Adiposity amplifies the genetic risk of fatty liver disease conferred by multiple loci

Stefan Stender1–3, Julia Kozlitina2, Børge G Nordestgaard4–6, Anne Tybjærg-Hansen3,5,6, Helen H Hobbs1,2,7 & Jonathan C Cohen8

Complex traits arise from the interplay between genetic and environmental factors. The actions of these factors usually appear to be additive, and few compelling examples of gene–environment synergy have been documented. Here we show that adiposity significantly amplifies the effect of three sequence variants (encoding PNPLA3 p.I148M, TM6SF2 p.E167K, and GCKR p.P446L) associated with nonalcoholic fatty liver disease (NAFLD). Synergy between adiposity and genotype promoted the full spectrum of NAFLD, from steatosis to hepatic inflammation to cirrhosis. We found no evidence of strong interaction between adiposity and sequence variants influencing other adiposity-associated traits. These results indicate that adiposity augments genetic risk of NAFLD at multiple loci that confer susceptibility to hepatic steatosis through diverse metabolic mechanisms.

For most complex traits, sequence variations identified by genome-wide association studies (GWAS) account for only a minor fraction of the heritable variation estimated from family studies1. The missing heritability has been attributed to rare variants that are not represented on commercial SNP arrays2–3, common variants that do not reach genome-wide significance4, and gene–gene and gene–environment interactions that amplify the phenotypic effects of individual sequence variations5–7. The contribution of gene–environment interactions remains controversial. Genetic variants are usually assumed to act in an additive manner5,8 such that the combined effect of two or more sequence variations equals the sum of their individual effects. Compelling examples of synergistic or context-dependent relationships between genetic variants and environmental exposures have been described, including susceptibility to adverse drug reactions9, infectious diseases10, and sun exposure11. But most reports of gene–environment interactions have proved poorly reproducible8,12,13.

Obesity has emerged as a major cause of morbidity owing to its role in metabolic disorders such as type 2 diabetes mellitus, hypertension, and dyslipidemia. More recently, obesity has been associated with NAFLD, a spectrum of disorders that includes excess liver fat (steatosis), inflammation (steatohepatitis), fibrosis (cirrhosis), and malignant transformation (hepatocellular carcinoma)14. Susceptibility to NAFLD is highly variable; not all individuals who are obese develop steatosis, and most cases of steatosis do not progress to chronic liver disease. Expression of the disorder is strongly influenced by heritable factors. One of the most powerful genetic risk factors for NAFLD is a SNP (rs738409, referred to here as the M variant) in PNPLA3 that changes residue 148 of patatin-like phospholipase 3 (PNPLA3) from isoleucine to methionine (p.I148M)15. Here we show that adiposity influences the effect of the M variant on hepatic triglyceride content (HTGC), as well as on serum alanine aminotransferase (ALT), a marker of hepatocellular injury, and on cirrhosis. Interactions with obesity were also observed for sequence variants in two other genes (TM6SF2 and GCKR) that contribute to NAFLD by different mechanisms16–18. We did not observe interactions of similar magnitude for a range of other traits that associate strongly with adiposity. Thus, gene–adiposity interaction on NAFLD appears to be a specific and robust phenomenon.

RESULTS

Effect of the M variant on HTGC is dependent on adiposity

Previously, we showed that steatosis (HTGC > 5.5%) was present in 33% of the participants in the Dallas Heart Study (DHS) and that body mass index (BMI) was strongly associated with increased HTGC (Spearman’s ρ = 0.4)19. To determine whether the effect of the M variant on HTGC is modified by adiposity, we analyzed the relationship between PNPLA3 genotype (wild-type (II), heterozygous (IM), or homozygous (MM) for the M variant) and HTGC after stratifying the DHS participants into four groups on the basis of BMI (Fig. 1). In the lean (BMI < 25 kg/m²) group, median HTGC increased modestly but significantly in a stepwise fashion in the II, IM, and MM groups (1.8%, 2.3%, and 2.8%, respectively; $P = 0.0003$). Steatosis was less common...
in the Genetics of Liver Disease study were associated with HTGC
that the M variant was amplified by increasing adiposity (Pinteraction = 4 × 10−5)
the Ninety-Fifth percentile of HTGC in the general population. Bottom, numbers (n) of
interaction remained significant (Pinteraction = 0.006) (TM6SF2). The dashed line marks the
interactions were a result of heteroscedasticity, we repeated the
interaction with the scale of the response variable. The BMI–
was threefold higher in MM than in II individuals (14.2% versus
The dashed line marks the ninety-fifth percentile of HTGC in the general population. Bottom, numbers (n) of individuals with each genotype in each BMI bin. Genotypes are described by their encoded amino acids (PP, LL, EE, and KK represent homozygotes; PL and EK indicate heterozygotes).

**Effect of M variant–adiposity interaction on liver enzymes**
Hepatic steatosis per se is considered to be benign23. A subset of individuals with steatosis develop hepatic inflammation, which can result in elevated serum levels of liver enzymes, especially ALT24. To determine whether adiposity exacerbates the effect of the M variant on liver inflammation, we tested for interaction between PNPLA3 genotype and BMI on serum ALT levels (Fig. 3). In the DHS, serum ALT levels were increased by the M variant, as previously reported15. Median ALT was 18 U/L in II homozygotes, 20 U/L in IM heterozygotes, and 22 U/L in MM homozygotes (Pinteraction = 9 × 10−5). As with HTGC, the effect of the M variant on ALT increased with increasing BMI. The M variant was associated with increased ALT in the obese (BMI 30–35 kg/m²) and very obese (BMI > 35 kg/m²) groups but not in the lean (BMI < 25 kg/m²) or overweight (BMI 25–30 kg/m²) groups. In the Danish Biobank and the Copenhagen cohorts (Online Methods), which are larger than the DHS, the effect of the M variant on ALT was also apparent in the overweight group. In all three cohorts, the interaction between BMI and PNPLA3 genotype on ALT was statistically significant (Fig. 3).

As for HTGC, we retested the BMI–PNPLA3 interaction in the DHS after various transformations of ALT. The interaction was robust regardless of transformation applied (Supplementary Fig. 2).
We also tested for interaction between BMI and the risk alleles at TM6SF2 and GCKR on ALT levels in the DHS, Dallas Biobank, and Copenhagen cohorts. The effect of the TM6SF2 p.E167K variant on ALT was significantly affected by BMI in the Copenhagen cohort ($P = 10^{-4}$, $t$-test) but not in the DHS ($P = 0.14$) or Dallas Biobank ($P = 0.39$) groups (Supplementary Fig. 3). The GCKR p.P446L variant showed marginal evidence for interaction with BMI on ALT in the DHS ($P = 0.01$) but not the other cohorts (Supplementary Fig. 4). Thus, in contrast to our observations with the M variant, we did not find reproducible evidence of an interaction between risk alleles at TM6SF2 or GCKR and BMI on ALT levels. This may be a result of reduced power to detect interactions due to reduced effect size on HTGC (as with GCKR) or low allele frequency (as with TM6SF2).

**Effect of adiposity and M variant on cirrhosis prevalence**

The DHS and Dallas Biobank contain too few subjects with cirrhosis to examine gene–environment interactions for this phenotype. The Copenhagen cohort includes 384 participants with cirrhosis due either to alcoholism or to NAFLD (Supplementary Table 2). The effect of the PNPLA3 M variant on the prevalence of cirrhosis increased with increasing BMI (Fig. 4). The risk of cirrhosis was higher among MM homozygotes in each BMI category, including the lean group (BMI < 25 kg/m²). Among people with BMI > 35 kg/m², the odds ratio (OR) for cirrhosis was 5.8 in MM homozygotes versus II homozygotes. The corresponding OR in those with BMI < 25 kg/m² was 2.4. Additional adjustments for alcohol–BMI or alcohol–PNPLA3 interactions (individually or simultaneously) did not materially change results (data not shown). Thus, interaction between adiposity and the PNPLA3 M variant appears to promote chronic liver disease as well as steatosis.

**Mendelian randomization implicates adiposity in NAFLD**

If adiposity contributes to NAFLD, then SNPs associated with BMI would be expected to associate with HTGC as well. Because the individual effects on BMI of these SNPs are small, we constructed a genetic risk score using 30 SNPs that were associated with adiposity in a previous GWAS (Supplementary Table 3) and tested for association with BMI and HTGC in the DHS. As expected, an increasing genetic risk score was associated with a modest but significant increase in BMI ($P$ for trend across SNP score = 0.001) (Supplementary Fig. 5). Subjects in the first quintile had a median BMI of 27.5, whereas those in the fifth quintile of the risk score had a median BMI of 29.3 kg/m². BMI risk score was also associated with an increase in HTGC ($P = 0.02$, $t$-test). The modest effect of the obesogenic risk score on HTGC was consistent with the small increase in BMI associated with these variants.

**Gene–adiposity effect on other BMI-associated traits**

To assess whether gene–environment interactions were commonly observed with genetic predictors of other metabolic traits related to obesity, we screened the DHS database for phenotypes that showed a correlation with BMI with an absolute $r$ value >0.2 after adjusting for age, gender, and ethnicity (Supplementary Table 4). We then determined whether these traits were associated with any SNP assayed using the Illumina Exome BeadChip array at an exome-wide significance level ($P < 3.6 \times 10^{-7}$). For comparison, we included SNPs found to associate with NAFLD in previous studies. A total of 13 traits and 21 SNPs meeting these criteria were identified (Supplementary Table 5). BMI was strongly associated with plasma levels of leptin, for example (partial $r = 0.74$; $P < 1 \times 10^{-300}$). A SNP located in the leptin gene ($LEP$) was strongly associated with leptin levels (per-allele change in standardized leptin level = $-0.22$ s.d.; $P = 2.28 \times 10^{-11}$). Despite the strengths of the SNP–leptin and BMI–leptin associations, no SNP–BMI interaction was seen ($P = 0.60$, $t$-test). A similar lack of SNP–BMI interaction was observed for an SNP in $CRP$, the second most strongly BMI-correlated trait. Of the 13 traits examined, HTGC showed robust and highly statistically significant interactions between trait-associated SNPs and BMI ($P = 0.006$ to $3.7 \times 10^{-7}$, $t$-test).

For the other 12 BMI-associated traits, 3 nominally significant interactions were identified (between BMI and SNPs in $APOA5$, $LPL$, and $GCKR$ on plasma triglyceride (TG) levels). None of these interactions were observed in the Dallas Biobank, and only one was replicated ($P = 0.02$) in the Copenhagen cohort (Supplementary Table 6).
The major finding of this paper is that adiposity amplifies the genetic risk of NAFLD. In a cohort from the general population (the DHS), the prevalence of hepatic steatosis ranged from 9% in lean individuals who did not carry the PNPLA3 M variant to 84% in very obese individuals who were homozygous for the M variant. Adiposity also amplified the effects of the M variant on serum ALT activity and the risk of cirrhosis. Taken together, these results indicate that gene–adiposity interaction has a major role in the development and progression of NAFLD in humans. Other traits that are strongly correlated with adiposity (for example, plasma leptin and CRP levels) were not influenced by gene–BMI interactions in our study, despite having associations with genetic variants that were comparable in magnitude to that of the M variant on HTGC.

The interaction with adiposity was not specific to the M variant of PNPLA3. We found similar gene–adiposity relationships for steatogenic alleles of GCKR (encoding p.P446L) and TM6SF2 (encoding p.E167K). These three variants promote steatosis by distinct metabolic mechanisms. The M variant of PNPLA3 accumulates on cytoplasmic lipid droplets and probably compromises TG mobilization. GCKR is a negative regulator of glucokinase; p.P446L is a loss-of-function variant that results in increased phosphorylation of glucose, glycolysis, and fatty acid synthesis in the liver. TM6SF2 is a polytopic endoplasmic reticulum protein that is required for secretion of very low-density lipoprotein from the liver. The E167K substitution is a loss-of-function variant that results in impaired hepatic TG secretion and accumulation of hepatic fat. Thus, obesity may augment genetic risk of NAFLD through at least three metabolic mechanisms.

It is possible that obesity amplifies the effects of the three risk alleles by altering their expression. PNPLA3 is a direct target of the insulin-regulated transcription factor sterol regulatory element binding protein-1c (SREBP-1c) and is regulated by fasting and refeeding. GCKR expression is also increased by glucose and insulin. The insulin resistance associated with obesity may therefore increase expression of these two genes. However, TM6SF2 does not respond to food intake. Thus, the gene–adiposity interaction appears not to be due simply to enhanced expression of the risk allele at these three loci.

Do the variants in PNPLA3, TM6SF2, and GCKR interact with other environmental risk factors associated with obesity? Higher visceral fat content augmented the effect of the PNPLA3 M variant on hepatic fat content in a study of 2,257 nondiabetic European Americans. A diet rich in carbohydrates is associated with an increase in risk of NAFLD. Among 158 Hispanic children, high carbohydrate intake increased HTGC in those homozygous for the PNPLA3 M variant but not in IM heterozygotes or II homozygotes. We observed a similar phenomenon in mice: knock-in mice expressing the PNPLA3 M variant do not develop hepatic steatosis on a low-fat chow diet but show 2- to 3-fold higher HTGC (comparing wild type) on a high-sucrose diet, which dramatically increases levels of insulin. These findings support the hypothesis that gene–diet interactions have a role in NAFLD. We speculate that energy surplus is an absolute requirement for the deposition of fat in the liver. In the absence of an energy surplus, there is no driver for hepatic fat accumulation, irrespective of genotype. This situation is analogous to a pharmacogenetics interaction, in which the effect of a genetic variant is contingent on the action of a drug.

An estimated 30% of individuals who develop steatosis have associated hepatic inflammation. In a subset of these individuals, liver enzymes are released into the blood. The PNPLA3 M variant was the most strongly associated with serum ALT levels in the first GWAS on serum liver enzyme levels, and this result has been confirmed in subsequent studies. We found that adiposity amplified the effect of the M variant on ALT levels in a manner that was similar to the effect on steatosis. This finding contrasts with those of Larrieta-Carrasco et al., who reported that the OR of elevated ALT associated with the M variant was greater in normal-weight children than in obese children, and Giudice et al., who reported that the effect of the M variant increased with waist-to-hip ratio, but not with BMI, in obese children. The reasons for these discrepancies are not known.

We also found a gene–adiposity interaction when we analyzed the effect of PNPLA3 p.I148M on cirrhosis due to NAFLD or alcoholic liver disease. In the very obese group, MM homozygotes had a 5.8-fold increased risk of cirrhosis compared to II homozygotes. Among lean people (BMI < 25 kg/m²), MM homozygotes had a 2.4-fold higher risk of cirrhosis than II homozygotes. Thus, adiposity appears to amplify the effect of the PNPLA3 M variant on the entire spectrum of NAFLD, from steatosis to steatohepatitis to end-stage liver disease. A limitation to the observed interaction on cirrhosis is that the number of cases was small and that cirrhosis was defined by registry-based International Statistical Classification of Diseases and Related Health Problems (ICD) codes. The interaction’s effect on cirrhosis should therefore be viewed as preliminary, pending independent replication in larger patient cohorts.

Adiposity has been found to amplify the effect of alcohol consumption on liver disease. Among obese men, those who drink >15 units of alcohol per week have an 18.9-fold higher risk of death from liver disease than nondrinkers. The corresponding risk increase among lean men is 3.2. Taken together, our data and the results of Hart et al. indicate that adiposity exacerbates the effects of both genetic and nongenetic factors on fatty liver disease. It is possible that adiposity exacerbates alcoholic liver disease through its actions on PNPLA3, as PNPLA3 p.I148M has been shown to confer risk of cirrhosis among alcoholics.
Whereas the burgeoning of obesity in the population is a result of changes in lifestyle factors (presumably diet and exercise), interindividual differences in adiposity are also partly heritable. We show here that genetic variants that associate with increased BMI also associate with increased hepatic fat content. This finding indicates that the sequence variants at nearly 100 loci that have been associated with BMI would be predicted to be associated with liver fat content in a PNPLA3, TM6SF2, and GCKR genotype-dependent manner. Thus, the heritability of HTGC is determined not only by the primary effect of PNPLA3, TM6SF2, and GCKR genotype but also by the secondary effects of variation at nearly 100 loci that influence HTGC indirectly via their effects on adiposity. The interaction between BMI and NAFLD risk variants reported here should, therefore, be viewed as a mixture of gene–gene and gene–environment interactions.

Obesity increases susceptibility to a wide variety of common complex diseases ranging from cancer (for example, breast and colon cancer) and hypertension to metabolic disorders (for example, type 2 diabetes mellitus). Few single gene–adiposity interactions have been robustly documented for any of these conditions. In an effort to probe the contribution of gene–adiposity interactions to adiposity-associated traits more generally, we screened the DHS database for interaction with BMI (variants in PNPLA3, TM6SF2, and GCKR interacted with BMI in their effect on plasma TG levels). The interactions were not robustly replicated in two larger cohorts. Limitations of this screen include the modest sample size and the comparison of candidate gene SNPs with those identified by an agnostic exome-wide approach. Nevertheless, these findings support the hypothesis that gene–adiposity interactions of comparable magnitude to those observed for PNPLA3, TM6SF2, and GCKR on HTGC are uncommon.

What distinguishes NAFLD from other adiposity-associated phenotypes such as hypertension and blood glucose levels? First, the common alleles that contribute to hypertension and blood glucose levels all have smaller phenotypic effects than do the fatty liver susceptibility alleles. For example, homozygotes for the PNPLA3 M variant have twofold higher HTGC than do II homozygotes, and the variant explains ~5–10% of the variance in HTGC in different ethnicities. In contrast, the alleles most robustly associated with blood pressure or blood glucose only increase these traits by ~1%, and each explains less than 1% of the total trait variance. These modest effect sizes limit the power to detect interactions with adiposity.

A second major difference between blood pressure or blood glucose and liver fat content is that blood pressure and blood glucose are both under homeostatic control. Consequently, the effect of any genetic variant on blood glucose level or blood pressure will be opposed by counter-regulatory effects. In contrast, there is no evidence that the concentration of TG in the liver is subject to feedback regulation. Therefore, sequence variants or environmental factors (such as increased food consumption) can promote the accumulation of large amounts of TG within lipid droplets in the liver without eliciting a counter-regulatory response. The frequency of the PNPLA3 p.I148M variant increases from sub-Saharan Africa to South America in a pattern that reflects human migration. This pattern suggests that the variant may have been under positive selective pressure. Could the M variant be part of the "thrifty genome," as has been suggested previously?

The findings reported here raise the possibility that consideration of adiposity and genotype jointly may improve prediction of individuals at highest risk of progressing from simple steatosis to chronic liver disease. The risk alleles of the three strongest NAFLD risk variants confer only moderate risk in lean individuals but are major risk factors in people with higher BMIs, suggesting that genetic screening would be especially valuable in this subgroup. Similarly, while all obese individuals would benefit from weight-loss intervention, our data suggest that individuals at high genetic risk of NAFLD are likely to benefit the most.

URLs. R statistical analysis software, https://www.R-project.org.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

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

ACKNOWLEDGMENTS

This work was supported by grants from the US National Institutes of Health (NIH) (PO1 HL20948 and RO1 DK90066 to H.H.H. and J.C.C. and UL1 TR001105 to H.H.H.) and The Danish Council for Independent Research, Medical Sciences (Sapere Aude 4004-00398 to S.S.). The Copenhagen cohort is supported by the Danish Council for Independent Research, the Research Fund at Rigshospitalet, Copenhagen University Hospital, Chief Physician Johan Bosstrup and Lise Boserup’s Fund, Ingeborg and Leo Dammin’s Grant, Henry Hansen and Wilse’s Grant, and a grant from the Odd Fellow Order (to A.T.-H.).

AUTHOR CONTRIBUTIONS

S.S.: study concept and design, analysis and interpretation of data, drafting of the manuscript, statistical analysis, and critical revision of the manuscript. J.K.: analysis and interpretation of data, statistical analysis, and critical revision of the manuscript. A.T.-H. and B.G.M.: acquisition of data and critical revision of the manuscript. H.H.H.: study concept and design, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript, acquisition of data, and study supervision. J.C.C.: study concept and design, analysis and interpretation of data, drafting of the manuscript, critical revision of the manuscript, acquisition of data, and study supervision.

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. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Manolio, T.A. et al. Finding the missing heritability of complex diseases. Nature 461, 74–75 (2009).
2. Pritchard, J.K. Are rare variants responsible for susceptibility to complex diseases? Am. J. Hum. Genet. 69, 124–137 (2001).
3. Cohen, J.C. et al. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science 305, 869–872 (2004).
4. Yang, J. et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat. Genet. 47, 1114–1120 (2015).
5. Zuk, O., Hechtler, E., Sunyaev, S.R. & Lander, E.S. The mystery of missing heritability: Genetic interactions create phantom heritability. Proc. Natl. Acad. Sci. USA 109, 1193–1198 (2012).
6. Purcell, S. Variance components models for gene–environment interaction in twin analysis. Twin Res. 5, 554–571 (2002).
7. Kaprio, J. Twins and the mystery of missing heritability: the contribution of gene–environment interactions. J. Intern. Med. 272, 440–448 (2012).
8. Sadee, W. et al. Missing heritability of common diseases and treatments outside the protein-coding exome. Hum. Genet. 133, 1199–1215 (2014).
9. Wang, L., McLeod, H.L. & Weinshilboum, R.M. Genomics and drug response. N. Engl. J. Med. 364, 1144–1153 (2011).
10. Samson, M. et al. Resistance to HIV-1 infection in Venezuelan individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 382, 722–725 (1996).
11. Barón, A.E. et al. Interactions between ultraviolet light and MC1R and OCA2 variants are determinants of childhood nevus and freckle phenotypes. Cancer Epidemiol. Biomarkers Prev. 23, 2829–2839 (2014).
12. Reddon, H., Guéant, J.L. & Meyre, D. The importance of gene–environment interactions in human obesity. Clin. Sci. (Lond.) 130, 1571–1597 (2016).
13. Manuck, S.B. & McCaffery, J.M. Gene-environment interaction. *Annu. Rev. Psychol.* 65, 41–70 (2014).
14. Ludwig, J., Viggiano, T.R., McGill, D.B. & Oh, B.J. Nonalcoholic steatohepatitis: Mayo Clinic experiences with a hitherto unnamed disease. *Mayo Clin. Proc.* 55, 434–438 (1980).
15. Romes, S. et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.* 40, 1461–1465 (2008).
16. Kozlitina, J. et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.* 46, 352–356 (2014).
17. Santoro, N. et al. Hepatic de novo lipogenesis in obese youth is modulated by a common variant in the GCKR gene. *J. Clin. Endocrinol. Metab.* 100, E1125–E1132 (2015).
18. 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).
19. 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).
20. Falconer, D.S. in *Introduction to Quantitative Genetics* 1st edn. 292–301 (Ronald Press Company, 1960).
21. Thompson, W.D. Effect modification and the limits of biological inference from epidemiologic data. *J. Clin. Epidemiol.* 44, 221–232 (1991).
22. Voorman, A., Lumley, T., McKnight, B. & Rice, K. Behavior of QQ-plots and genomic control in studies of gene-environment interaction. *PLoS One* 6, e19416 (2011).
23. Nguyen, T.A. & Sanyal, A.J. Pathophysiology guided treatment of nonalcoholic steatohepatitis. *J. Gastroenterol. Hepatol.* 27 (suppl. 2), 58–64 (2012).
24. Cohen, J.C., Horton, J.D. & Hobbs, H.H. Human fatty liver disease: old questions and new insights. *Science* 332, 1519–1523 (2011).
25. Speliotes, E.K. et al. Association analyses of 249,795 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* 42, 937–948 (2010).
26. Mancina, R.M. et al. The MBOAT7–TMC4 variant rs641738 increases risk of nonalcoholic fatty liver disease in individuals of European descent. *Gastroenterology* 150, 1219–1230 (2016).
27. Smagris, E. et al. *PNPLA3* I148M knockin mice accumulate PNPLA3 on lipid droplets and develop hepatic steatosis. *Hepatology* 61, 108–118 (2015).
28. Rees, M.G. et al. Cellular characterisation of the GCKR P446L variant associated with type 2 diabetes risk. *Diabetologia* 55, 114–122 (2012).
29. Huang, Y. et al. A feed-forward loop amplifies nutritional regulation of PNPLA3. *Proc. Natl. Acad. Sci. USA* 107, 7892–7897 (2010).
30. Arden, C. et al. Elevated glucose represses liver glucokinase and induces its regulatory protein to safeguard hepatic phosphate homeostasis. *Diabetes* 60, 3110–3120 (2011).
31. Smagris, E., Gilyard, S., BasuRay, S., Cohen, J.C. & Hobbs, H.H. Inactivation of Tm6sf2, a gene defective in fatty liver disease, impairs lipidation but not secretion of very low density lipoproteins. *J. Biol. Chem.* 291, 10659–10676 (2016).
32. Graff, M. et al. PNPLA3 gene–by-visceral adipose tissue volume interaction and the pathogenesis of fatty liver disease: the NHLBI family heart study. *Int. J. Obes. (Lond.)* 37, 432–438 (2013).
33. York, L.W., Puthalapattu, S. & Wu, G.Y. Nonalcoholic fatty liver disease and low-carbohydrate diets. *Annu. Rev. Nutr.* 29, 365–379 (2009).
34. Davis, J.N. et al. Increased hepatic fat in overweight Hispanic youth influenced by interaction between genetic variation in PNPLA3 and high dietary carbohydrate and sugar consumption. *Am. J. Clin. Nutr.* 92, 1522–1527 (2010).
35. Williams, C.D. et al. Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study. *Gastroenterology* 140, 124–131 (2011).
36. 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).
37. Chambers, J.C. et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat. Genet.* 43, 1131–1138 (2011).
38. Larrieta-Carrasco, E. et al. Association of the I148M/PNPLA3 variant with elevated alanine transaminase levels in normal-weight and overweight/obese Mexican children. *Gene* 520, 185–188 (2013).
39. Giudice, E.M. et al. The association of PNPLA3 variants with liver enzymes in childhood obesity is driven by the interaction with abdominal fat. *PLoS One* 8, e27933 (2011).
40. Hart, C.L., Morrison, D.S., Baty, G.D., Mitchell, R.J. & Davey Smith, G. Effect of body mass index and alcohol consumption on liver disease: analysis of data from two prospective cohort studies. *Br. Med. J.* 340, c1240 (2010).
41. Tian, C., Stokowski, R.P., Kershnerobich, D., Ballinger, D.G. & Hinds, D.A. Variant in PNPLA3 is associated with alcoholic liver disease. *Nat. Genet.* 42, 21–23 (2010).
42. Locke, A.E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 518, 197–205 (2015).
43. Ehret, G.B. et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* 478, 103–109 (2011).
44. Dupuis, J. et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* 42, 105–116 (2010).
45. Neel, J.V. Diabetes mellitus: a “thrifty” genotype rendered detrimental by “progress”? *Am. J. Hum. Genet.* 14, 353–362 (1962).
46. Browning, J.D., Cohen, J.C. & Hobbs, H.H. Patatin-like phospholipase domain-containing 3 and the pathogenesis and progression of pediatric nonalcoholic fatty liver disease. *Hepatology* 52, 1189–1192 (2010).
ONLINE METHODS

Ethics. Studies were approved by institutional review boards and ethics committees of the University of Texas Southwestern Medical Center and by Danish institutional review boards and ethics committees and were conducted according to the Declaration of Helsinki. Written informed consent was obtained from participants. There was no overlap of individuals between the studies.

Participants. We included participants from four studies: the Dallas Heart Study (DHS), the Dallas Biobank, the Copenhagen City Heart Study (CCHS), and the Copenhagen General Population Study (CGPS). The DHS is a multiethnic, probability-based sample of Dallas County that was collected between 2000 and 2002 and between 2007 and 2009 (refs. 15,19). Ethnicity was self-reported in accordance with US census categories. From the DHS, we included 2,675 participants in whom hepatic triglyceride content (HTGC) was measured\(^47\), and up to 1,786 additional individuals were added to the sample size for the analysis of other traits. We included 5,434 people from the Dallas Biobank, a general population cohort of African Americans and Hispanic Americans from Dallas, Texas\(^16\). The CCHS and CGPS are prospective studies of the Danish general population initiated in 1976 and 2003, respectively\(^16,48\). All participants from the CCHS and CGPS were white and of Danish descent, as determined by the Danish Civil Registration System. We combined the CCHS and CGPS into one cohort, totaling 93,719 people, referred to here as the Copenhagen cohort.

Measurements. BMI was measured as weight (in kilograms) divided by height (in meters) squared. Hepatic triglyceride content (HTGC) was measured in the DHS using proton magnetic resonance spectroscopy\(^47\). Hepatic steatosis was defined as an HTGC of 5.5% or greater; 5.5% represents the ninety-fifth percentile of the distribution of HTGC in a population with no risk factors for steatosis\(^47\). Serum levels of ALT were measured as described\(^49\). PNPLA3 rs738409 (p.I148M; NC_000002.11:g.43928847C>G, p.Ile148Met), TM6SF2 rs58542926 (p.E167K; NC_000009.10:g.19268740C>T, p.Glu167Lys), and GCKR rs12063026 (p.P446L; NC_000002.11:g.27730940C>T, p.Pro446Leu), the 30 BMI-associated variants, and the exome-wide variants used to screen for associations and interactions with other phenotypes were genotyped in the Dallas Heart Study by an exome chip as previously described\(^49\). PNPLA3 p.I148M, TM6SF2 p.E167K, and GCKR p.P446L were genotyped by TaqMan in the Dallas Biobank and by TaqMan and PCR-based KASP genotyping in the Copenhagen cohort. Alcohol intake in the Copenhagen cohort was self-reported.

Cirrhosis. In the Copenhagen cohort, diagnoses of cirrhosis (International Statistical Classification of Diseases and Related Health Problems, revision 8 (ICD8): 57109 (alcoholic cirrhosis), K74.0 (hepatic fibrosis), K74.6 (unspecified cirrhosis)) were recorded from the National Danish Patient Registry and the National Danish Causes of Death Registry from 1 January 1977 to 10 November 2014. The National Danish Patient Registry has information on all patient contacts with all clinical hospital departments in Denmark, including emergency wards and outpatient clinics (from 1994). The National Danish Causes of Death Registry contains data on the causes of all deaths in Denmark, as reported by hospitals and general practitioners. A validation study in the Danish registry found that 85.4% of patients with an ICD code for cirrhosis fulfilled the diagnostic criteria for cirrhosis\(^50\).

Statistical analysis. All analyses were performed using Stata/SE 12 (Stata Corp.) and/or R statistical analysis software v3.2.3. A two-sided \(P < 0.05\) was considered statistically significant in all main analyses, whereas \(P < 3.6 \times 10^{-7}\) was considered significant in the exome-wide screen. For statistical tests, genotypes were coded 0, 1, or 2. BMI was entered as a continuous variable in all analyses (apart from a sensitivity test for interaction, in which BMI groups were entered as an ordered categorical variable, encoded 0–3). To depict the interaction between genotype and BMI visually, participants were divided into four groups of BMI: lean (\(\leq 25\) kg/m\(^2\)), overweight (25–30 kg/m\(^2\)), obese (30–35 kg/m\(^2\)), and very obese (>35 kg/m\(^2\)). The distributions of HTGC and ALT were highly skewed to the right\(^19\). Therefore, before entering these variables into regression analyses, we transformed them to HTGC\(^0.3\) and 1/(ALT\(^0.25\)) to approximate normality and constant variance of the residuals. These transformations were selected by using Tukey’s ladder of power transformations, and by visual inspection of Q-Q plots of residuals after the transformation. To assess the robustness of the interactions on different scales, we also used untransformed, inverse normally transformed, logarithmically transformed, and dichotomized HTGC or ALT. For each transformation, we plotted distributions of the variable, the normal Q-Q plot of the residuals, and distribution of the residuals by BMI category, and tested for BMI–SNP interactions (Supplementary Figs. 1 and 2). To account for a higher variance in HTGC in the most obese compared to lean subjects (heteroscedasticity), we repeated all interaction tests using a heteroscedasticity–robust model\(^22\).

We considered whether adjusting for BMI and PNPLA3, GCKR, or TM6SF2 in the models could introduce collider bias\(^51\). This was deemed unlikely given that none of the three genetic variants associate with BMI and that NAFLD is not known to causally influence adiposity. Prevalence of cirrhosis and steatosis were evaluated by logistic regression models adjusted for sex and age (and ethnicity in the DHS). We evaluated the interactions between BMI and SNPs by the inclusion of interaction terms between BMI and SNPs in the linear or logistic regression models, adjusted for sex, age, and ethnicity (encoded African American = 1, European American = 2, Hispanic American = 3, and entered as a factorial variable in the regression). BMI and SNPs were entered as continuous variables in the interaction term (i.e., all interaction tests are 1 degree of freedom). In a sensitivity analysis, the interaction on cirrhosis was retested after further adjustment for alcohol–BMI and alcohol–PNPLA3 interaction, entered individually or simultaneously into the regression.

To test whether adiposity is a likely causal risk factor for increased HTGC, we genotyped 30 SNPs known to be associated with BMI in individuals of European ancestry\(^25\) (Supplementary Table 3). For each SNP, the BMI-increasing alleles were weighted by the per-allele effect size reported in the GWAS\(^22\). A gene score was calculated for each European American participant of the DHS by summation of weighted alleles across all 30 BMI-associated SNPs. The gene score was tested for association with BMI and HTGC using linear regression, with the gene score included as a continuous variable. To depict the association between the genotype score and BMI and HTGC visually, the genotype score was divided into quintiles (Supplementary Fig. 5). Instrumental variable analysis was conducted to compare the observational association between BMI and HTGC with the effect of genetically increased BMI on HTGC\(^22\). The observational association between BMI and HTGC\(^0.3\) was determined using linear regression, adjusted for age and sex. For the genetic causal analysis, two-stage least-squares regression was used to assess the effect of a 1 kg/m\(^2\) increase in genetically modeled BMI on HTGC\(^0.3\) (ref. 52). Strength of the genetic instrument was evaluated by \(F\)-statistics and \(R^2\), where \(F > 10\) was considered sufficient to avoid weak-instrument bias, and \(R^2\) indicates the fraction of variation in BMI explained by the instrument.

To determine whether gene–environment interactions were commonly observed with other obesity-associated traits, we screened phenotypes relevant to metabolism (plasma lipids, glucose and insulin homeostasis, blood pressure, liver enzymes, sterols, biomarkers) for correlation with BMI in the DHS. Phenotypes showing a partial correlation with BMI (after adjustment for age, gender, and ethnicity) exceeding 0.2 in absolute value were further screened for association with genetic variants present on the Illumina HumanExome BeadChip (12v1_A)\(^19\). Variants exceeding our exome-wide significance threshold (\(P < 3.6 \times 10^{-7}\) and variants in established genetic loci from a previously published NAFLD GWAS\(^18\) were then tested for SNP–BMI interaction using linear regression adjusted for age, gender, and ethnicity. All significant SNP–BMI interactions (\(P < 0.05\)) were retested in the Dallas Biobank and in the Copenhagen cohort (where both phenotype and genotype data were available).
Data availability. The data that support the findings of this study are available within the article and its supplementary information and from the corresponding authors upon reasonable request.

47. Szczepaniak, L.S. et al. Magnetic resonance spectroscopy to measure hepatic triglyceride content: prevalence of hepatic steatosis in the general population. Am. J. Physiol. Endocrinol. Metab. 288, E462–E468 (2005).

48. Stender, S., Frikke-Schmidt, R., Nordestgaard, B.G. & Tybjaerg-Hansen, A. The ABCG5/8 cholesterol transporter and myocardial infarction versus gallstone disease. J. Am. Coll. Cardiol. 63, 2121–2128 (2014).

49. 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).

50. Vestberg, K. et al. Data quality of administratively collected hospital discharge data for liver cirrhosis epidemiology. J. Med. Syst. 21, 11–20 (1997).

51. Day, F.R., Loh, P.R., Scott, R.A., Ong, K.K. & Perry, J.R. A robust example of collider bias in a genetic association study. Am. J. Hum. Genet. 98, 392–393 (2016).

52. Burgess, S., Small, D.S. & Thompson, S.G. A review of instrumental variable estimators for Mendelian randomization. Stat. Methods Med. Res. http://dx.doi.org/10.1177/0962280215597579 (2015).