the Glu167Lys variant form is misfolded and undergoes accelerated intracellular degradation.

To directly assess the effect of loss of TM6SF2 function on HTGC, we used recombinant adeno-associated viral (AAV) vectors expressing short hairpin RNAs (shRNAs) to selectively reduce TM6sf2 transcript levels in the livers of mice. Expression of two different shRNAs targeting mouse Tm6sf2 resulted in >90% knockdown of Tm6sf2 mRNA (Fig. 3a, left), without any changes in Tm6sf2 mRNA levels in adipose tissue or small intestine (Supplementary Fig. 1). Empty AAV and AAV expressing nonfunctional shRNA were used as controls (Supplementary Fig. 1b). Hepatic inhibition of Tm6sf2 increased HTGC by threefold (Fig. 3a, middle) and significantly decreased plasma levels of cholesterol in mice fed a chow diet ad libitum (Fig. 3a, right). Plasma triglyceride levels tended to be lower in knockdown mice fed ad libitum (Fig. 3a, right) and were consistently lower after a 4-h fast (Fig. 3c). Fast-performance liquid chromatography (PLPC) fractionation of plasma lipoproteins isolated after a 4-h fast showed that Tm6sf2 knockdown reduced the cholesterol content of both the LDL and HDL fractions and the triglyceride content of the VLDL fraction (Fig. 3b). Mean serum ALT levels were unchanged in knockdown mice compared to controls (mean ± s.d.: 136 ± 8 U (shRNA6) and 131 ± 4 U (shRNA8) versus 132 ± 8 U (control); P = 0.33 and P = 0.68, respectively).

Higher HTGC together with lower plasma cholesterol and triglyceride levels is consistent with a defect in VLDL secretion. To determine the effect of Tm6sf2 knockdown on VLDL secretion, we inhibited intravascular lipoprotein lipase, which hydrolyzes the triglyceride content of VLDL, and measured the rate of accumulation of triglycerides in plasma. Triglyceride accumulation was markedly reduced in the Tm6sf2 knockdown mice (mean ± s.d.: 4.5 ± 1.6 mg/dl/min (shRNA8) versus 11.1 ± 0.6 mg/dl/min (control); Fig. 3c). These data indicate that Tm6sf2 normally acts to promote VLDL secretion and suggest that the increased HTGC associated with the Glu167Lys TM6SF2 variant in humans results from a reduction in TM6SF2 function.

High-sucrose diets, which increase hepatic triglyceride synthesis, exacerbated the effects of Tm6sf2 knockdown on HTGC (Fig. 4). In sucrose-fed mice, levels of triglycerides and cholesterol esters were higher in the liver (Fig. 4a) and lower in plasma (Fig. 4b). Oil Red O staining, which stains neutral lipids, showed an increase in the number and size of lipid droplets in the knockdown mice (Fig. 4c). Therefore, knockdown of Tm6sf2 selectively in mouse liver recapitulated the effects on HTGC and plasma lipids of the Glu167Lys TM6SF2 variant observed in humans.

TM6SF2 is expressed at the highest levels in the human intestine (Fig. 1e), which is the source of dietary-derived triglyceride-rich lipoproteins (chylomicrons). ALP is also expressed in the intestine. Reduction in ALP activity in carriers of the TM6SF2 variant encoding p.Glu167Lys might be due to intestinal action of TM6SF2. Kallianian et al. found that deletion of intestinal ALP produces features of metabolic syndrome in mice17. We are currently testing the hypothesis that TM6SF2 has a role in lipoprotein synthesis and ALP activity in the intestine and that reduced intestinal ALP activity contributes to the increased HTGC associated with the TM6SF2 variant encoding p.Glu167Lys.

METHODS
Methods and any associated references are available in the online version of the paper.

Accession codes. TM6SF2 cDNA, NM_001001524; TM6SF2 protein, NP_001001524.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS
We thank F. Xu for excellent technical assistance. This work was supported by grants from the US National Institutes of Health (HL20948, HL109250 and DK090066) and the National Center for Advancing Translational Sciences (UL1TR001105). T.F.V acknowledges the contributions of G. Gao (University of Massachusetts) for advice and assistance with AAV vector production.

AUTHOR CONTRIBUTIONS
The manuscript was prepared by all of the authors. E.S. and H.H.Z. performed the experiments. J.K. and S.S. performed the genetic analysis and association studies. T.F.V, A.T.-H., B.G.N., J.C.C. and H.H.H. provided experimental design and coordination.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Tiniakos, D.G., Vos, M.B. & Brunt, E.M. Nonalcoholic fatty liver disease: pathology and pathogenesis. Annu. Rev. Pathol. 5, 145–171 (2010).
2. Browning, J.D. et al. Prevalence of hepatic steatosis in an urban population in the United States: impact of ethnicity. Hepatology 40, 1387–1395 (2004).
3. McCullough, A.J. The clinical features, diagnosis and natural history of nonalcoholic fatty liver disease. Clin. Liver Dis. 8, 521–533, viii (2004).
4. Romeo, S. et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat. Genet. 40, 1461–1465 (2008).
5. Speliotes, E.K. et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet.* **7**, e1001324 (2011).

6. Yuan, X. et al. Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes. *Am. J. Hum. Genet.* **83**, 520–528 (2008).

7. Chambers, J.C. et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat. Genet.* **43**, 1131–1138 (2011).

8. Victor, R.G. et al. The Dallas Heart Study: a population-based probability sample for the multidisciplinary study of ethnic differences in cardiovascular health. *Am. J. Cardiol.* **93**, 1473–1480 (2004).

9. Gorden, A. et al. Genetic variation at NCAN locus is associated with inflammation and fibrosis in non-alcoholic fatty liver disease in morbid obesity. *Hum. Hered.* **75**, 34–43 (2013).

10. Hernaez, R. et al. Association between variants in or near PNPLA3, GCKR, and PPP1R3B with ultrasound-defined steatosis based on data from the Third National Health and Nutrition Examination Survey. *Clin. Gastroenterol. Hepatol.* **11**, 1183–1190 e2 (2013).

11. Stender, S., Frikke-Schmidt, R., Nordestgaard, B.G. & Tybjaerg-Hansen, A. Extreme bilirubin levels as a causal risk factor for symptomatic gallstone disease. *JAMA Intern. Med.* **173**, 1222–1228 (2013).

12. Teslovich, T.M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).

13. Kathiresan, S. et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat. Genet.* **40**, 189–197 (2008).

14. Kathiresan, S. et al. Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.* **41**, 56–65 (2009).

15. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).

16. Marchler-Bauer, A. et al. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* **39**, D225–D229 (2011).

17. Kaliannan, K. et al. Intestinal alkaline phosphatase prevents metabolic syndrome in mice. *Proc. Natl. Acad. Sci. USA* **110**, 7003–7008 (2013).
ONLINE METHODS

Study populations. The Dallas Heart Study (DHS) is a multi-ancestry, population-based probability sample of Dallas County residents, weighted to include 50% African-American and 50% non-African-American participants3. Sampling design and recruitment procedures were previously described in detail11. The study was initiated in 2000 and converted from a cross-sectional to a longitudinal study in 2007, when all participants were invited for a repeat evaluation (DHS-2). The sampling process was designed to obtain a probability sample of residents that was representative of the target population and to identify subjects to be recruited later for mechanistic substudies of hypertension and dyslipidemia. On the basis of power calculations for several such studies, the recruitment target was set at 3,000 subjects completing all phases of initial data collection. The DHS-2 cohort was augmented by voluntary participation of friends and spouses of the original participants. DHS was approved by the Institutional Review Board (IRB) of the University of Texas Southwestern Medical Center, and all subjects provided written informed consent. Each participant completed a detailed staff-administered survey, including questions about socioeconomic status, medical history and medication use. Ancestry was self-reported. A total of 4,708 DHS participants (52% non-Hispanic, African Americans, 29% non-Hispanic whites, 16% Hispanics and 3% other ancestries) who underwent clinical examination during DHS-1 and 3% other ancestries) who underwent clinical examination during DHS-2 were included in the present analysis.

Findings in DHS were replicated in two independent cohorts: the Dallas Biobank and the Copenhagen Study. The Dallas Biobank is a repository of DNA and plasma samples from individuals ascertained at various locations in north-central Texas. All participants were over 18 years of age and gave written informed consent for inclusion in the database. Individuals for this study were European Americans who completed a preventive medicine examination at the Cooper Clinic in Dallas between 2008 and 2012 after signing an informed consent that was approved by the Cooper Research Institute IRB. The Copenhagen Study combines participants in two studies of the Danish population: the Copenhagen General Population Study (CGPS)12 and the Copenhagen City Heart Study (CCHS)12.CGPS and CCHS are prospective studies of the Danish general population initiated, respectively, in 2003 and 1976–1978, with ongoing enrollment. Individuals were selected on the basis of the National Danish Civil Registration System to reflect the adult Danish population aged 20 to 100+ years of age. Data were obtained from a self-administered questionnaire reviewed together with an investigator on the day of attendance, a physical examination and blood samples including DNA extraction. Blood samples for DNA extraction and biochemical analyses were done on the day of enrollment in CGPS (2003–2011) and on the day of enrollment at the CCHS examinations in 1991–1994 and 2001–2003. Studies were approved by IRBs and Danish ethical committees and were conducted according to the Declaration of Helsinki. Written informed consent was obtained from participants. All participants were white and of Danish descent, as determined by the National Danish Person Registration System. There was no overlap of individuals between the studies. We included 64,000 consecutive participants from CGPS and 9,532 participants from CCHS, yielding a total of 73,532 participants.

Clinical measurements. BMI was calculated as weight measured in kilograms divided by squared height in meters. HOMA-IR was calculated from fasting glucose and insulin. Plasma levels of HDL-C and triglycerides were determined by β quantification, and LDL-C concentrations were estimated using the Friedewald equation14. Liver enzyme levels (ALT, AST and ALP) were determined by enzymatic assays. Hepatic triglyceride content was measured with proton magnetic resonance spectroscopy (1H-MRS) as previously described15. Hepatic triglyceride measurements were available for a total of 2,815 participants (including 2,287 of the DHS-1 and 528 of the DHS-2 participants).

Genotyping and quality control. Genomic DNA was extracted from circulating leukocytes. A total of 4,625 DHS-1 and DHS-2 participants were genotyped using the Illumina Infinium HumanExome BeadChip, which captured a total of 247,870 markers, including functional exonic variants (>90%), disease-associated tag markers from recently published GWAS, ancestry-informative markers and other markers. Genotypes were called using Illumina GenomeStudio software. Several quality control filters were applied to samples and variants before analysis. Individuals were excluded for the following reasons: a call rate of <99% (n = 25) or duplicate discordance (n = 1). Variants were excluded on the basis of a call rate of <99% (n = 1,795) or a deviation from Hardy-Weinberg equilibrium in African Americans with P < 0.0001 (n = 221). In addition, variants that were monomorphic in our study population (n = 61,874) or had a single heterozygote (n = 27,705) were removed from analysis. We did not further filter variants on the basis of minor allele frequency. After exclusions, a total of 156,202 variants and 4,591 individuals were available for analysis (including 138,374 variants in 2,735 individuals with measures of HTGC). Genotyping of the Dallas Biobank Sample and the Copenhagen Study was performed using a TaqMan assay (dbSNP rs58542926; Applied Biosystems, C_89463510_10). Genotyping of the Copenhagen Study was performed by TaqMan assay using primers 1 and 2 (Supplementary Table 6).

Association analysis. To account for possible population stratification, we computed principal components of ancestry on the basis of markers with a minor allele frequency of >0.1% in the combined sample using EIGENSTRAT20 software version 4.2. Each sequence variant was tested for association with HTGC using linear regression, assuming an additive genetic model (with genotypes coded as 0, 1 or 2) and adjusted for age, sex, BMI and the four leading principal components of ancestry. For very rare variants (<8 carriers), the analysis was performed using the Wilcoxon rank-sum test, without covariate adjustment. A power transformation (1/4) was applied to liver fat measurements to achieve approximate normality of the residuals. On the basis of the number of tests performed, we set a significance threshold at 1 × 10−7 for each variant to maintain a family-wise error rate of 5%. A quantile-quantile plot of −log10 P values did not show systematic deviation from the expected distribution.

In the follow-up and replication analyses, association between the TM6SF2 variant encoding p.Glu167Lys and clinical traits was tested using linear regression with adjustment for age, sex, BMI and ancestry (where appropriate). For hepatic fat content, we tested an additional model adjusted for HOMA-IR and alcohol consumption. For variables with skewed distributions, a logarithm (BMI, ALT, AST, ALP, triglycerides) or square-root (LDL) transformation was applied before analysis to ensure that the residuals were approximately normal and had constant variance. All reported P values are two-sided.

Mouse studies. All experimental protocols were approved by the University of Texas Southwest Medical Center Institutional Animal Care and Research Committee. Mice were housed in a vivarium and maintained under a standard 12-h light/12-h dark cycle and fed a Chow diet (Teklad Mouse/Rat Diet 7001, Harlan Teklad) ad libitum unless otherwise stated. Experiments were performed in 8-week-old male C57BL/6J mice (Jackson Laboratory). For AAV experiments, half of the mice were selected at random and injected via the tail vein with 1 × 1011 genome copies of AAV8 or TM6sf2 shRNA, respectively. Initial studies were performed in eight mice per group to provide 80% power to detect 1.5 s.d. in liver triglyceride content. The investigator was not blinded to group allocation during the experiment. After 14 d, mice were fasted for 4 h at the end of the dark cycle and then sacrificed. Tissues and plasma were collected and frozen at −80 °C before analysis. Hepatic and plasma lipids were measured as described below. The experiment was repeated except that the mice were fed a high-sucrose diet (MP Biomedicals, 901683) for 4 weeks. In each experiment, data from all mice in the experiment were included in the analysis.

Quantitative RT-PCR assays of mRNA abundance. Total RNA was isolated from the livers of mice using RNA STAT-60 (Tel-Test), and RT-PCR measurements were performed. Expression levels of mouse Tm6sf2 mRNA were measured using oppositely oriented primers (primers 3 and 4; Supplementary Table 6). Mouse RPLP0 mRNA was used as an internal control. Oligonucleotides specific for each gene that was analyzed were used for PCR amplification in 2× SYBR Green PCR Master Mix (Applied Biosystems) in a volume of 66 μl, according to the manufacturer’s instructions.

Relative expression levels of TM6SF2 in human tissues were assayed in cDNA samples prepared from the RNA of 22 human tissues (Human Total RNA Master Panel II, Clontech) by RT-PCR using two oppositely oriented primers (primers 5 and 6; Supplementary Table 6). Human RPLP0 mRNA was used as an internal control.
Plasma fast-performance liquid chromatography. Eight-week-old C57BL/6J male mice (Jackson Laboratory) were injected via the tail vein with 1 × 10^{11} genome copies of AAV8 control or Tm6sf2 shRNA, respectively (four mice per group) and fed with regular chow for 4 d. After a 4-h fast, plasma was isolated from blood samples collected from the tail veins of four mice. Plasma samples were pooled (total volume of 400 µl per group), and lipoproteins were separated by FPLC using a Superose 6 column (GE Healthcare). A total of 42 fractions (300 µl each) were collected. The cholesterol and triglyceride content of each fraction was measured using enzymatic assays (Infinity, Thermo Scientific).

Liver and plasma chemistries. Lipids were extracted from liver tissue (−100 mg) using the method of Folch and Lees. Hepatic triglyceride, cholesterol and phosphatidylcholine concentrations were measured using enzymatic assays (Infinity, Thermo Electron and Wako) and normalized by sample weight. Serum levels of ALT, AST, ALP, triglycerides and cholesterol were measured using the Vitros 250 system (GM).

Histological studies. Mice were fed with a high-sucrose diet for 4 weeks after Tm6sf2 knockdown by AAV-encoding shRNA. The slice of liver was fixed in 4% paraformaldehyde for 48 h and then equilibrated in 10% sucrose for 24 h and in 18% sucrose for 24 h at 4 °C before cryosectioning by the Molecular Pathology Core facility of the University of Texas Southwestern Medical Center. Cryosections of livers stored at −80 °C were brought to room temperature and air-dried for at least 2 h and were then fixed in methanol-free 4% paraformaldehyde. Slides were washed with distilled water three times and incubated for 10 min in 0.18% Oil Red O (Sigma) prepared in 60% isopropanol alcohol. Slides were washed in distilled water five times. Nuclei were counterstained with hematoxylin, and coverslips were affixed with aequous mounting medium (Vector Laboratories). Slides were viewed using a Leica microscope (DM2000; original magnifications, ×20 and ×63).

Expression of TM6SF2 in cultured hepatocytes. The cDNA for human TM6SF2 was cloned downstream of the cytomegalovirus promoter/enhancer elements in pCMV6-XL5 plasmid ( OriGene Technologies). Single-nucleotide changes were introduced using the QuikChange site-directed mutagenesis kit (Stratagene) and confirmed by Sanger sequencing. A sequence encoding a C-terminal V5 epitope tag (GKPIPNPLLGLDST) was placed in each plasmid construct.

Human hepatoma (HuH-7) cells, which tested negative for mycoplasma, were obtained from Sigma and were transfected with plasmids expressing wild-type or mutant human TM6SF2-V5 using Lipofectamine 2000 (Life Technologies). Proteins were visualized using Super-Signal ECL (Pierce Biotechnology).

VLDL synthesis. To measure the effect of Tm6sf2 knockdown on VLDL secretion, 8-week-old C57BL/6J male mice (four per group) were transduced with 1 × 10^{11} genome copies of control AAV or AAV-shRNA (eight mice per group). Mice were maintained on a 12-h light/12-h dark cycle and fed the Teklad Mouse/Rat Diet 7001 chow diet (Harlan Teklad) (eight mice per group). After 14 d, mice were fasted for 4 h at the end of the dark cycle and then sacrificed.

Liver and plasma chemistries. Lipids were extracted from liver tissue (−100 mg) using the method of Folch and Lees. Hepatic triglyceride, cholesterol and phosphatidylcholine concentrations were measured using enzymatic assays (Infinity, Thermo Electron and Wako) and normalized by sample weight. Serum levels of ALT, AST, ALP, triglycerides and cholesterol were measured using the Vitros 250 system (GM).