patient-oriented and epidemiological research

Novel association of *TM6SF2* rs58542926 genotype with increased serum tyrosine levels and decreased apoB-100 particles in Finns

Daniel Seung Kim,[1,*] Anne U. Jackson,[*] Yatong K. Li,[*] Heather M. Stringham,[*] FinMetSeq Investigators, Johanna Kuusisto,[†,§] Antti J. Kangas,[**] Pasi Soininen,[,††] Mika Ala-Korpela,[**,††,§§] Charles F. Burant,[***,†††] Veikko Salomaa,[§§§] Michael Boehnke,[1,2,*] Markku Laakso,[1,2,†,§] and Elizabeth K. Speliotes[1,2,†††,****]**

Department of Biostatistics and Center for Statistical Genetics,* Division of Metabolism, Endocrinology, and Diabetes, Department of Medicine,*** Center for Computational Medicine and Bioinformatics,[†††] and Division of Gastroenterology, Department of Medicine,**** University of Michigan, Ann Arbor, MI; Institute of Clinical Medicine, Internal Medicine,[†] and Nuclear Magnetic Resonance Metabolomics Laboratory, School of Pharmacy,[††] University of Eastern Finland, Kuopio, Finland; Department of Medicine,[§] Kuopio University Hospital, Kuopio, Finland; Computational Medicine,** Faculty of Medicine, University of Oulu and Biocenter Oulu, Oulu, Finland; Computational Medicine,[§§] School of Social and Community Medicine and Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom; and National Institute for Health and Welfare,[§§§] Helsinki, Finland

Abstract A glutamate-to-lysine variant (rs58542926-T) in transmembrane 6 superfamily member 2 (*TM6SF2*) is associated with increased fatty liver disease and diabetes in conjunction with decreased cardiovascular disease risk. To identify mediators of the effects of *TM6SF2*, we tested for associations between rs58542926-T and serum lipoprotein/metabolite measures in cross-sectional data from nondiabetic statin-naïve participants. We identified independent associations between rs58542926-T and apoB-100 particles ($\beta = -0.057$ g/l, $P = 1.99 \times 10^{-14}$) and tyrosine levels ($\beta = 0.0020$ mmol/l, $P = 1.10 \times 10^{-8}$), controlling for potential confounders, in 6,929 Finnish men. The association between rs58542926-T and apoB-100 was confirmed in an independent sample of 2,196 Finnish individuals from the FINRISK study ($\beta_{replication} = -0.029$, $P_{replication} = 0.029$). Secondary analyses demonstrated an rs58542926-T dose-dependent decrease in particle concentration, cholesterol, and triglyceride (TG) content for VLDL and LDL particles ($P < 0.001$ for all). No significant associations between rs58542926-T and HDL measures were observed. *TM6SF2* SNP rs58542926-T and tyrosine levels were associated with increased incident T2D risk in both METSIM and FINRISK. Decreased liver production/secretion of VLDL, decreased cholesterol and TGs in VLDL/LDL particles in serum, and increased tyrosine levels identify possible mechanisms by which rs58542926-T exerts its effects on increasing risk of fatty liver disease, decreasing cardiovascular disease, and increasing diabetes risk, respectively.—Kim, D. S., A. U. Jackson, Y. K. Li, H. M. Stringham, FinMetSeq Investigators, J. Kuusisto, A. J. Kangas, P. Soininen, M. Ala-Korpela, C. F. Burant, V. Salomaa, M. Boehnke, M. Laakso, and E. K. Speliotes. **Novel association of *TM6SF2* rs58542926 genotype with increased serum tyrosine levels and decreased apoB-100 particles in Finns. *J. Lipid Res.* 2017.** 58: **1471–1481.**

Supplementary key words transmembrane 6 superfamily member 2 • nonalcoholic fatty liver disease • epidemiology • cardiovascular disease • type 2 diabetes • apolipoprotein B-100 • very low density lipoprotein

This work was supported by Academy of Finland Grants 77299, 124243, and 141226 (M.L.); the Finnish Heart Foundation (M.L.); the Finnish Diabetes Foundation (M.L.); the Juselius Foundation (ML); National Institutes of Health Grants R01 DK062370 (M.B.), R01 DK106621 (E.K.S.), and R01 DK107904 (E.K.S); and American Heart Association Grant 16POST27250048 (D.S.K.). Additional support was provided by the Sigrid Juselius Foundation, Strategic Research Funding from the University of Oulu, Finland, the University of Bristol and UK Medical Research Council (MC_UU_12013/1) (M.A.K.), and the Finnish Foundation for Cardiovascular Research (V.S.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. A.J.K. and P.S. are shareholders and report employment relation for Nightingale Health, a company offering NMR-based metabolite profiling. The remaining authors declare no conflicts of interest.

Manuscript received 3 March 2017 and in revised form 12 May 2017.

Published, JLR Papers in Press, May 24, 2017
DOI https://doi.org/10.1194/jlr.P076034

Abbreviations: ALT, alanine aminotransferase; C/C, homozygous for the major allele of *TM6SF2* SNP rs58542926; C/T, heterozygous genotype (one major and one minor allele) for *TM6SF2* SNP rs58542926; FDR, false discovery rate; HR, hazard ratio; ISI, index of insulin sensitivity; LDL-C, cholesterol content of LDL particles; LDL-P, particle content of LDL particles; LDL-TG, triglyceride content of LDL particles; NAFLD, nonalcoholic fatty liver disease; OGTT, oral glucose tolerance test; TG, triglyceride; *TM6SF2*, transmembrane 6 superfamily member 2; T/T, homozygous genotype for the minor allele for *TM6SF2* SNP rs58542926; VLDL-C, cholesterol content of VLDL particles; VLDL-P, particle content of VLDL particles; VLDL-TG, triglyceride content of VLDL particles.

[1] To whom correspondence should be addressed.
e-mail: markku.laakso@kuh.fi (M.L.); espeliot@umich.edu (E.K.S.)

[2] M. Boehnke, M. Laakso, and E. K. Speliotes contributed equally to this work.

The online version of this article (available at http://www.jlr.org) contains a supplement.

Nonalcoholic fatty liver disease (NAFLD) is defined by the presence of more than 5% hepatic steatosis in the absence of other known causes of liver fat accumulation (e.g., alcohol use). NAFLD is common in developed countries, with an estimated prevalence of 30–60% in adults, with significant differences by genetic ancestry (1). Hepatic steatosis can predispose to development of advanced liver disease, including cirrhosis, liver failure, and hepatocellular carcinoma (2). NAFLD is associated with the metabolic syndrome; some consider it the hepatic manifestation of the metabolic syndrome (3). NAFLD has been shown to be predictive of incident T2D and cardiovascular disease, after controlling for obesity (4), suggesting that hepatic factors may contribute to the development of metabolic syndrome, possibly through influencing hepatic glucose and lipid metabolism (3).

Genome-wide association studies have identified risk loci associated with NAFLD [see recent review in (5)]. Variants at the 19p13.11 locus encompassing *NCAN-TM6SF2-SUGP1-CILP2* were identified early on to associate with both serum lipid levels (6) and NAFLD (7) and have since been fine mapped to a rare missense variant (rs58542926-T, encoding p.Glu167Lys) in transmembrane 6 superfamily member 2 (*TM6SF2*). This region and variant associate with decreased serum LDL and triglyceride (TG) levels (6, 8), decreased risk of cardiovascular disease (8), and recently increased risk of T2D (9) and NAFLD (10).

In vitro experiments in human Huh7 and HepG2 cell lines demonstrated that *TM6SF2* was localized to the endoplasmic reticulum and the endoplasmic reticulum-Golgi intermediate compartment (11). In addition, *TM6SF2* silencing via RNA inhibition resulted in decreased TG and apoB secretion suggestive of decreased total VLDL secretion and increased intracellular lipid droplet content (11). Conversely, overexpression of *TM6SF2* reduced liver cell steatosis, indicating a role for *TM6SF2* in liver fat metabolism (11). Separate short hairpin RNA knockdown of *Tm6sf2* in mice resulted in increased liver TG accumulation (10) in conjunction with decreased plasma total cholesterol and decreased TG content in VLDL fractions (8, 10). Transient overexpression of *Tm6sf2* in the liver of mice using adenoviruses from a separate group showed increased total and LDL cholesterol and increased TG, whereas knockdown decreased total serum cholesterol (8). These lipid perturbations from functional gene silencing experiments, increased hepatic steatosis from intracellular TG accumulation in conjunction with decreased serum total cholesterol and TG levels, are consistent with human population-level associations with rs58542926-T (6–8, 10). These data together suggest a possible loss-of-function role for rs58542926-T with regard to TM6SF2 protein function in causing increased hepatic steatosis while decreasing total serum cholesterol (5).

Follow-up studies of *tm6sf2*$^{-/-}$ KO mice, however, have led to different proposed molecular models of how *TM6SF2* may affect lipid metabolism. The first study found that *tm6sf2*$^{-/-}$ KO mice had altered gene expression and plasma lipid levels suggestive of a role of *TM6SF2* in cholesterol biosynthesis (12). The second study using *tm6sf2*$^{-/-}$ KO mice reported decreases in plasma cholesterol and TG levels and a decrease in VLDL particle size without a concurrent decrease in apoB levels (13). These results from mouse models were suggestive of either decreased cholesterol biosynthesis (12) or decreased TG loading of VLDL without altered secretion of apoB reflecting total VLDL particle secretion (13). This latter result in mice conflicts with the originally reported findings of reduced apoB secretion in two human cell lines when *TM6SF2* function is reduced (11).

Further investigations into the metabolic effects of the *TM6SF2* p.Glu167Lys mutation in humans are necessary to better understand the role of *TM6SF2* in NAFLD pathogenesis and lipid metabolism. Here, we leveraged high-throughput NMR lipoprotein and metabolic data from nondiabetic and statin-naïve men from the Metabolic Syndrome in Men (METSIM) study and identified two metabolic variables (apoB and tyrosine levels) that are altered in carriers of rs58542926-T. We replicated the association of decreased serum apoB and rs58542926-T in the independent FINRISK study. In secondary analyses, we demonstrated an association of rs58542926-T with increased T2D risk, which may be mediated by tyrosine levels; both rs585429260T and tyrosine levels were significantly associated with increased hazard of developing diabetes. These results help to identify possible metabolic mechanisms by which the *TM6SF2* p.Glu167Lys mutation promotes liver disease and T2D, while protecting against cardiovascular disease.

MATERIALS AND METHODS

Ethics statement

Both the METSIM and FINRISK studies were performed in accordance with the Helsinki Declaration and were approved by the ethics committee of the University of Kuopio and Kuopio University and the National Public Health Institute of Finland, respectively. Informed, written consent was obtained for all METSIM participants.

METSIM study participants

The METSIM study is a population-based study, with participants between the ages of 45 and 70 years randomly selected from the population register of the town of Kuopio in Eastern Finland. Each participant had a 1 day outpatient visit, from which history of previous diseases and current medication list were obtained. Evaluation of metabolic syndrome and cardiovascular risk factors were also assessed at this outpatient visit. Fasting blood samples were drawn after 12 h of fasting, followed by an oral glucose tolerance test (OGTT) (14). Of the 10,197 METSIM participants, we examined 6,926 subjects chosen to exclude those with diabetes at baseline (n = 915) and those undergoing statin pharmacotherapy (n = 2,353), as both conditions are associated with altered metabolic and lipid characteristics (15, 16). A review of the METSIM study and its major findings was recently reported in the *Journal of Lipid Research* (17).

Insulin sensitivity calculation

Results from the OGTT were used to calculate the Matsuda index of insulin sensitivity (ISI) as 10,000/sqrt (fasting insulin × fasting glucose × mean insulin during OGTT × mean glucose during OGTT) (18).

Serum NMR metabolomics

Fasting serum samples for METSIM were collected at enrollment, stored at −80°C, and thawed overnight in a refrigerator before sample preparation. A high-throughput serum NMR metabolomics platform was used to quantify the levels of individual metabolites and lipoprotein measures. The NMR-based metabolic profiling was previously used in multiple large-scale epidemiological and genetic studies (19) and the experimentation described elsewhere (20, 21).

The NMR-based quantification of lipoprotein subclasses was calibrated using HPLC (21, 22). The lipoprotein subclass sizes measured were as follows: six VLDL subclasses, ranging from extremely large (average particle diameter 64.0 nm) to very small (31.3 nm); three LDL subclasses, ranging from large (25.5 nm) to small (18.7 nm); and four HDL subclasses, ranging from very large (14.3 nm) to small (8.7 nm). For each lipoprotein subclass, particle concentration, TG content, and cholesterol content were quantified (see supplemental Table S1).

Genotyping, imputation, and quality control

METSIM participant samples were genotyped on the HumanOmniExpresss-12v1_C BeadChip (OmniExpress) and Infinium HumanExome-12 v1.0 BeadChip (Exome Chip) platforms. Quality controls in the METSIM study included sample-level controls for identifying sex and relatedness confirmation, sample duplication, and detection of sample genetic ancestry outliers using principal-components analysis. Based on these quality control measures, we removed 14 samples with sex chromosome anomalies, 18 with evidence of participant duplication, and 12 that were population outliers. Additionally, we removed one individual from seven monozygotic twin pairs. Variants in the METSIM study were filtered for low mapping quality of probes to genome build GRCh37, low genotype completeness (<95% and <98% for the OmniExpress and ExomeChip, respectively), or Hardy-Weinberg equilibrium $P < 10^{-6}$.

OmniExpress variants passing quality control were then phased with Shape-It v2 (23) and imputed using minimac v2 (24). Imputation used a reference panel of 26.7 M variants from the GoT2D study (including SNPs, indels, and large deletions) based on the whole genome sequence of 2,874 Europeans, including 1,004 Finnish individuals, the largest panel of Finnish genomes available (9). Following imputation, variants directly genotyped on the ExomeChip were added. In cases of discordance between imputed and genotyped variants, the directly genotyped call from the ExomeChip was used.

FINRISK replication study participants

We performed replication analyses in the FINRISK study, a cross-sectional general population study with data collected every 5 years to assess the health of the Finnish population between 25 and 74 years of age (25). The statin-naïve and nondiabetic 2,196 participants analyzed in this work were recruited in 1997. NMR-based metabolite and lipoprotein measurements have previously been described (26–28), and were analyzed by the same laboratory using the same methods that produced the aforementioned METSIM NMR data. Genotyping in FINRISK was performed on Illumina platforms (26, 29) and imputation was performed similarly to METSIM (26).

Analysis

We performed all analyses using standard R (https://www.r-project.org/) packages. Due to the right skew of the lipoprotein variables, we Winsorized extreme observations of all lipoprotein variables to three SDs from the mean (30), to allow for natural variation in lipid traits while minimizing the effect of potentially influential outliers. Genotypes of genetic variants were coded using an additive model.

Association testing. We used the *TM6SF2* SNP rs58542926 genotype as the outcome variable, as it allowed for use of stepwise linear regression to identify independent predictors of genotype (which would not be possible in the converse, with rs58542926 as the predictor). Due to the cross-sectional design of the study, the association (*P* value) did not change between two variables when dependent/independent roles were switched. Thus, for interpretability, we present the rescaled estimated β coefficients (β) and SEs for the effect of each minor allele of rs58542926 on each covariate.

As lipoprotein variables are highly correlated, we used a two-step approach to identify independent metabolic and lipoprotein predictors of *TM6SF2* SNP rs58542926. First, we performed univariate tests of association for each metabolic and lipoprotein variable across genotype group, using the Pearson test for categorical variables and the Kruskal-Wallis test for continuous variables. Forty-three lipoprotein and amino acid variables that were nominally significantly different by *TM6SF2* SNP rs58542926 genotype were carried forward to the second multivariable step.

We used stepwise linear regression to identify independent lipoprotein and amino acid predictors of *TM6SF2* SNP rs58542926 genotype with the aforementioned 43 variables entering the model. Model comparison was performed using Akaike's information criterion to examine the fit of each model, beginning with a base model that included age, BMI, current smoking status, serum alanine aminotransferase (ALT) activity, and Matsuda ISI. Lipoprotein and amino acid variables that were included in the final stepwise linear regression model independently increased the ability of the model to predict *TM6SF2* SNP rs58542926 genotype. Robust SE estimates were used to account for the lack of homoscedasticity in lipoprotein variables across genotype groups. We used a false discovery rate (FDR) threshold of $<10^{-4}$ for all variables for inclusion in the final regression model to correct for multiple comparisons. A sensitivity analysis considering recessive genotype coding of rs58542926 was performed.

Replication analyses. We attempted to replicate our primary findings from METSIM in 2,196 FINRISK participants with demographic, clinical, NMR metabolite, and lipoprotein data, and *TM6SF2* SNP rs58542926 genotype. Due to the lack of Matsuda ISI and serum ALT data, we adjusted our replication analyses for age, sex (FINRISK is composed of both men and women), BMI, current smoking status, and tyrosine (when apoB-100 was the outcome studied) and apoB-100 (when tyrosine was the outcome studied). We used linear regression with robust SEs, as described above. Reanalysis of the association of rs58542926-T with apoB-100 and tyrosine was performed in METSIM with the updated replication covariates.

Association with incident T2D in METSIM and FINRISK. We performed analyses of the association of both *TM6SF2* SNP rs58542926-T and tyrosine levels with incident T2D in statin-naïve and nondiabetic participants in both METSIM and FINRISK. Longitudinal analyses used a Cox proportional hazards model with robust SEs, adjusting for baseline age, female gender (for FINRISK), BMI, and current smoking status. To visualize the

association of tyrosine and incident T2D, we plotted covariate-adjusted cumulative T2D incidence of the top 50th percentile as compared with the bottom 50th percentile for tyrosine levels.

RESULTS

Baseline clinical, lipid, and lipoprotein variables are presented in **Table 1**, stratified by *TM6SF2* SNP rs58542926 genotype. Univariate associations of specific lipoprotein variables (e.g., particle concentration of extremely large VLDL) with rs58542926-T are presented in supplemental Table S1. We observed a difference in apoB-100 levels ($P < 0.001$), but not in apoA-I levels ($P = 0.53$) in univariate analyses across genotype group. Accordingly, we observed nominally significant ($P < 0.001$) differences in apoB-100-associated particles: VLDL and LDL, both of which had genotype-dependent differences in their particle concentration (VLDL-P and LDL-P), TG content (VLDL-TG and LDL-TG), and cholesterol content (VLDL-C and LDL-C). We did identify differences by rs58542926 genotype for the mean diameter of VLDL ($P = 0.027$), but not LDL or HDL. For metabolite measures, citrate ($P = 0.034$), phenylalanine ($P = 0.007$), and tyrosine ($P < 0.001$) were nominally significant for differences by *TM6SF2* SNP rs58542926 genotype. Pairwise correlation coefficients for *1*) broad serum lipid and lipoprotein variables, *2*) particle concentrations, *3*) TG content, *4*) cholesterol content, and *5*) metabolite and amino acid variables are presented in supplemental Figs. S1−S5, respectively.

A total of 43 lipoprotein and metabolite variables were nominally significant ($P < 0.05$) for association with *TM6SF2* SNP rs58542926-T (see Table 1, supplemental Table S1) and were carried forward to stepwise linear regression to identify which of the 43 variables represented the strongest independent signals in multivariate analysis. Beginning with a base regression model containing age, BMI, smoking status, ALT, and Matsuda ISI, we identified two additional variables that improved model prediction of rs58542926 with $FDR_{stepwise} < 10^{-4}$: apoB-100 ($\beta = -0.057$ g/l per minor T allele, $P = 1.99 \times 10^{-14}$) and tyrosine ($\beta = 0.0020$ mmol/l per minor T allele, $P = 1.10 \times 10^{-8}$). We noted that once apoB-100 was in the model, no other lipoprotein variable

TABLE 1. Baseline clinical, lipoprotein, and metabolic measures of the subset of the METSIM study (n = 6,929), stratified by *TM6SF2* SNP rs58542926 genotype group

Trait Description	Trait Abbreviation	Units	C/C (n = 6,119)	C/T (n = 784)	T/T (n = 26)	Two-sided P^a
Baseline and clinical characteristics						
Age at study enrollment	Age	years	56.5 (7.0)	56.6 (7.0)	58.6 (6.1)	0.23
BMI	BMI	kg/m^2	26.8 (3.9)	26.9 (3.8)	27.4 (3.6)	0.44
Current smoking status	Current smoking status	%	1058 (17)	120 (15)	5 (19)	0.35
Serum alanine transaminase	ALT	units/l	32 (22)	33 (21)	32 (23)	0.16
Matsuda index of insulin sensitivity	Matsuda ISI		7.1 (4.2)	7.1 (4.4)	6.6 (4.3)	0.77
Apolipoproteins						
apoA-I	apo-AI	g/l	1.58 (0.21)	1.57 (0.22)	1.56 (0.26)	0.53
apoB-100	apoB-100	g/l	1.08 (0.22)	1.03 (0.21)	0.96 (0.21)	<0.001
Cholesterol measures						
Serum total cholesterol	C	mmol/l	5.5 (1.1)	5.3 (1.0)	5.0 (1.2)	<0.001
Total cholesterol in VLDL	VLDL-C	mmol/l	0.92 (0.32)	0.85 (0.29)	0.75 (0.27)	<0.001
Total cholesterol in LDL	LDL-C	mmol/l	2.23 (0.58)	2.14 (0.58)	2.04 (0.61)	<0.001
Total cholesterol in HDL	HDL-C	mmol/l	1.43 (0.35)	1.45 (0.35)	1.45 (0.35)	0.37
TG measures						
Serum total TGs	TG	mmol/l	1.53 (0.88)	1.39 (0.79)	1.20 (0.52)	<0.001
Total TGs on VLDL	VLDL-TG	mmol/l	1.02 (0.63)	0.91 (0.56)	0.77 (0.41)	<0.001
Total TGs on LDL	LDL-TG	mmol/l	0.21 (0.069)	0.20 (0.067)	0.19 (0.072)	<0.001
Total TGs on HDL	HDL-TG	mmol/l	0.14 (0.046)	0.14 (0.044)	0.14 (0.048)	0.26
Lipoprotein particle concentrations						
Concentration of VLDL particles	VLDL-P	uM/l	0.11 (0.039)	0.10 (0.035)	0.093 (0.030)	<0.001
Concentration of LDL particles	LDL-P	uM/l	0.64 (0.16)	0.62 (0.16)	0.59 (0.16)	<0.001
Concentration of HDL particles	HDL-P	uM/l	8.7 (1.4)	8.7 (1.4)	8.9 (1.6)	0.77
Lipoprotein particle size						
Mean diameter for VLDL particles	VLDL-D	nm	36.6 (1.5)	36.4 (1.4)	36.2 (0.9)	0.027
Mean diameter for LDL particles	LDL-D	nm	23.52 (0.20)	23.53 (0.22)	23.52 (0.19)	0.77
Mean diameter for HDL particles	HDL-D	nm	9.85 (0.28)	9.87 (0.28)	9.81 (0.20)	0.35
Amino acid and metabolite measures						
Alanine	Ala	mmol/l	0.42 (0.063)	0.42 (0.062)	0.42 (0.068)	0.92
Citrate	Cit	mmol/l	0.11 (0.021)	0.11 (0.022)	0.12 (0.014)	0.034
Glutamine	Gln	mmol/l	0.52 (0.074)	0.52 (0.073)	0.54 (0.10)	0.81
Glucose	Glu	mmol/l	4.63 (0.75)	4.64 (0.64)	4.78 (0.65)	0.12
Glycine	Gly	mmol/l	0.27 (0.042)	0.27 (0.042)	0.27 (0.041)	0.52
Histidine	His	mmol/l	0.067 (0.0098)	0.068 (0.0098)	0.071 (0.010)	0.16
Isoleucine	Ile	mmol/l	0.059 (0.017)	0.058 (0.016)	0.058 (0.0094)	0.12
Leucine	Leu	mmol/l	0.091 (0.018)	0.091 (0.018)	0.090 (0.012)	0.99
Phenylalanine	Phe	mmol/l	0.077 (0.012)	0.078 (0.013)	0.082 (0.010)	0.007
Pyruvate	Pyr	mmol/l	0.068 (0.022)	0.066 (0.020)	0.060 (0.016)	0.085
Tyrosine	Tyr	mmol/l	0.057 (0.012)	0.058 (0.012)	0.067 (0.012)	<0.001
Valine	Val	mmol/l	0.22 (0.037)	0.22 (0.036)	0.22 (0.030)	0.45

[a]The Pearson test was used to test for differences of proportions (current smoking status) across genotype groups. The Kruskal-Wallis test was used to test for differences in means across genotype groups. No adjustment for potential confounders was used for these tests.

(including the specific subclass measures) significantly increased model prediction (**Table 2**). A sensitivity analysis using recessive coding of *TM6SF2* SNP rs58542926 found consistent direction of effects and nominally significant associations for apoB-100 [β = −0.13 g/l for those with the homozygous for the minor (NAFLD risk increasing) allele (T/T) genotype as compared with the heterozygous (C/T) and homozygous for the major allele (C/C) genotype groups, $P = 0.0012$] and tyrosine (β = 0.0089 mmol/l for those with the T/T genotype as compared with the C/T and C/C genotype group, $P = 4.71 \times 10^{-6}$) levels (see supplemental Table S2).

To determine the significance of our apoB-100 and tyrosine associations with *TM6SF2* SNP rs58542926-T, we present replication analyses from 2,196 nondiabetic and statin-naïve participants of the independent FINRISK study in **Table 3**. The association of apoB-100 and rs58542926-T was significant in FINRISK (β = −0.029, $P = 0.029$). The association of tyrosine with rs58542926-T was not significant in FINRISK (β = 0.0010, $P = 0.15$); however, the direction of effect was consistent with METSIM.

We performed secondary multivariate regression analyses in METSIM to identify potential lipid pathway perturbations associated with the *TM6SF2* SNP rs58542926-T and apoB-100 association (see supplemental Table S3). Visualization of the covariate-adjusted association of decreasing apoB-100 levels per T allele of rs58542926 ($P = 1.99 \times 10^{-14}$) is presented in **Fig. 1A**. We noted a lack of association between genotype and apoA-I ($P = 0.67$). Examination of covariate-adjusted serum lipid levels revealed a nominally significant dose-dependent decrease for both TG ($P = 2.70 \times 10^{-12}$) and cholesterol ($P = 1.61 \times 10^{-8}$) levels (see Fig. 1B). **Figure 2** shows the association of the particle concentrations, TG levels, and cholesterol content of VLDL (Fig. 2A), LDL (Fig. 2B), and HDL (Fig. 2C). Consistent with the univariate analyses, we saw no association across rs58542926 genotype for any HDL measure ($P > 0.05$). Particle concentration, TG levels, and cholesterol content of both VLDL and LDL all decreased with each T allele of *TM6SF2* SNP rs58542926 ($P < 0.001$ for all observations; Fig. 2A, B). Examination of univariate genotype association with particle diameter revealed nominally significant differences in VLDL diameter ($P = 0.027$). Follow-up examination of covariate-adjusted VLDL particle diameter did not reveal differences by genotype ($P = 0.13$, see supplemental Fig. S6).

TABLE 2. Final stepwise regression model of lipoprotein and metabolic variables associating with *TM6SF2* SNP rs58542926-T (n = 6,929)

Variable	B ± SE[a]	FDR	*P*
Age, years	−0.10 ± 0.24	0.77	0.67
BMI, kg/m^2	0.0016 ± 0.12	0.99	0.99
ALT, IU	0.58 ± 0.72	0.73	0.42
Current smoker, %	−0.0099 ± 0.013	0.73	0.46
Matsuda ISI	0.051 ± 0.12	0.77	0.68
ApoB-100, g/l	−0.057 ± 0.0074	1.59×10^{-13}	1.99×10^{-14}
Tyrosine, mmol/l	0.0020 ± 0.00036	4.41×10^{-8}	1.10×10^{-8}

B, estimated β coefficient.
[a]As this is a cross-sectional study, presented B values are rescaled to represent the effect of each minor allele of rs58542926.

To elucidate the role of *TM6SF2* in T2D pathogenesis, we analyzed the association of rs58542926-T and tyrosine levels with incident T2D in both METSIM and FINRISK. Using Cox proportional hazards modeling, we found that *TM6SF2* SNP rs58542926-T was significantly associated with incident T2D in METSIM [hazard ratio (HR) = 1.24, $P = 0.046$] and FINRISK (HR = 1.42, $P = 0.0059$; see **Fig. 3A**, supplemental Table S4). Analyses of the association of tyrosine levels demonstrated a strong pro-diabetes effect in both METSIM and FINRISK (see supplemental Table S5): those with higher tyrosine levels (>50th percentile) had a 63% and 118% increased hazard of incident diabetes as compared with the lower tyrosine levels group in METSIM (HR = 1.63, $P = 8.22 \times 10^{-6}$) and FINRISK (HR = 2.18, $P = 0.00014$), respectively (see Fig. 3B).

DISCUSSION

Since its discovery, the *TM6SF2* p.Glu167Lys missense variant has been associated with decreased total serum cholesterol, LDL and TG concentration, and decreased myocardial infarction risk (8). Separately, knockdown of *tm6sf2* expression in mouse liver has been causally linked to increased hepatic steatosis (10) and rs58542926 has since been associated with a wide spectrum of liver disease (31–33). This paradoxical association, where rs58542926-T decreases serum lipid levels and therefore decreases cardiovascular disease risk, but increases liver disease risk through increased hepatic fat accumulation, has been confirmed in meta-analyses (34) and recently has been the topic of a commentary (35). This same variant has now also been associated with increased risk of T2D (9). Despite the importance of rs58542926-T in metabolic disease risk, how rs58542926-T exerts its effect to predispose to these metabolic changes is unknown (36).

To help elucidate its mechanism of action, we aimed to identify metabolites associated with the *TM6SF2* p.Glu167Lys missense variant. Using high-throughput NMR lipoprotein and metabolite profiling of a large sample (n = 6,929) of genetically homogeneous, nondiabetic, and statin-naïve Finnish men, we identified two highly significant associations in our data: a novel genotype-dependent increase in tyrosine levels (β = 0.0020 mmol/l, $P = 1.10 \times 10^{-8}$) and a strong decrease in apoB-100 levels (β = −0.057 g/l, $P = 1.99 \times 10^{-14}$) per T allele of rs58542926. The association of apoB-100 levels with rs58542926-T was subsequently replicated in the independent FINRISK study (β = −0.029, $P = 0.029$). The association of tyrosine levels with rs58542926-T was not replicated in FINRISK (β = 0.0010, $P = 0.15$); however, given the same direction of effect in both cohorts and the smaller more diverse sample size of FINRISK (2,129 men and women in FINRISK vs. 6,929 men in METSIM), the failure of replication in this work may reflect a lack of statistical power to detect an association. Moreover, given the strong association of tyrosine levels with incident T2D in both METSIM and FINRISK, alterations of tyrosine levels by *TM6SF2* SNP rs58542926 genotype may reflect a potential causal pathway through which *TM6SF2* affects diabetes risk.

TABLE 3. Association of *TM6SF2 SNP* rs58542926-T with apoB-100 replicates in the independent FINRISK cohort (n = 2,196)

Replication Variable	n	β ± SE	P^a
ApoB-100, g/l			
METSIM	6,929	−0.057 ± 0.0081	2.44×10^{-12}
FINRISK	2,196	−0.029 ± 0.013	0.029
Tyrosine, mmol/l			
METSIM	6,929	0.0021 ± 0.00039	6.53×10^{-8}
FINRISK	2,196	0.0010 ± 0.00074	0.15

[a]Replication analyses in both METSIM and FINRISK excluded for those with known diabetes and on lipid-lowering medication at time of enrollment. Linear regression models adjusted for age, sex (in FINRISK), BMI, current smoking status, and tyrosine (when apoB-100 was the outcome) and apoB-100 (when tyrosine was the outcome).

Sensitivity analyses of associations with *TM6SF2* SNP rs58542926-T showed decreases in total apoB-100-containing particles (VLDL-P and LDL-P) and their TG (VLDL-TG and LDL-TG) and cholesterol content (VLDL-C and LDL-C). We also observed a lack of effect of rs58542926 genotype with apoA-I levels and HDL measures, which are the primary carriers of apoA-I. Finally, the particle diameters of VLDL, LDL, and HDL were not affected by rs58542926 genotype in multivariate analyses.

These lipoprotein data are consistent with the potential mechanism of decreased production or secretion of VLDL particles from the liver in individuals carrying the T (minor) allele of rs58542926. This hypothesis of a decrease in VLDL particle concentration would also result in decreased LDL particle concentration and decreased TG/cholesterol levels of these particle classes in the serum [we note that in our data that apoB, VLDL-P, VLDL-TG, and VLDL-C are all highly correlated (supplemental Fig. S1)]. As with apoB mutations causing familial hypobetalipoproteinemia (37), the replicated finding that *TM6SF2* p.Glu167Lys is associated with significantly decreased apoB-100 levels suggests that the variant may act to promote accumulation of lipids and cholesterol in liver to cause fatty liver disease and advanced liver disease, while decreasing the levels of these lipids in the serum. Although these data are consistent with the hypothesis of decreased VLDL secretion (11), we cannot exclude from our cross-sectional analyses the possibility that production of VLDL is affected.

Within the context of prior literature on *TM6SF2*, our data are in apparent conflict with recent work by Smagris et al. (13) using a *tm6sf2*$^{-/-}$ KO mouse model from which the authors concluded that, while *TM6SF2* was implicated in lipidation of VLDL particles in the liver, it was not involved in secretion of VLDL. Data from Smagris et al. (13) that supported this conclusion was the decreased hepatic secretion of VLDL TGs and the lack of decreased hepatic apoB-100 secretion in conjunction with decreased VLDL diameter, all of which suggests decreased TG content of VLDL, but no decrease in absolute VLDL particle concentration in serum. With regard to another recent report using a *tm6sf2*$^{-/-}$ KO mouse, our data cannot refute the hypothesis that *TM6SF2* is involved in cholesterol biosynthesis (12). Further work in large-scale human studies of gene expression with enough participants with the minor allele homozygous genotype at rs58542926 (of which we had 17) will be needed to determine whether *TM6SF2* loss-of-function results in altered expression in cholesterol biosynthesis genes (12).

The conflict between our results, which are consistent with prior studies in humans (8, 11), and with the recent work using *tm6sf2*$^{-/-}$ KO mice (12, 13), may reflect differences between the species. First, the lipoproteome of VLDL in mice is poorly understood as compared with humans (38). It is possible that regulation and secretion of VLDL differs in mice compared with humans, similar to the example of mice carrying the majority of cholesterol on HDL due to a lack of cholesterol ester transport protein (39). Second, humans have significantly different genetic regulation as compared with mice, as reflected by data from the Mouse ENCODE project, which found that approximately half of the transcription factor binding sites in the mouse genome were not found in the orthologous human genetic regions, while a quarter of transcription sites migrated to different positions within the regulatory element (40). Third, the effect of metabolic stressors, particularly the effects of diet, may differ between species. A 10 week high-fat diet commonly used to induce fatty liver disease, for example, did not result in the development of excess hepatic steatosis in mice; an effect that may be observed by the significantly higher daily intake and liver biosynthesis of cholesterol in mice (41).

In addition to our lipoprotein findings, we report the novel discovery that tyrosine levels are associated with *TM6SF2* SNP rs58542926-T in METSIM and that tyrosine levels are independently associated with increased incidence of T2D in both METSIM and FINRISK. We note that while the association of tyrosine and rs58542926-T did not replicate in FINRISK, this may be due to smaller sample size and cohort heterogeneity, as the effect size direction was consistent with what was observed in METSIM and the *P* value was suggestive of significance. Tyrosine, an aromatic amino acid, has previously been associated in a partially overlapping subset of METSIM (n = 9,369) to be predictive of incident diabetes at 4.7 year follow-up (15). We further confirmed this association in a smaller statin-naïve subset of METSIM (n = 6,929) with approximately 12 years of follow-up time, and found that individuals in the top 50th percentile of tyrosine levels have a 1.69 greater hazard of incident T2D as compared with individuals in the lower 50th percentile. We similarly report that tyrosine levels are strongly predictive of incident T2D in a statin-naïve subset of the FINRISK study, with individuals having higher tyrosine levels at 2.18 greater hazard of incident T2D. Furthermore, in an independent and nonoverlapping population-based study of Finnish individuals, tyrosine levels were associated with baseline measures of insulin resistance, and at 6 year follow-up in men only (42), and separately, with measures of oral glucose tolerance at baseline and 6.5 year follow-up (43).

These tyrosine association data in aggregate suggest a role of rs58542926-T in the development of diabetes via alterations to aromatic amino acid metabolism. This hypothesis is strengthened by recent genetic fine-mapping of low frequency variants by the GoT2D and T2D-GENES

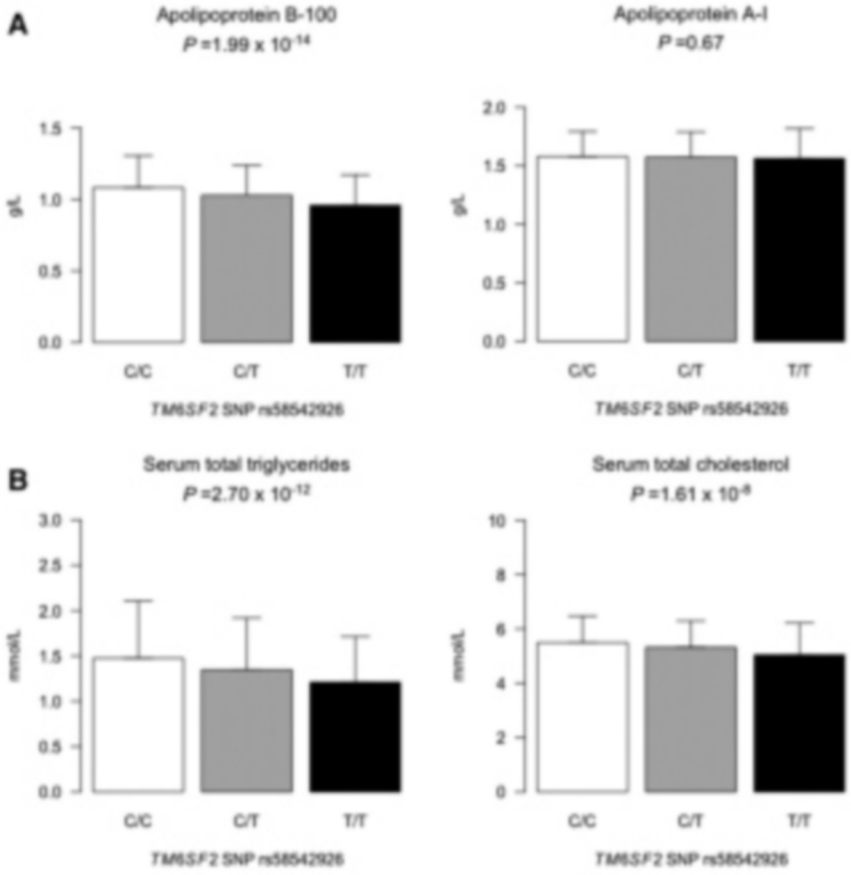

Fig. 1. Associations between *TM6SF2* SNP rs58542926-T and apolipoprotein levels (A) and serum lipid levels (B). Bars show the unadjusted means and SEs. *P* values were adjusted for age, BMI, ALT, current smoking status, Matsuda ISI, and tyrosine levels using linear regression with robust SEs.

consortia, from which they concluded that *TM6SF2* SNP rs58542926 was the likely causal variant for diabetes in the 19p13.11 locus (9). In this work, each T minor allele of rs58542926 was associated with 21% increased odds of T2D (9). We validate this association: each T minor allele rs58542926 was significantly associated with 24% and 42% increased hazard of incident T2D in METSIM and FINRISK, respectively. As a result of our data and prior work, it is plausible to speculate that *TM6SF2* rs58542926-T increases tyrosine levels, which in turn increase T2D risk. However, it remains to be determined whether the effect of rs58542926 on tyrosine is direct on its metabolism or indirect via effects on lipid levels (44).

Some limitations of this study should be considered. First, this analysis was of cross-sectional data; therefore, we could not prove proposed mechanisms of action from these data alone. Second, the METSIM population in which discovery analyses were performed consisted solely of men; therefore, generalizability to women will need further evaluation in similarly large cohorts. However, using just men resulted in greater homogeneity across important risk factors for chronic disease and likely increased sensitivity to detect true positives (see prior discussion on the lack of replication of the tyrosine association in FINRISK). Third, concerns with multiple collinearity are important with analyses of highly correlated lipoprotein data. In our analyses, we used stepwise linear regression to identify independent predictors from among the numerous correlated lipoprotein variables. We note that our two significant variables, apoB-100 and tyrosine levels, are not highly correlated ($r = 0.20$). Fourth, we lacked liver biopsy and NMR liver fat measurements; thus, we were unable to make inferences as to the effect of rs58542926 on hepatic histologic processes. Future work is needed to assess this genotype-dependent effect on hepatic pathology on a large population scale.

The strengths and weakness of NMR metabolomic and lipoprotein profiling should be reviewed as compared with MS methods [see (19) for a recent review]. The strengths of NMR include lower cost and higher throughput, as compared with MS, which allow NMR metabolic and lipoprotein profiling to be used in large-scale studies such as METSIM. Weaknesses of NMR as compared with MS include less specific data, as MS separates signals from individual molecular identities with mass differences and thus leads to rich and complicated spectrometric data. Moreover, there is a strong focus on lipid species with NMR metabolomic profiling. Our study was designed to identify changes in lipoprotein particles and their cholesterol/TG content to address recent hypotheses by Smagris et al. (13) and Fan et al. (12) that TM6SF2 affected VLDL lipidation, but not secretion and cholesterol biosynthesis, respectively. As a result, our study would not have been able to assess

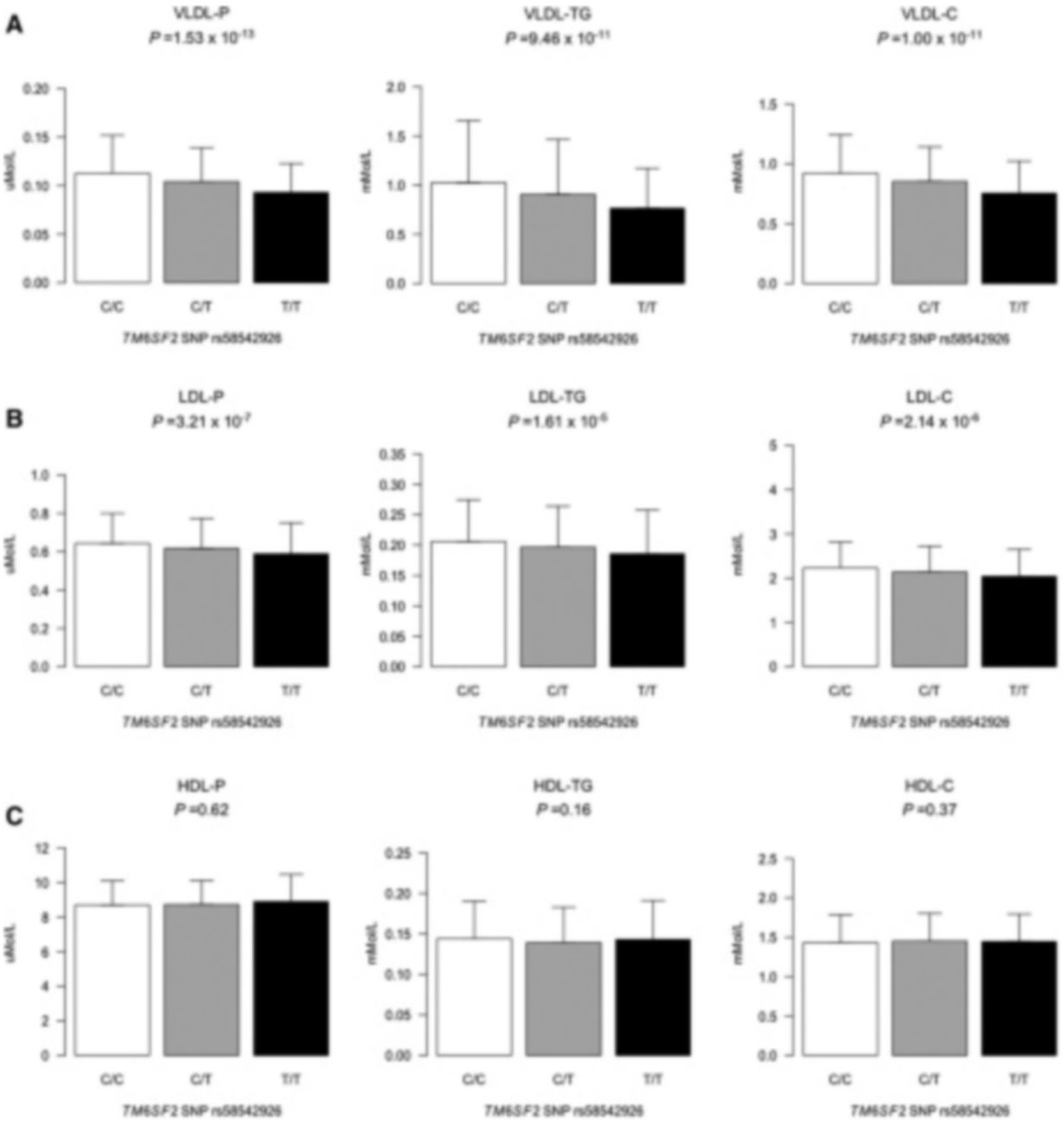

Fig. 2. Associations between *TM6SF2* SNP rs58542926-T and particle concentrations, TG content, and cholesterol of VLDL (A), LDL (B), and HDL (C). Bars show the unadjusted means and SEs. *P* values were adjusted for age, BMI, ALT, current smoking status, Matsuda ISI, and tyrosine levels using linear regression with robust SEs.

these potential mechanisms with MS metabolomic data, as lipoproteins are not captured by the method.

In summary, we performed a novel study associating detailed lipoprotein subclass data and multiple circulating metabolite measures with *TM6SF2* SNP rs58542926-T. We identified genotype-dependent decreases in the concentrations of apoB-100 particles: VLDL and LDL, and their TG and cholesterol contents. These findings are most consistent with a defect in VLDL production or secretion from hepatocytes. We also report the novel association of rs58542926-T with tyrosine levels and validate prior reports of tyrosine levels increasing incident T2D risk. This finding that rs58542926-T affects tyrosine levels may reflect a pathway through which *TM6SF2* is associated with insulin resistance and incident T2D through alteration of tyrosine levels. Due to the importance of *TM6SF2* in lipid metabolism and disease risk, further molecular and large-scale studies are warranted.

The authors thank the participants of the METSIM study.

FinMetSeq Investigators: Haley J. Abel,[3,4] Michael Boehnke,[5] Lei Chen,[4] Charleston W. K. Chiang,[6] Colby C. Chiang,[4] Susan K. Dutcher,[3,4] Nelson B. Freimer,[6] Robert S. Fulton,[4] Liron Ganel,[3,4] Ira M. Hall,[4] Aki S. Havulinna,[7,8] Anne U. Jackson,[5] Krishna L.

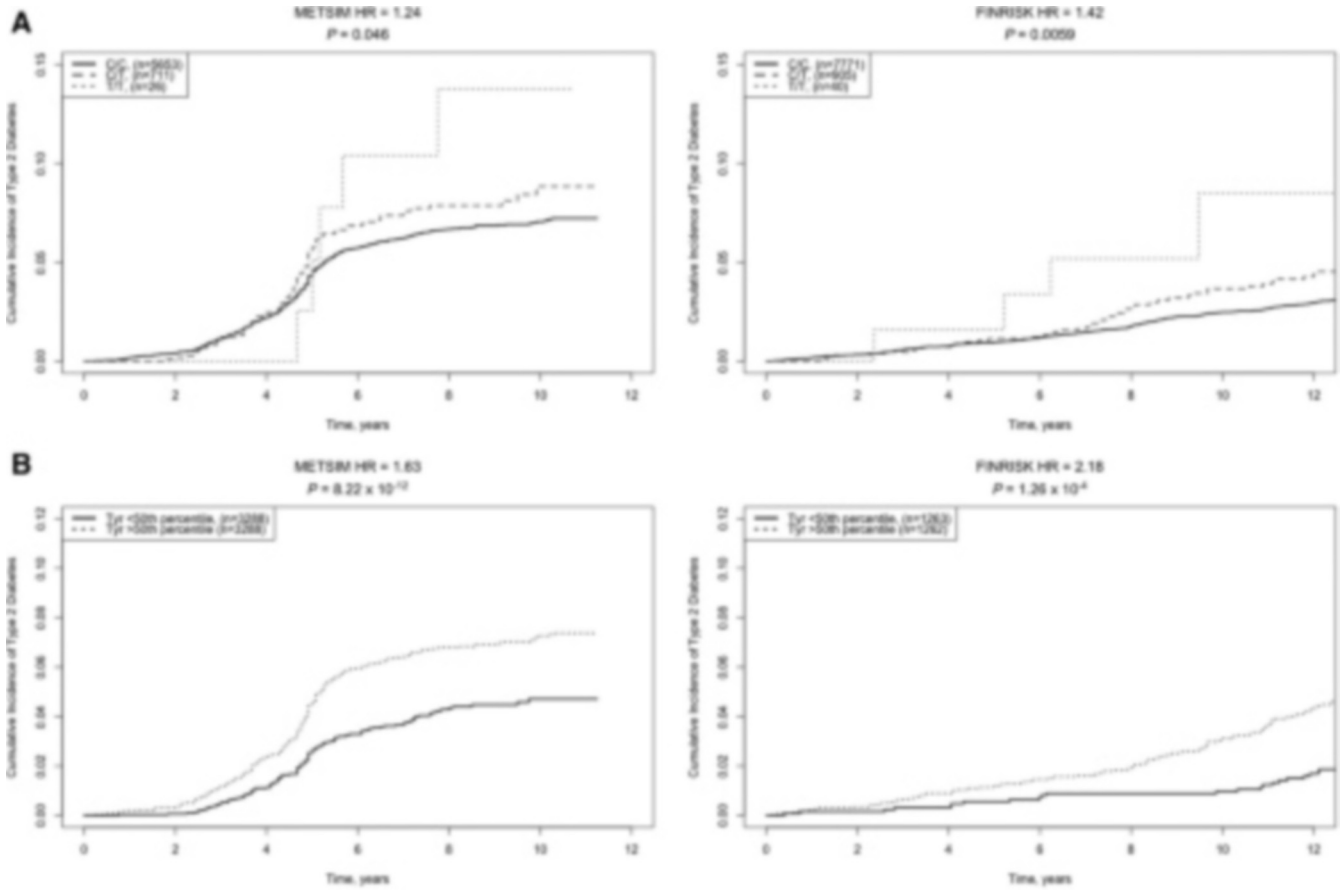

Fig. 3. Association of *TM6SF2* SNP rs58542926-T (A) and tyrosine levels (B) with incident T2D in the METSIM cohort. Plotted Cox proportional hazards output was adjusted for age, BMI, current smoking status, and physical activity level with robust SEs. See supplemental Tables S4 and S5 for full Cox proportional hazards model output for *TM6SF2* rs58542926 and tyrosine associations with incident diabetes, respectively.

Kanchi,[4] Chul Joo Kang,[4] Daniel C. Koboldt,[4,9,10] Mitja I. Kurki,[7,11,12] Markku Laakso,[13] Hannele Laivuori,[7,14,15] David E. Larson,[3,4] Adam E. Locke,[4,5] Karyn Meltz Steinberg,[4] Joanne Nelson,[4] Thomas J. Nicholas,[4] Aarno Palotie,[7,16,17] Arto Pietilä,[8] Matti Pirinen,[7,18,19] Vasily Ramensky,[6] Debashree Ray,[5] Samuli Ripatti,[7,18,20] Chiara Sabatti,[21,22] Veikko Salomaa,[8] Laura J. Scott,[5] Susan Service,[6] Laurel Stell,[21] Nathan O. Stitziel,[4,23] Heather M. Stringham,[5] Ryan Welch,[5] Richard K. Wilson,[4,9,24] and Pranav Yajnik[5].

[3]Department of Genetics, Washington University School of Medicine, St. Louis, MO; [4]McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO; [5]Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI; [6]Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA; [7]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; [8]National Institute for Health and Welfare, Helsinki, Finland; [9]Department of Pediatrics, Ohio State University College of Medicine, Columbus, OH; [10]Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH; [11]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; [12]Psychiatric and Neurodevelopmental Genetics Unit, Departments of Psychiatry and Neurology, Massachusetts General Hospital, Boston, MA; [13]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio, Finland; [14]Medical and Clinical Genetics, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; [15]Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; [16]Analytical and Translational Genetics Unit (ATGU), Psychiatric and Neurodevelopmental Genetics Unit, Departments of Psychiatry and Neurology, Massachusetts General Hospital, Boston, MA; [17]Broad Institute of MIT and Harvard, Cambridge, MA; [18]Department of Public Health, University of Helsinki, Helsinki, Finland; [19]Helsinki Institute for Information Technology HIIT and Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland; [20]Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; [21]Department of Biomedical Data Science, Stanford University, Stanford, CA; [22]Department of Statistics, Stanford University, Stanford, CA; [23]Cardiovascular Division, Department of Medicine, Washington University School of Medicine, St. Louis, MO; and [24]Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH.

REFERENCES

1. Williams, C. D., J. Stengel, M. I. Asike, D. M. Torres, J. Shaw, M. Contreras, C. L. Landt, and S. A. Harrison. 2011. 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.

2. McCullough, A. J. 2004. The clinical features, diagnosis and natural history of nonalcoholic fatty liver disease. *Clin. Liver Dis.* **8:** 521–533.
3. Yki-Järvinen, H. 2014. Non-alcoholic fatty liver disease as a cause and a consequence of metabolic syndrome. *Lancet Diabetes Endocrinol.* **2:** 901–910.
4. Anstee, Q. M., G. Targher, and C. P. Day. 2013. Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. *Nat. Rev. Gastroenterol. Hepatol.* **10:** 330–344.
5. Kahali, B., B. Halligan, and E. K. Speliotes. 2015. Insights from genome-wide association analyses of nonalcoholic fatty liver disease. *Semin. Liver Dis.* **35:** 375–391.
6. Teslovich, T. M., K. Musunuru, A. V. Smith, A. C. Edmondson, I. M. Stylianou, M. Koseki, J. P. Pirruccello, S. Ripatti, D. I. Chasman, C. J. Willer, et al. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature.* **466:** 707–713.
7. Speliotes, E. K., L. M. Yerges-Armstrong, J. Wu, R. Hernaez, L. J. Kim, C. D. Palmer, V. Gudnason, G. Eiriksdottir, M. E. Garcia, L. J. Launer, et al. 2011. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet.* **7:** e1001324.
8. Holmen, O. L., H. Zhang, Y. Fan, D. H. Hovelson, E. M. Schmidt, W. Zhou, Y. Guo, J. Zhang, A. Langhammer, M-L. Løchen, et al. 2014. Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nat. Genet.* **46:** 345–351.
9. Fuchsberger, C., J. Flannick, T. M. Teslovich, A. Mahajan, V. Agarwala, K. J. Gaulton, C. Ma, P. Fontanillas, L. Moutsianas, D. J. McCarthy, et al. 2016. The genetic architecture of type 2 diabetes. *Nature.* **536:** 41–47.
10. Kozlitina, J., E. Smagris, S. Stender, B. G. Nordestgaard, H. H. Zhou, A. Tybjaerg-Hansen, T. F. Vogt, H. H. Hobbs, and J. C. Cohen. 2014. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.* **46:** 352–356.
11. Mahdessian, H., A. Taxiarchis, S. Popov, A. Silveira, A. Franco-Cereceda, A. Hamsten, P. Eriksson, and F. van't Hooft. 2014. TM6SF2 is a regulator of liver fat metabolism influencing triglyceride secretion and hepatic lipid droplet content. *Proc. Natl. Acad. Sci. USA.* **111:** 8913–8918.
12. Fan, Y., H. Lu, Y. Guo, T. Zhu, M. T. Garcia-Barrio, Z. Jiang, C. J. Willer, J. Zhang, and Y. E. Chen. 2016. Hepatic transmembrane 6 superfamily member 2 regulates cholesterol metabolism in mice. *Gastroenterology.* **150:** 1208–1218.
13. Smagris, E., S. Gilyard, S. BasuRay, J. C. Cohen, and H. H. Hobbs. 2016. 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.
14. Stancáková, A., M. Javorský, T. Kuulasmaa, S. M. Haffner, J. Kuusisto, and M. Laakso. 2009. Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6,414 Finnish men. *Diabetes.* **58:** 1212–1221.
15. Stancáková, A., M. Civelek, N. K. Saleem, P. Soininen, A. J. Kangas, H. Cederberg, J. Paananen, J. Pihlajamäki, L. L. Bonnycastle, M. A. Morken, et al. 2012. Hyperglycemia and a common variant of GCKR are associated with the levels of eight amino acids in 9,369 Finnish men. *Diabetes.* **61:** 1895–1902.
16. Cholesterol Treatment Trialists' (CTT) Collaborators, B. Mihaylova, J. Emberson, L. Blackwell, A. Keech, J. Simes, E. H. Barnes, M. Voysey, A. Gray, R. Collins, et al. 2012. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. *Lancet.* **380:** 581–590.
17. Laakso, M., J. Kuusisto, A. Stančáková, T. Kuulasmaa, P. Pajukanta, A. J. Lusis, F. S. Collins, K. Mohlke, and M. Boehnke. 2017. The Metabolic Syndrome in Men study: a resource for studies of metabolic and cardiovascular diseases. *J. Lipid Res.* **58:** 481–493.
18. Matsuda, M., and R. A. DeFronzo. 1999. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. *Diabetes Care.* **22:** 1462–1470.
19. Soininen, P., A. J. Kangas, P. Würtz, T. Suna, and M. Ala-Korpela. 2015. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet.* **8:** 192–206.
20. Soininen, P., A. J. Kangas, P. Würtz, T. Tukiainen, T. Tynkkynen, R. Laatikainen, M-R. Jarvelin, M. Kähönen, T. Lehtimäki, J. Viikari, et al. 2009. High-throughput serum NMR metabonomics for cost-effective holistic studies on systemic metabolism. *Analyst.* **134:** 1781–1785.
21. Inouye, M., J. Kettunen, P. Soininen, K. Silander, S. Ripatti, L. S. Kumpula, E. Hämäläinen, P. Jousilahti, A. J. Kangas, S. Männistö, et al. 2010. Metabonomic, transcriptomic, and genomic variation of a population cohort. *Mol. Syst. Biol.* **6:** 441–450.
22. Okazaki, M., S. Usui, M. Ishigami, N. Sakai, T. Nakamura, Y. Matsuzawa, and S. Yamashita. 2005. Identification of unique lipoprotein subclasses for visceral obesity by component analysis of cholesterol profile in high-performance liquid chromatography. *Arterioscler. Thromb. Vasc. Biol.* **25:** 578–584.
23. Delaneau, O., J-F. Zagury, and J. Marchini. 2013. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods.* **10:** 5–6.
24. Fuchsberger, C., G. R. Abecasis, and D. A. Hinds. 2015. minimac2: faster genotype imputation. *Bioinformatics.* **31:** 782–784.
25. Vartiainen, E., T. Laatikainen, M. Peltonen, A. Juolevi, S. Mannisto, J. Sundvall, P. Jousilahti, V. Salomaa, L. Valsta, and P. Puska. 2010. Thirty-five-year trends in cardiovascular risk factors in Finland. *Int. J. Epidemiol.* **39:** 504–518.
26. Marttinen, P., M. Pirinen, A-P. Sarin, J. Gillberg, J. Kettunen, I. Surakka, A. J. Kangas, P. Soininen, P. O'Reilly, M. Kaakinen, et al. 2014. Assessing multivariate gene-metabolome associations with rare variants using Bayesian reduced rank regression. *Bioinformatics.* **30:** 2026–2034.
27. Vogt, S., S. Wahl, J. Kettunen, S. Breitner, G. Kastenmüller, C. Gieger, K. Suhre, M. Waldenberger, J. Kratzsch, M. Perola, et al. 2016. Characterization of the metabolic profile associated with serum 25-hydroxyvitamin D: a cross-sectional analysis in population-based data. *Int. J. Epidemiol.* **45:** 1469–1481.
28. Würtz, P., A. S. Havulinna, P. Soininen, T. Tynkkynen, D. Prieto-Merino, T. Tillin, A. Ghorbani, A. Artati, Q. Wang, M. Tiainen, et al. 2015. Metabolite profiling and cardiovascular event risk: a prospective study of 3 population-based cohorts. *Circulation.* **131:** 774–785.
29. Sulkava, S., H. M. Ollila, K. Ahola, T. Partonen, K. Viitasalo, J. Kettunen, M. Lappalainen, M. Kivimäki, J. Vahtera, J. Lindstrom, et al. 2013. Genome-wide scan of job-related exhaustion with three replication studies implicate a susceptibility variant at the UST gene locus. *Hum. Mol. Genet.* **22:** 3363–3372.
30. Dixon, W. J., and J. W. Tukey. 1968. Approximate behavior of the distribution of Winsorized t (Trimming/Winsorization 2). *Technometrics.* **10:** 83–98.
31. Buch, S., F. Stickel, E. Trépo, M. Way, A. Herrmann, H. D. Nischalke, M. Brosch, J. Rosendahl, T. Berg, M. Ridinger, et al. 2015. A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. *Nat. Genet.* **47:** 1443–1448.
32. Milano, M., A. Aghemo, R. M. Mancina, J. Fischer, P. Dongiovanni, S. De Nicola, A. L. Fracanzani, R. D'Ambrosio, M. Maggioni, R. De Francesco, et al. 2015. Transmembrane 6 superfamily member 2 gene E167K variant impacts on steatosis and liver damage in chronic hepatitis C patients. *Hepatology.* **62:** 111–117.
33. Coppola, N., Z. Rosa, G. Cirillo, M. Stanzione, M. Macera, A. Boemio, A. Grandone, M. Pisaturo, A. Marrone, L. E. Adinolfi, et al. 2015. TM6SF2 E167K variant is associated with severe steatosis in chronic hepatitis C, regardless of PNPLA3 polymorphism. *Liver Int.* **35:** 1959–1963.
34. Pirola, C. J., and S. Sookoian. 2015. The dual and opposite role of the TM6SF2-rs58542926 variant in protecting against cardiovascular disease and conferring risk for nonalcoholic fatty liver: a Meta-Analysis. *Hepatology.* **62:** 1742–1756.
35. Kahali, B., Y-L. Liu, A. K. Daly, C. P. Day, Q. M. Anstee, and E. K. Speliotes. 2015. TM6SF2: catch-22 in the fight against nonalcoholic fatty liver disease and cardiovascular disease? *Gastroenterology.* **148:** 679–684.
36. Sookoian, S., and C. J. Pirola. 2016. The exact mechanism by which hepatic transmembrane 6 superfamily member 2 modulates triglyceride metabolism is still uncertain. *Gastroenterology.* **151:** 1033–1034.
37. Elias, N., B. W. Patterson, and G. Schonfeld. 1999. Decreased production rates of VLDL triglycerides and ApoB-100 in subjects heterozygous for familial hypobetalipoproteinemia. *Arterioscler. Thromb. Vasc. Biol.* **19:** 2714–2721.
38. Gordon, S. M., H. Li, X. Zhu, A. S. Shah, L. J. Lu, and W. S. Davidson. 2015. A comparison of the mouse and human lipoproteome: suitability of the mouse model for studies of human lipoproteins. *J. Proteome Res.* **14:** 2686–2695.

39. Camus, M. C., M. J. Chapman, P. Forgez, and P. M. Laplaud. 1983. Distribution and characterization of the serum lipoproteins and apoproteins in the mouse, Mus musculus. *J. Lipid Res.* **24:** 1210–1228.
40. Stergachis, A. B., S. Neph, R. Sandstrom, E. Haugen, A. P. Reynolds, M. Zhang, R. Byron, T. Canfield, S. Stelhing-Sun, K. Lee, et al. 2014. Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature.* **515:** 365–370.
41. Dietschy, J. M., and S. D. Turley. 2002. Control of cholesterol turnover in the mouse. *J. Biol. Chem.* **277:** 3801–3804.
42. Würtz, P., P. Soininen, A. J. Kangas, T. Rönnemaa, T. Lehtimäki, M. Kähönen, J. S. Viikari, O. T. Raitakari, and M. Ala-Korpela. 2013. Branched-chain and aromatic amino acids are predictors of insulin resistance in young adults. *Diabetes Care.* **36:** 648–655.
43. Würtz, P., M. Tiainen, V-P. Mäkinen, A. J. Kangas, P. Soininen, J. Saltevo, S. Keinanen-Kiukaanniemi, P. Mäntyselkä, T. Lehtimäki, M. Laakso, et al. 2012. Circulating metabolite predictors of glycemia in middle-aged men and women. *Diabetes Care.* **35:** 1749–1756.
44. Würtz, P., Q. Wang, A. J. Kangas, R. C. Richmond, J. Skarp, M. Tiainen, T. Tynkkynen, P. Soininen, A. S. Havulinna, M. Kaakinen, et al. 2014. Metabolic signatures of adiposity in young adults: Mendelian randomization analysis and effects of weight change. *PLoS Med.* **11:** e1001765.