JHEP|Reports

# Deep metabolic phenotyping of humans with protein-altering variants in *TM6SF2* using a genome-first approach

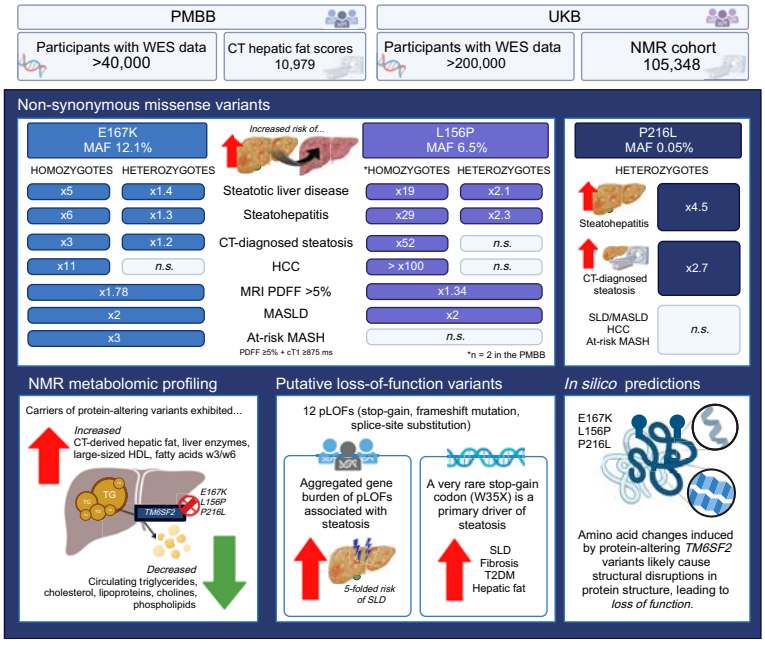

## Authors

Helen Ye Rim Huang, Cecilia Vitali, David Zhang, …, Kai Markus Schneider, Daniel J. Rader, Carolin Victoria Schneider

## Correspondence

rader@pennmedicine.upenn.edu (D.J. Rader), cschneider@ukaachen.de (C.V. Schneider).


## Graphical abstract




## Highlights:

- Unbiased genome-first approaches permit deeper metabolic phenotyping of select genes.
- Protein-altering variants in *TM6SF2* are drivers of hepatic steatosis and advanced fibrosis.
- Risk of liver disease in *TM6SF2* is independent of common *PNPLA3* I48M allele.
- Loss of protein function potentially causes aberrations in hepatic secretion of VLDL.

## Impact and implications:

The genome-first approach expands insights into genetic risk factors for steatotic liver disease with *TM6SF2* being a focal point due to its known association with plasma lipid traits. Our findings validated the association of two missense variants (E167K and L156P) with increased risk of hepatic steatosis on CT and MRI scans, as well as the risk of clinically diagnosed hepatocellular carcinoma independent of the common *PNPLA3* I48M risk variant. Notably, we also identified a predicted deleterious missense variant (P216L) linked to steatotic risk and demonstrated that an aggregated gene burden of rare putative loss-of-function variants was associated with the risk of hepatic steatosis. Combined, this study sets the stage for future mechanistic investigations into the functional consequences of *TM6SF2* variants in metabolic dysfunction-associated steatotic liver disease.

https://doi.org/10.1016/j.jhepr.2024.101243

Research article

JHEP|Reports

# Deep metabolic phenotyping of humans with protein-altering variants in *TM6SF2* using a genome-first approach

Helen Ye Rim Huang[1], Cecilia Vitali[1,2], David Zhang[1], Nicholas J. Hand[1,2], Michael C. Phillips[1], Kate Townsend Creasy[3], Eleonora Scorletti[2,4], Joseph Park[1,5], Regeneron Centre[6], Kai Markus Schneider[7,8], Daniel J. Rader[1,2,4,*,‡], Carolin Victoria Schneider[4,7,*,‡]

JHEP Reports **2025**. vol. 7 | 1–16



**Background & Aim:** An unbiased genome-first approach can expand the molecular understanding of specific genes in disease-agnostic biobanks for deeper phenotyping. *TM6SF2* represents a good candidate for this approach due to its known association with steatotic liver disease (SLD).

**Methods:** We screened participants with whole-exome sequences in the Penn Medicine Biobank (PMBB, n >40,000) and the UK Biobank (UKB, n >200,000) for protein-altering variants in *TM6SF2* and evaluated their association with liver phenotypes and clinical outcomes.

**Results:** Missense variants in *TM6SF2* (E167K, L156P, P216L) were associated with an increased risk of clinically diagnosed and imaging-proven steatosis, independent of the *PNPLA3* I48M risk allele and hepatitis B/C ($p$ <0.001). E167K homozygotes had significantly increased risk of SLD (odds ratio [OR] 5.38, $p$ <0.001), steatohepatitis (OR 5.76, $p$ <0.05) and hepatocellular carcinoma (OR 11.22, $p$ <0.0001), while heterozygous carriers of L156P and P216L were also at an increased risk of steatohepatitis. In addition, carriers of E167K are at a 3-fold increased risk of at-risk MASH (OR 2.75, $p$ <0.001). CT-derived liver fat scores were higher in E167K and L156P in an allele-dose manner ($p$ <0.05). This corresponded with the UKB nuclear magnetic resonance-derived lipidomic analyses (n = 105,348), revealing all carriers to exhibit lower total cholesterol, triglycerides and total choline. In silico predictions suggested that these missense variants cause structural disruptions in the EXPERA domain, leading to reduced protein function. This hypothesis was supported by the association of rare loss-of-function variants in *TM6SF2* with an increased risk of SLD (OR 4.9, $p$ <0.05), primarily driven by a novel rare stop-gain variant (W35X) with the same directionality.

**Conclusion:** The functional genetic study of protein-altering variants provides insights on the association between loss of *TM6SF2* function and SLD and provides the basis for future mechanistic studies.



## Introduction

Metabolic dysfunction-associated steatotic liver disease (MASLD) is one of the most common liver conditions, affecting 25% of the population worldwide and strongly linked to features of metabolic syndrome including insulin resistance and obesity.[1,2] In many cases, MASLD can progress into metabolic dysfunction-associated steatohepatitis (MASH) and more serious phenotypes such as fibrotic liver disease, cirrhosis, and hepatocellular carcinoma (HCC).[3] The newly revised nomenclature supports the diagnosis of MASLD in the presence of at least one of five affirmative cardiometabolic risk factors and encompasses multiple parameters that are implicated in lipid metabolism.[4] Such changes comes with persistent discussions on the delineation and terminology of liver disease classifications for steatotic liver disease (SLD).

Genome-wide association studies have greatly contributed to our understanding of MASLD.[5,6] Of the multiple genomic loci associated with an increased risk of MASLD, variants in transmembrane 6 superfamily member 2 (*TM6SF2*) located on chromosome 19 have been significantly associated with MASLD and plasma lipid traits.[7] *TM6SF2* encodes a protein of 351 amino acids and is mostly expressed in the liver and intestine.[8] *TM6SF2* has been shown to modulate hepatic secretion of very-low density lipoprotein (VLDL), which is thought to be a mechanism by which it influences both liver and plasma lipids.[9]

Genome-wide association studies are by definition a 'phenotype-first' approach. An unbiased 'genome-first' approach has the potential to expand the understanding of gene-phenotype associations by starting with individuals who carry protein-altering variants in specific genes and investigating a diverse range of phenotypes.[10–12] Utilizing this approach, we investigated selected protein-altering variants in *TM6SF2*, leveraging the Penn Medicine Biobank (PMBB), a

* Corresponding authors. Address: Division of Translational Medicine and Human Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; (D.J. Rader), or Department of Medicine III, Gastroenterology, Metabolic diseases and Intensive Care, University Hospital RWTH Aachen, 52074 Aachen, Germany. (C.V. Schneider).
*E-mail addresses:* rader@pennmedicine.upenn.edu (D.J. Rader), cschneider@ukaachen.de (C.V. Schneider).
‡ These authors share the last authorship
https://doi.org/10.1016/j.jhepr.2024.101243

large medical biobank with sequencing data linked to electronic health record data, and the UK Biobank (UKB), a large population-based biobank with whole-exome sequencing (WES) and extensive phenotype data.

## Patients and methods

### *TM6SF2* variant selection

All available non-synonymous missense and putative loss-of-function (pLOF) *TM6SF2* variants were extracted from the Genome Aggregation Database (gnomAD) and annotated using ANNOVAR. For non-synonymous missense variants, individual regression models were applied to those with carriers over >40 in the PMBB pLOF variants were defined as stop-gain codons, frame-shift substitutions, and disruption of canonical splice site dinucleotides. pLOF carriers were first aggregated into a gene burden, followed by regression studies of individual variants.

### Penn Medicine Biobank

*Study population (discovery cohort)*

The PMBB comprises data from clinical practice sites of the University of Pennsylvania Health System from over 60,000 participants. Participants in the PMBB consented for access to all available electronic health record data and genetic sequencing. This study was approved by the Institutional Review Board and complies with the principles set out by the Declaration of Helsinki. The PMBB was utilized as the discovery cohort of non-synonymous missense variants and pLOFs associated with steatosis.

*Genotyping*

WES data for participants in the PMBB were generated from DNA extracted from stored buffy coats by the Regeneron Genetics Center. Sequences were mapped to the Genome Reference Consortium Build 38 (GRCh38). Samples with low exome sequencing coverage and other quality metrics were removed from the PMBB as previously described.[12] After quality control measures, the *TM6SF2* gene was analyzed in a total of 41,759 participants with WES data.

*Clinical data collection*

Between the baseline assessment and July 2022, ongoing hospital and outpatient records were analyzed to determine diagnoses using International Classification of Diseases Tenth Revision (ICD-10) codes. The presence of the following primary ICD10 codes was evaluated: Liver disease (K71), hepatic failure (K72), chronic hepatitis (K73), fibrosis and cirrhosis (K74), inflammatory liver diseases (K75), steatohepatitis (K75.81), SLD (K76.0), and malignant neoplasm of the liver and/or bile ducts (C22). For the discovery of steatosis-associated variants, K76.0 and K75.81 were used as baseline variables. As well as ICD-10 codes, baseline characteristics (age, sex, BMI) procedural billing codes, medication, and laboratory measurements were extracted from the electronic health record database in PMBB. Serum parameters were extracted for participants from the time of enrolment in the PMBB until April 28, 2022. Because data on alcohol intake was not available in the PMBB, participants with alcoholic liver disease (571.0, K70.0), alcoholic hepatitis (571.1, K70.1), alcoholic fibrosis and sclerosis of the liver (571.2, K70.3), alcoholic cirrhosis of liver and/or ascites (571.2, K70.2), alcoholic hepatic failure, coma, and unspecified alcoholic liver disease (571.3, K70.4, K70.40, K70.41, K70.9) were excluded (n = 439). Participants with a history of chronic viral hepatitis B and C were also excluded (n = 1,006).

*CT-derived hepatic fat quantification*

PMBB participants who had both CT-derived hepatic fat quantitation and WES available, were analyzed (n = 10,979). Liver fat was analyzed using a neuronal network and the techniques used are described elsewhere (Maclean *et al.* 2022).[13] Hepatic fat was quantitated in PMBB by subtracting the mean attenuation of all voxels contained within the liver from the mean attenuation of all voxels contained in the spleen, which are quantified by the spleen-liver Hounsfield Unit (spleen HU − liver HU). Values were identified by CT density determination to create a measure that is directly proportional to intrahepatic fat. Median, and maximum measurements of hepatic fat were recorded per individual given the multiple independent CT scans available per patient.

*NLP-derived imaging- and biopsy-proven steatosis cohorts*

In a study described elsewhere, natural language processing (NLP; Linguamatics) was used to interrogate 2.17 million radiology reports involving the liver in PMBB participants for positive mention of 'steatosis'.[14] Among PMBB participants with WES data, 2,865 cases of imaging-proven steatosis were identified and used for analyses. NLP was also used to interrogate 2.15 million pathology reports in PMBB participants for positive identification of hepatic steatosis biopsy results.[14] A total of 430 cases of biopsy-proven hepatic steatosis were identified, including steatotic liver disease (n = 224) and steatohepatitis (n = 119), which were used for analyses.

### United Kingdom Biobank

*Study population (replication cohort)*

The UKB is a large population scale study, which recruited 502,511 participants aged 37 to 73 at enrolment in 22 assessment centres across the United Kingdom. The study was conducted under UK Biobank access number 71300.

#### *Clinical data collection*

After enrolment during 2006 and 2010, participants underwent an initial examination, which was followed by a long-term follow up until January 2023. The baseline examination included blood sampling and physical examination, as well as socio-demographical and lifestyle data collection. As part of the enrolment all participants who were included in the biobank gave electronic signed consent for genotyping and data linkage to medical reports. Diagnoses were classified using the ICD-10 codes and inpatient hospital records from 1996 onward.

#### *Genotype data*

Genome-wide genetic data and analyses were available for 488,000 participants. For the genotype data the Haplotype Reference Consortium and UK10K were used.

### *Metabolomics data in the UKB*

In a subgroup of UKB participants, metabolomic profiling was performed with 168 normalized lipidomic parameters that were measured via nuclear magnetic resonance and normalized (n = 105,348). Logistic regression analyses were performed with Bonferroni-correction to account for multiple testing of major metabolic categories, using allele number as explanatory variable and metabolite levels as dependent variable ($p$ <0.05/168). Results (regression estimate, standard error and $p$ value) were graphically illustrated in a circle plot.

### *Hepatic fat quantification and cohort selection of steatotic liver disease phenotypes*

Liver MRI scans were performed according to a standardized protocol in the UKB imaging subgroup. The presence of steatosis was assessed based on MRI-derived proton density fat fraction (PDFF) measurements (data field 40061), which have been shown to be reliable and accurate for quantifying liver fat content.[15] Steatosis was defined as a PDFF >5% in accordance with established cut-off values for the presence of hepatic steatosis.[16] Fat accumulation on MRI was reported as a categorical outcome, with "1" indicating that a participant had PDFF >5% on MRI.

Amongst UKB participants with steatosis as defined by MRI-PDFF values, additional criterions utiziling alcohol consumption cut-offs were placed to identify MASLD and combined metabolic alcoholic liver disease (MetALD) utilizing AASLD guidelines.[17] In addition, we also leveraged "at-risk MASH" phenotypes construted in the UKB based on PDFF ≥5% and cT1 ≥875 ms.[18] Specific criterions for MASLD, MetALD and at-risk MASH are detailed in a previously pusbliehd study in the UKB population[19] and in the supplemental files.

### Analysis of additional phenotypes

A phenome-wide association study (PheWAS) in the PMBB and UKB was used to determine multiple phenotypes associated with selected variants in *TM6SF2* carried by participants. Phenotypes for each individual genotype were determined by mapping ICD-10 codes to distinct disease entities with methods described elsewhere.[13] Each disease phenotype was tested for association with *TM6SF2* variants using a logistic regression model adjusted for confounding factors: age, gender, and the first ten principal components of genetic ancestry. Bonferroni correction was used to adjust for multiple testing.

In addition, we evaluated cardiovascular ICD-10 codes of the identified *TM6SF2* variants in the PMBB given its known associations with lipids. The following ICD-10 codes were evaluated: Hypertensive heart disease (I11), acute myocardial infarction (I21), subsequent ST elevation (STEMI) and non-ST elevation (NSTEMI) myocardial infarction (I22), other acute ischemic heart disease (I24), chronic ischemic heart disease (I25) and atherosclerosis (I70). Logistic regression models were fitted to evaluate these associations, adjusted for age, gender, BMI and principal components of ancestry.

### *In silico* prediction of *TM6SF2* structure

*In silico* prediction of the structure of *TM6SF2* was generated using ColabFold Alphfold2 notebook with MMseqs2.[20–23] The amino acid sequence of wild-type (WT) *TM6SF2* were obtained from Uniprot[24] (accession number Q9BZW4). Visualization and analysis were performed using UCSF ChimeraX 1.4 (add ref PMID 32881101). Steatosis-associated non-synonymous missense variants were mapped on the structure for visualization based on their amino acid positions.

### Statistical analysis

Continuous variables are reported as mean ± standard deviation, while categorical variables were presented as relative frequencies (%). For age and BMI, a univariate unpaired two-tailed T-test was applied to test for differences. Other continuous variables such as serum parameters and CT-derived hepatic fat were fitted with a multivariate linear regression model. For sex and ethnicity, a univariate chi-square test was used while other categorical variables such as ICD-10 codes and imaging/biopsy reports were fitted with a multivariate binomial logistics regression model. Multivariate analyses were adjusted for age, sex, BMI, and principal components of ancestry 1-10 (PC1-10). Significant $p$ values from the binomial logistic regression are presented with corresponding odds ratios (ORs) and 95% CIs. A cut-off of $p$ <0.05 was deemed statistically significant for all analyses except for the PheWAS and metabolomics analysis, where Bonferroni correction thresholds were applied to adjust for multiple testing for PheCodes in the PheWAS and metabolites for the UKB metabolomics analysis. The following statistical programs were used to analyze our collected data: R version 4.0.2 (R Foundation for Statistical Computing; Vienna, Austria) and Prism version 8 (GraphPad, LaJolla, CA, USA). To create the graphical abstract and flow charts in Fig. 1, BioRender was used.

## Results

### *TM6SF2* missense variant selection

In PMBB participants, we filtered 121 non-synonymous missense variants in *TM6SF2* and selected variants that had over 40 carriers for variant-specific analyses (Table S1). Of the nine eligible variants, rs58542926 (E167K; minor allele frequency [MAF] 0.0126) and rs187429064 (L156P; MAF 0.0657) were significantly associated with increased risk of both ICD-diagnosed SLD and steatohepatitis in the PMBB ($p$ <0.001, Table 1). We also identified a rare variant rs186811910 (P216L; MAF 5.0e-04) in the PMBB to be associated with a significantly increased risk of MASH (Table 1). However, there were no significant associations with SLD or liver enzymes in P216L carriers (Tables S2–S3).

### Liver phenotypes of missense variants

The univariate PheWAS of *TM6SF2* E167K in PMBB identified Bonferroni-significant associations with the PheCodes of "chronic non-alcoholic liver disease" and "chronic liver disease and cirrhosis" (Bonferroni significance $p$ <0.0001, Fig. 2A). The L156P variant was nominally associated with "chronic non-alcoholic liver disease" ($p$ <0.001, Fig. 2B). The P216L variant was only associated with secondary thrombocytopenia (Bonferroni significance $p$ <0.0001, Fig. S1A).

Next, we interrogated ICD-10 codes and liver enzymes after multivariable adjustments. Amongst homozygote carriers of E167K, there was a significant increase in the risk of SLD,

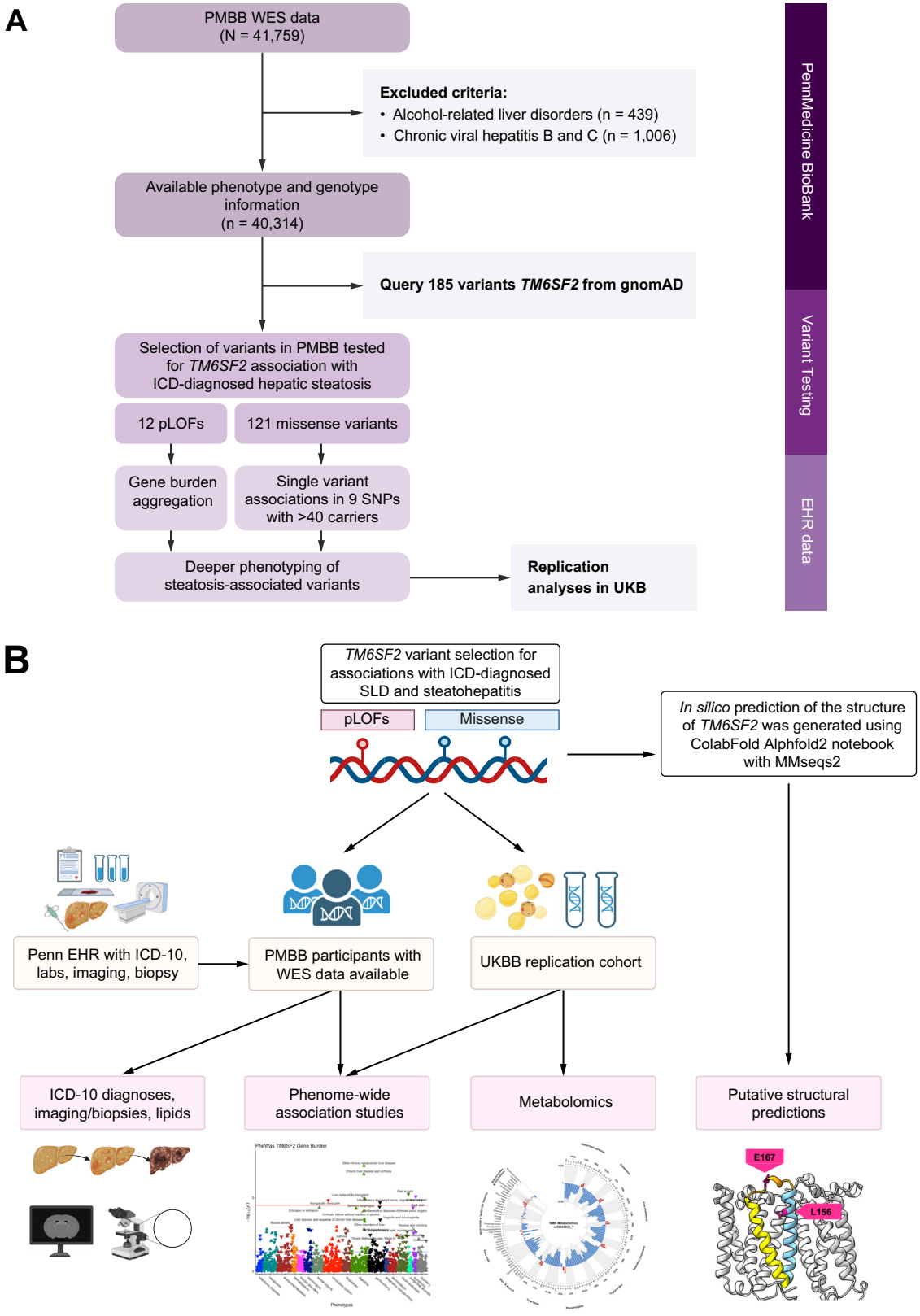



**Fig. 1. Analytical flowchart of study design.** (A) Exclusion criteria and variant selection in the PMBB discovery cohort. All non-synonymous missense variants and pLOFs were extracted from gnomAD to determine carriers in the PMBB with SNPs. The 12 pLOFs were aggregated into a gene burden and only 9 non-synonymous missense variants were identified based on the criteria that there must be over 40 carriers in the PMBB for adequate power. (B) Descriptive diagram of analyses conducted in the PMBB and UKB with structural predictions of steatosis-associated variants. The UKB was used as the replication cohort with additional insight in metabolomics data. Structural predictions of TM6SF2 were generated via ColabFold and only steatosis-associated missense variants that were identified were plotted. (p)LOF, (putative) loss of function; PMBB, Penn Medicine Biobank; SNP, single nucleotide polymorphism; UKB, UK Biobank.

**Table 1.** *TM6SF2* non-synonymous missense variants with >40 carriers in the PMBB and its associations with ICD-10 coded SLD and steatohepatitis.

| Variants (GChr38) | Predicted AA change | SNP | Carriers (n)* | MASLD *p* value | OR | 95% CI | | MASH *p* value | OR | 95% CI | | REVEL score | gnomAD MAF All |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 19:19269704:A:G | p.L156P | rs187429064 | 613 | **1.05e-05** | 2.18 | 1.54 | 3.07 | **3.58e-05** | 2.28 | 1.54 | 3.38 | 0.28 | 0.0657 |
| 19:19268740:C:T | p.E167K | rs58542926 | 4,540 | **1.08e-06** | 1.48 | 1.27 | 1.72 | **0.0006** | 1.39 | 1.15 | 1.68 | 0.075 | 0.0126 |
| 19:19268050:G:A | p.P216L | rs186811910 | 55 | 0.093 | 2.50 | 0.86 | 7.28 | **0.002** | 4.83 | 1.83 | 12.77 | 0.596 | 0.0005 |
| 19:19271076:C:A | p.V49L | rs200492531 | 156 | 0.63 | 0.79 | 2.9e-01 | 2.15 | 0.40 | 0.55 | 0.13 | 2.23 | 0.009 | 0.0012 |
| 19:19271072:G:A | p.A50V | rs184254301 | 129 | 0.97 | - | - | - | 0.48 | 1.44 | 0.52 | 4.00 | 0.063 | 0.0013 |
| 19:19269754:A:C | p.N139K | rs201189528 | 107 | 0.82 | 1.13 | 4.1e-01 | 3.11 | 0.74 | 1.22 | 0.38 | 3.89 | 0.022 | 0.0014 |
| 19:19270373:A:G | p.V90A | rs201830619 | 46 | 0.80 | 0.78 | 1.1eE-01 | 5.72 | 0.75 | 1.38 | 0.19 | 10.23 | 0.016 | 0.0001 |
| 19:19264774:G:T | p.L342M | rs188787025 | 113 | 0.50 | 1.38 | 5.5eE-01 | 3.43 | 0.79 | 1.17 | 0.37 | 3.74 | 0.029 | 0.0011 |
| 19:19264791:A:C | p.F336C | rs141184770 | 43 | 0.94 | - | - | - | 0.94 | - | - | - | 0.354 | 0.0005 |

AA, amino acid; MAF, minor allele frequency; OR, odds ratio; PMBB, Penn Medicine Biobank; REVEL, rare exome variant ensemble learner; SLD, steatotic liver disease; SNP, single nucleotide polymorphism.
Categorical/quantitative measures are expressed as number of participants (n) and relative frequencies (%), and fitted with a multivariable logistics regression model adjusted for age, sex, BMI, PC1-10. *All carriers were heterozygous except rs187429064 and rs58542926 carriers, which were calculated by combining the number of homozygotes and heterozygotes. Level of significance: $p$ <0.05 (Binomial multivariate logistics regression model). Values below the level of significance are bolded.

steatohepatitis, and HCC ($p$ <0.001, Table 2). E167K heterozygous carriers also had a significantly increased risk of SLD, fibrosis/cirrhosis and HCC ($p$ <0.05). After multivariable adjustment, individuals who were heterozygous for L156P were at significantly increased risk of SLD and steatohepatitis (Table 3). Though the sample size of L156P homozygotes was relatively small, there was a trend towards increased risk of HCC and fibrosis while the increased risks for SLD and steatohepatitis were nominal. Interestingly, the effect sizes were strongest in L156P homozygotes regarding liver diagnoses, suggesting that there is an apparent allele-dose trend in the risk of SLD and advanced fibrosis. When interrogating the liver biopsy cohort, we found that E167K heterozygotes had significantly increased risk of biopsy-proven steatohepatitis ($p$ = 0.03, Table 2). However, there were no associations with biopsy-proven SLD or steatohepatitis in L156P carriers, likely due to a low number of participants with biopsy results. In addition, E167K heterozygotes had higher serum ALT levels compared to non-carriers (Table 4).

### Additional metabolic risk factors and cardiovascular outcomes

We assessed the contribution of known genetic and metabolic risk factors on the observed clinical outcomes. To this end, we stratified carriers of E167K, L156P and non-carriers based on clinical SLD, steatohepatitis and HCC diagnoses, and we evaluated the frequency of obesity (BMI ≥30 kg/m$^2$), and homo and heterozygosity for the *PNPLA3* I148M risk allele in each subgroup (Fig. S3). The frequency of obesity (BMI ≥30 kg/m$^2$) was higher in homozygous E167K carriers who were diagnosed with clinical SLD, steatohepatitis and HCC, and in L156P heterozygotes carriers who were diagnosed with SLD and steatohepatitis (Fig. S3). Additionally, the frequency of homozygosity for the *PNPLA3* I148M risk allele was higher in E167K homozygotes diagnosed with SLD and steatohepatitis and in L156P heterozygotes diagnosed with HCC. These observations suggest that the *PNPLA3* I148M risk allele may have also contributed to the clinical outcomes. To determine the independent effects of E167K and L156P, we additionally adjusted for *PNPLA3* I148M as well as age, sex, BMI and PC1-10 in our regression analyses shown in Table S4 which demonstrated significant associations with liver disease independent of *PNPLA3* and obesity.

### Associations with imaging steatosis

We interrogated CT-derived hepatic fat quantification data of participants where available. Here, we found further evidence of increased risk of hepatic steatosis for both E167K and L156P variants. Homozygous and heterozygous E167K carriers showed significantly higher liver fat accumulation in an allele-dose manner compared to non-carriers (Table 2). The same trend was found in L156P heterozygotes, confirming that carriers of these variants exhibited higher hepatic fat content. This was further validated in the PMBB NLP imaging steatosis cohort amongst *TM6SF2* E167K and L156P carriers (Tables 2 and 3). There was an increased risk of NLP imaging-proven steatosis amongst all carriers of E167K and L156P heterozygotes compared with non-carriers. We also found that P216L heterozygotes were at an increased risk of imaging-proven steatosis (OR 2.52, $p$ <0.05, Table S2).

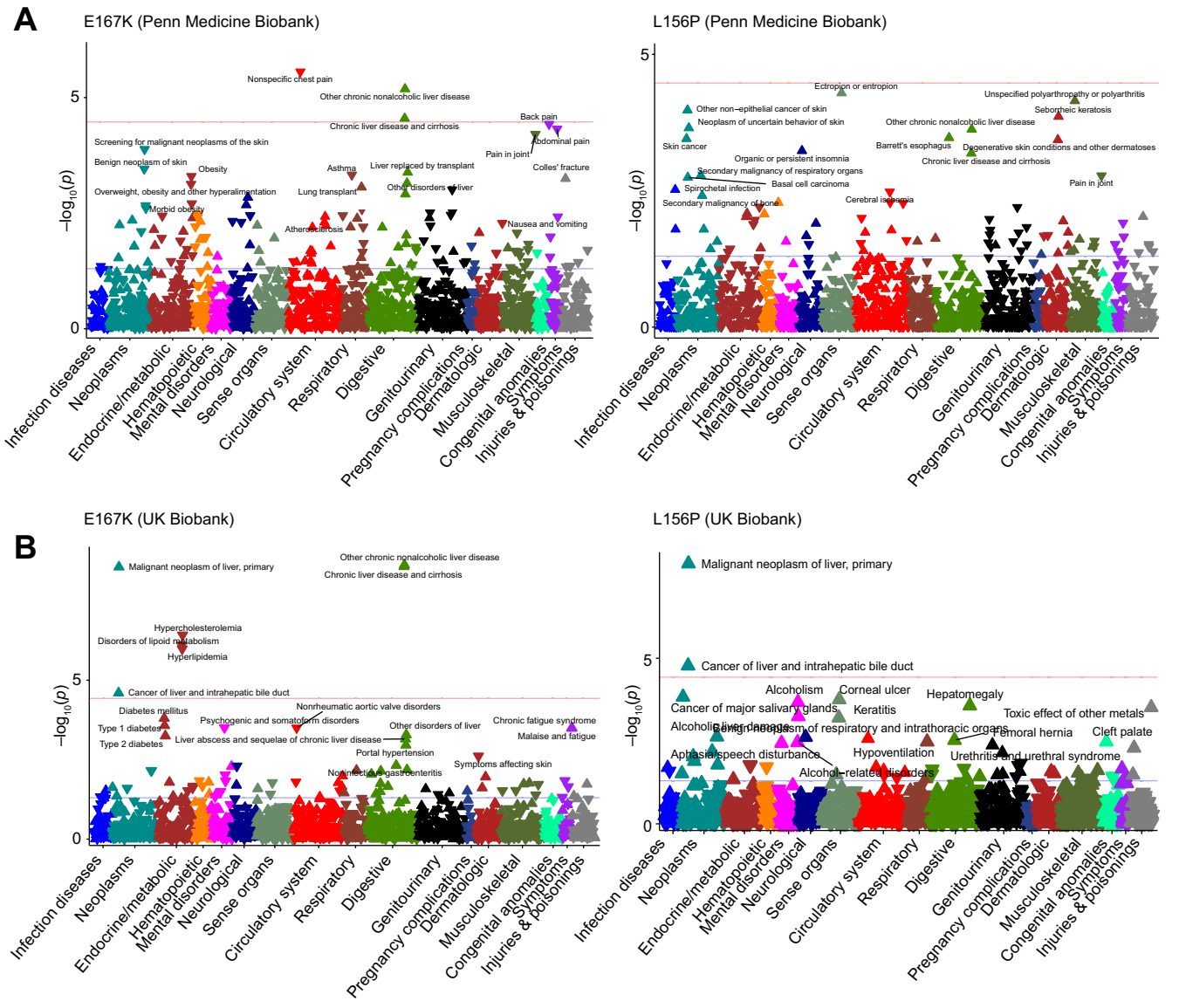



**Fig. 2. Univariate phenome-wide association studies of *TM6SF2* missense variants in PMBB and UKB.** The blue line represents a $p$ value of 0.05, and the red line represents the Bonferroni corrected significance threshold to adjust for multiple testing ($p \approx 0.05/1{,}800$). Level of significance = $p$ <0.05 (blue line), $p \approx 0.05/1{,}800$ (red line) using binomial logistic regression models. PMBB, Penn Medicine Biobank; UKB, UK Biobank.

### Identifying rare predicted loss of function variants in discovery cohort

Next, we screened the PMBB WES data for pLOF variants and identified 12 pLOF variants carried by 18 heterozygotes in the PMBB; two were stop-gain codons, four disrupted canonical splice site dinucleotides, and six were frame-shift mutations (Table S5). One participant in the PMBB carried two separate pLOF variants. A gene burden-association analysis revealed that pLOF variants were associated with a 5x higher risk of SLD ($p$ <0.01, OR 5.34, 95% CI 1.63-17.48) and imaging-proven steatosis compared to non-carriers (Table 5). A PheWAS analysis of pLOF carriers revealed nominal associations with liver abscess and sequelae of chronic liver disease and liver replaced by transplant phenotypes ($p$ <0.01, Fig. S4). There were no significant findings in relation to liver enzymes and lipids (Table S6). We identified that one very rare stop-gain variant W35X carried by four PMBB participants was the primary driver of associations with SLD amongst the pLOF gene burden (Table S7). A separate analysis in those four PMBB participants carrying W35X revealed an increased risk of ICD-diagnosed SLD, steatohepatitis, type 2 diabetes and fibrosis/cirrhosis compared to non-carriers ($p$ <0.05, Table S8). Additionally, W35X heterozygotes exhibited elevated hepatic fat scores and additional phenotypes in the PheWAS analysis related to sulphur-bearing amino-acid metabolism ($p$ <0.001, Fig. S5).

### Replication of hepatic steatosis phenotypes in the UKB

In the univariate and multivariable PheWAS, we noted that both E167K and L156P were associated with malignant neoplasms

**Table 2. Baseline characteristics and liver phenotypes of *TM6SF2* E167K homozygous/heterozygous carriers compared with non-carriers in the PMBB.**

| E167K Carriers | Homozygotes (T/T) n = 113 | Heterozygotes (C/T) n = 4,427 | Non-carriers (C/C) n = 35,774 | *p* value T/T *vs.* C/C | *p* value C/T *vs.* C/C | Odds ratio T/T *vs.* C/C | Odds ratio C/T *vs.* C/C |
|---|---|---|---|---|---|---|---|
| **Baseline characteristics** | | | | | | | **Univariate** |
| Age (years) | 56.66 ± 16.08 | 55.37 ± 16.69 | 55.10 ± 16.71 | 0.32 | 0.32 | - | - |
| Men (n, %) | 60 (53.1) | 2,272 (51.3) | 17,710 (49.5) | 0.42 | **0.009** | - | - |
| White ethnicity (n, %) | 96 (85.0) | 3,363 (76.1) | 24,067 (67.4) | **<2.2e-16** | **<2.2e-16** | - | - |
| Black ethnicity (n, %) | 4 (3.5) | 594 (13.4) | 8,708 (24.4) | **<2.2e-16** | **<2.2e-16** | - | - |
| BMI (kg/m$^2$) | 27.73 ± 5.27 | 28.91 ± 6.87 | 29.46 ± 7.23 | **0.015** | **3.5e-06** | - | - |
| *PNPLA3* rs738409:G (n, %)* | 0.57 (0.68) | 0.44 (0.59) | 0.43 (0.59) | **0.038** | 0.31 | - | - |
| **ICD-10 diagnoses (n, %)** | | | | | | | **Multivariate** |
| Type 2 diabetes (E11.00) | 0 (0.0) | 27 (0.6) | 283 (0.8) | 0.97 | 0.89 | 0.01 | 0.98 |
| Toxic liver disease (K71) | 0 (0.0) | 35 (0.8) | 164 (0.5) | 0.97 | 0.009 | 0.01 | 1.62 [1.13-2.32] |
| Hepatic failure (K72) | 1 (0.9) | 49 (1.1) | 420 (1.2) | 0.53 | 0.86 | 0.53 | 0.98 |
| Chronic hepatitis (K73) | 0 (0.0) | 2 (0.0) | 17 (0.0) | 0.99 | 0.59 | 9.49E-06 | 1.3 |
| Fibrosis and cirrhosis (K74) | 5 (4.4) | 58 (1.3) | 344 (1.0) | **0.02** | **0.02** | **2.99 [1.21-7.39]** | 1.3 [1.04-1.64] |
| Inflammatory liver disease (K75) | 0 (0.0) | 7 (0.2) | 64 (0.2) | **0.01** | **0.02** | **3.42 [1.82-6.44]** | 1.22 [1.04-1.43] |
| Steatohepatitis (K75.81) | 10 (8.8) | 116 (2.6) | 743 (2.1) | **2.68E-07** | **0.03** | **5.76 [2.96-11.21]** | 1.26 [1.03-1.55] |
| SLD (K76.0) | 13 (10.8) | 179 (4.0) | 1,113 (3.1) | **2.99E-08** | **3.5E-04** | **5.38 [2.97-9.74]** | 1.37 [1.15-1.62] |
| HCC (C22.0) | 2 (1.8) | 13 (0.3) | 64 (0.2) | **9.7E-04** | 0.06 | **11.22 [2.67-47.16]** | 1.79 |
| **Clinical imaging and biopsy data (n, %)** | | | | | | | **Multivariate** |
| CT-proven steatosis | 17 (15.0) | 361 (8.2) | 2,298 (6.4) | **6.28E-05** | **8.39E-07** | **3.04 [1.77-5.22]** | **1.36 [1.21-1.54]** |
| Biopsy-proven MASLD | 0 (0.0) | 17 (0.4) | 24.58 (43.49) | 0.97 | 0.67 | 8.09E-06 | 1.13 |
| Biopsy-proven MASH | 1 (0.9) | 21 (0.5) | 90 (0.3) | 0.17 | **0.02** | 4.00 | **1.8 [1.09-2.98]** |
| **Biomarkers of liver injury (n,%)** | | | | | | | **Multivariate** |
| Upper limits of AST | 7 (7.1) | 201 (5.6) | 1,445 (4.9) | 0.51 | 0.19 | 1.33 | 1.12 |
| Upper limits of ALT | 10 (10.2) | 294 (8.1) | 1,864 (6.3) | 0.15 | **0.001** | 1.69 | **1.26 [1.1-1.44]** |

ALT, alanine aminotransferase; AST, aspartate aminotransferase; HCC, hepatocellular carcinoma; PMBB, Penn Medicine Biobank; SLD, steatotic liver disease.
A univariate t-test was applied on age and BMI, whereas Chi-square test was used for gender, race and *PNPLA3* rs738409:G carriage. Categorical/quantitative measures are expressed as number of participants (n) and relative frequencies (%), and fitted with a multivariable logistics regression model adjusted for age, sex, BMI, PC1-10. Significant *p* values in multivariate analyses were reported with its corresponding odds ratios and confidence intervals of 95%. *2 = G/G Homozygotes, 1 = G/A heterozygotes for *PNPLA3*. Level of significance: *p* <0.05 (Chi square test for gender, race and *PNPLA3* rs738409:G carriage, univariate t-test for age and BMI, binomial multivariate logistics regression model for categorical outcomes). Values below the level of significance are bolded.

**Table 3. Baseline characteristics and liver phenotypes of L156P homozygous/heterozygous carriers compared with non-carriers in the PMBB.**

| L156P carriers | Homozygotes (G/G) n = 2 | Heterozygotes (A/G) n = 611 | Non-carriers (A/A) n = 39,701 | *p* value G/G *vs.* A/A | *p* value A/G *vs.* A/A | Adjusted odds ratio T/T *vs.* C/C | Adjusted odds ratio C/T *vs.* C/C |
|---|---|---|---|---|---|---|---|
| **Baseline characteristics** | | | | | | | **Univariate** |
| Age (years) | 54.52 ± 17.78 | 56.74 ± 15.72 | 55.04 ± 16.57 | 0.96 | **0.017** | - | - |
| Men (n, %) | 1 (50.0) | 313 (51.2) | 20,698 (50.3) | 0.99 | 0.31 | - | - |
| White ethnicity (n, %) | 2 (100.0) | 539 (88.4) | 27,699 (67.5) | **<2.2e-16** | **<2.2e-16** | - | - |
| Black ethnicity (n, %) | 0 (0.0) | 37 (6.1) | 9,867 (24.0) | **<2.2e-16** | **<2.2e-16** | - | - |
| BMI (kg/m$^2$) | 38.50 ± 2.12 | 28.57 ± 6.53 | 29.40 ± 7.17 | 0.073 | **0.0068** | - | - |
| *PNPLA3* rs738409:G (n, %)* | 0.50 (0.50) | 0.46 (0.63) | 0.43 (0.59) | 0.98 | 0.47 | - | - |
| **ICD-10 diagnoses (n, %)** | | | | | | | **Multivariate** |
| Type 2 diabetes (E11.00) | 0 (0.0) | 2 (0.3) | 308 (0.8) | 0.98 | 0.65 | 0.01 | 0.73 |
| Toxic liver disease (K71) | 0 (0.0) | 5 (0.8) | 194 (0.5) | 0.98 | 0.3 | 0.01 | 1.62 |
| Hepatic failure (K72) | 0 (0.0) | 6 (1.0) | 464 (1.2) | 0.98 | 0.81 | 0.01 | 0.93 |
| Chronic hepatitis (K73) | 0 (0.0) | 0 (0.0) | 19 (0.0) | 1 | 0.99 | 0.01 | 0.01 |
| Fibrosis and cirrhosis (K74) | 1 (50.0) | 7 (1.1) | 399 (1.0) | **0.01** | 0.2 | **52.36 [3.22-853.38]** | 1.43 |
| Inflammatory liver disease (K75) | 0 (0.0) | 3 (0.5) | 68 (0.2) | **0.04** | 0.01 | **18.41 [1.12-302.94]** | **1.89 [1.35-2.66]** |
| Steatohepatitis (K75.81) | 1 (50.0) | 26 (4.3) | 842 (2.1) | **0.03** | 8.86E-05 | **29.1 [1.71-497.84]** | **2.26 [1.51-3.39]** |
| SLD (K76.0) | 1 (50.0) | 36 (5.9) | 1,268 (3.2) | **0.04** | 2.69E-05 | **19.01 [1.15-315.49]** | **2.14 [1.51-3.06]** |
| HCC (C22.0) | 1 (50.0) | 2 (0.3) | 76 (0.2) | **0.01** | 0.39 | **480.57 [20.35-11,348.59]** | 1.88 |
| **Clinical imaging and biopsy data from EHR** | | | | | | | **Multivariate** |
| CT-proven steatosis | 0 (0.0) | 63 (10.3) | 2,613 (6.6) | 0.95 | **3.46E-04** | 0.01 | 1.68 [1.27-2.23] |
| Biopsy-proven MASLD | 0 (0.0) | 2 (0.3) | 148 (0.4) | 0.99 | 0.94 | 0.01 | 1.07 |
| Biopsy-proven MASH | 0 (0.0) | 0 (0.0) | 112 (0.3) | 1.00 | 0.98 | 0.01 | 0.01 |
| **Biomarkers of liver injury** | | | | | | | **Multivariate** |
| Upper limits of AST | 1 (50.0) | 30 (6.0) | 1,622 (5.0) | **0.04** | 0.94 | 21.18 [1.32-341.45] | 1.17 |
| Upper limits of ALT | 1 (50.0) | 44 (8.8) | 2,123 (6.5) | 0.07 | 0.98 | 13.34 | 1.33 |

ALT, alanine aminotransferase; AST, aspartate aminotransferase; HCC, hepatocellular carcinoma; MASLD, metabolic dysfunction-associated steatotic liver disease; PMBB, Penn Medicine Biobank; SLD, steatotic liver disease.
A univariate t-test was applied on age and BMI, whereas Chi-square test was used for gender. Categorical/quantitative measures are expressed as number of participants (n) and relative frequencies (%), and fitted with a multivariable logistics regression model adjusted for age, sex, BMI, PC1-10. Significant p-values were reported with its corresponding odds ratios and confidence intervals of 95%. *2 = G/G [Homozygotes, 1 = G/A heterozygotes for *PNPLA3.* Level of significance: *p* <0.05 (Chi square test for gender, race and *PNPLA3* rs738409:G carriage, univariate t-test for age and BMI, binomial multivariate logistics regression model for categorical outcomes). Values below the level of significance are bolded.

Table 4. Liver enzymes, fat retention, plasma lipids and glucose parameters in homo- and heterozygote E167K and L156P carriers in comparison with non-carriers in PMBB.

| | E167K | | | | | | L156P | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Carriers | Homozygotes (T/T) n = 113 | Heterozygotes (C/T) n = 4,427 | Non-carriers (C/C) n = 35,774 | *p* value T/T *vs.* C/C | *p* value C/T *vs.* C/C | Homozygotes (G/G) n = 2 | Heterozygotes (A/G) n = 611 | Non-carriers (A/A) n = 39,701 | *p* value G/G *vs.* A/A | *p* value A/G *vs.* A/A |
| **Liver status** | | | | | | | | | | |
| CT hepatic fat (HU) | 1.67 ± 13.20 | -4.99 ± 10.09 | -6.90 ± 8.87 | 9.68E-09 | 7.75E-11 | -16.10 (N/A) | -4.76 ± 10.18 | -6.69 ± 9.04 | 0.18 | 0.01 |
| AST (U/L) | 24.91 ± 9.30 | 25.39 ± 30.28 | 24.68 ± 40.35 | 0.87 | 0.53 | 60.00 ± 42.43 | 26.56 ± 34.33 | 24.73 ± 39.39 | 0.18 | 0.37 |
| ALT (U/L) | 26.34 ± 16.57 | 26.27 ± 43.08 | 24.58 ± 43.49 | 0.8 | 0.06 | 40.50 ± 0.71 | 27.46 ± 35.14 | 24.72 ± 43.50 | 0.59 | 0.26 |
| ALP (U/L) | 71.92 ± 31.52 | 74.60 ± 41.16 | 76.16 ± 41.60 | 0.79 | 0.4 | 108.00 ± 55.15 | 73.58 ± 46.49 | 76.01 ± 41.45 | 0.32 | 0.5 |
| GGT (U/L) | 60.81 ± 53.54 | 102.00 ± 132.74 | 120.13 ± 252.08 | 0.55 | 0.45 | NA | 118.62 ± 232.52 | 117.81 ± 241.29 | N/A | 0.79 |
| **Lipid metabolism** | | | | | | | | | | |
| LDL (mg/dl) | 91.39 ± 30.59 | 97.65 ± 32.59 | 99.32 ± 32.52 | 0.12 | 0.31 | 99.00 ± 52.33 | 95.41 ± 27.26 | 99.18 ± 32.59 | 1.00 | 0.14 |
| HDL (mg/dl) | 53.25 ± 22.79 | 51.20 ± 15.60 | 51.37 ± 15.55 | 0.31 | 0.46 | 40.00 ± 2.83 | 52.45 ± 16.45 | 51.34 ± 15.57 | 0.49 | 0.54 |
| TG (mg/dl) | 110.89 ± 58.26 | 118.92 ± 71.41 | 121.63 ± 83.82 | 0.17 | **0.001** | 81.00 ± 28.28 | 105.08 ± 49.07 | 121.56 ± 82.89 | 0.35 | 3.73E-06 |
| TC (mg/dl) | 166.68 ± 41.64 | 172.98 ± 39.42 | 175.31 ± 38.95 | 0.12 | **0.03** | 158.50 ± 58.69 | 169.54 ± 34.76 | 175.12 ± 39.07 | 0.58 | 0.01 |
| **Glucose parameters** | | | | | | | | | | |
| HbA1C (%) | 6.70 ± 1.75 | 6.28 ± 1.35 | 6.30 ± 2.03 | 0.09 | 0.51 | NA | 6.12 ± 1.03 | 6.30 ± 1.98 | 0.5 | 0.93 |
| Fasting glucose (mmol/L) | 138.33 ± 101.16 | 105.91 ± 40.34 | 96.91 ± 33.65 | 0.04 | 0.06 | NA | 95.86 ± 22.37 | 97.92 ± 35.03 | N/A | 0.77 |
| Random glucose (mmol/L) | 122.77 ± 46.57 | 110.02 ± 37.48 | 108.53 ± 35.40 | 0.01 | 0.01 | 86.75 ± 6.01 | 108.97 ± 34.62 | 108.73 ± 35.69 | 0.24 | 0.42 |

ALP, alkaline phosphatase; ALT, alanine aminotransferases; AST, aspartate aminotransferase; GGT, gamma-glutamyltransferase; HDL, high-density lipoprotein; HbA1C, glycated hemoglobin; HU, Hounsfield units; LDL, low-density lipoprotein; PMBB, Penn Medicine Biobank.
Multivariable analysis of serum parameters were adjusted for age, sex, BMI, and PC1-10. All continuous variables were fitted with linear regression models and expressed as means and standard deviations. Level of significance: $p$ <0.05 (generalized linear regression model). Values below the level of significance are bolded.

of the liver and cancer of the liver/intrahepatic bile ducts (Bonferroni significance $p$ <0.0001, Fig. 2). Specific to E167K carriers, PheWAS analyses in both cohorts revealed significant associations with SLD, liver abscess and sequelae of chronic liver disease, portal hypertension, and alcohol-related liver damage (Bonferroni significance $p$ <0.0001, Fig. S6). E167K and L156P carriers also exhibited higher levels of liver enzymes (Table S9). The significant association with an increased risk of ICD-diagnosed SLD, steatohepatitis and fibrosis/cirrhosis was only replicated in E167K carriers while increased HCC risk was seen for carriers of both E167K and L156P (Table S10).

We also investigated 266 carriers of the P216L variant in the UKB and did not find a significant association with SLD identified through ICD-10 codes and imaging reports (Fig. S6; Table S11). To characterize loss-of-function variants in the UKB, we interrogated a gene burden association with SLD (K76.0) using Genebass, an online platform which uses SAIGE-GENE to perform gene-burden tests and SKAT-O[25] (Table S12). Here, we confirmed associations with an elevated risk of liver disease. However, the rare stop-gain variant W35X was found in only seven UKB participants, and was associated with additional phenotypes such as dysthymic disorder reported in Fig. S7. None of the UKB participants carrying W35X were ICD-diagnosed with liver-related phenotypes or exhibited elevated levels of aminotransferases.

### Alcohol-related and metabolic-associated liver disease spectrum

Based on the newly revised nomenclature defining the spectrum of conditions under SLD, we leveraged UKB-specific data on MRI reports with PDFF values and utilized different diagnostic cut-offs to capture novel phenotypes in the European population. Here, we reported carriers of E167K and L156P to exhibit a higher risk of hepatic steatosis quantified as >5% accumulation of fat on MRI (Table S13). After stratifying carriers by alcohol consumption, we demonstrated an almost 2-fold elevated risk of MASLD and MetALD in carriers *vs*. non-carriers. In addition, we identified that E167K carriers have a 3-fold increased risk of at-risk MASH compared to non-carriers. In previous studies, the prevalence of at-risk MASH in the general UK population was less than 2%.[19]

### Plasma lipids and glucose parameters amongst *TM6SF2* variant carriers

In PMBB, heterozygotes of E167K and L156P had significantly lower circulating triglycerides and total cholesterol compared to non-carriers (Table 4). As such UKB E167K carriers were significantly less likely to have ICD-10 diagnoses of hyperlipidemia or hypercholesterolemia ($p$ <0.0001, Fig. 2). Both E167K and L156P carriers exhibited reduced circulating LDL-cholesterol, triglycerides, and total cholesterol consistent with PMBB carriers (Table S9). In addition, there were significant associations with decreased apolipoprotein B levels compared to non-carriers in both variants. E167K carriers exhibited evidence of insulin resistance due to significantly higher random and fasting glucose levels compared to non-carriers in an allele-dose manner ($p$ <0.05, Table 4). However, the frequencies of ICD-diagnosed type 2 diabetes mellitus and in the levels of HbA1c were not significantly higher in E167K carriers. Amongst

**Table 5. Gene-burden association analysis of TM6SF2 pLOF variants with demographics characteristics and liver disease diagnoses (PMBB).**

| Carriers | *TM6SF2*$^{-/-}$ carrier of two pLOF variants (n = 1) | *TM6SF2*$^{+/-}$ carriers of one pLOF variant (n = 18) | *TM6SF2*$^{+/+}$ non-carriers of pLOF (n = 40,289) | | *p* values *TM6SF2* pLOF carriers *vs.* non-carriers |
|---|---|---|---|---|---|
| **Baseline characteristics** | | | | | **Univariate** |
| Age (years) | 67.14 [67.14, 67.14] | 53.56 [36.31, 58.79] | 57.45 [42.46, 67.71] | | 0.352 |
| Men (n, %) | 0 (0.0) | 2 (11.1) | 20,036 (49.7) | | **0.003** |
| White ethnicity (n, %) | 1 (100.0) | 12 (66.7) | 27,512 (68.4) | | 1.000 |
| BMI (kg/m$^2$) | 29.00 [29.00, 29.00] | 28.00 [24.00, 35.00] | 28.00 [24.00, 33.00] | | 0.943 |
| **ICD-10 diagnoses (n, %)** | | | | **Univariate (Chi-square)** | **Multivariate (logistic regression model)** |
| Type 2 diabetes (E11.00) | 0 (0.0) | 1 (5.6) | 309 (0.8) | 0.067 | 0.07 |
| Toxic liver disease (K71) | 0 (0.0) | 0 (0.0) | 199 (0.5) | 0.954 | 0.98 |
| Hepatic failure (K72) | 0 (0.0) | 0 (0.0) | 470 (1.2) | 0.894 | 0.96 |
| Chronic hepatitis (K73) | 0 (0.0) | 0 (0.0) | 19 (0.0) | 0.996 | 0.99 |
| Fibrosis and cirrhosis (K74) | 0 (0.0) | 1 (5.6) | 406 (1.0) | 0.155 | 0.23 |
| Inflammatory liver disease (K75) | 0 (0.0) | 0 (0.0) | 71 (0.2) | 0.983 | 0.67 |
| Steatohepatitis (K75.81) | 0 (0.0) | 1 (5.6) | 868 (2.2) | 0.604 | 0.35 |
| SLD (K76.0) | 1 (100.0) | 1 (5.6) | 1,303 (3.2) | **<0.001** | **0.01** OR:4.9, 95%CI [1.51-15.91] |
| HCC (C22.0) | 0 (0.0) | 0 (0.0) | 79 (0.2) | 0.982 | 0.99 |
| **Clinical imaging and biopsy data from EHR** | | | | | |
| CT-proven steatosis | 1 (100.0) | 1 (5.6) | 2,674 (6.6) | **0.001** | 0.11 |
| Biopsy-proven SLD | 0 (0.0) | 0 (0.0) | 150 (0.4) | 0.965 | 0.98 |
| Biopsy-proven steatohepatitis | 0 (0.0) | 0 (0.0) | 112 (0.3) | 0.974 | 0.98 |
| **Biomarkers of liver injury (n,%)** | | | | | |
| Upper limits of AST | 0 (0.0) | 1 (7.1) | 1,652 (5.0) | 0.911 | 0.95 |
| Upper limits of ALT | 1 (100.0) | 2 (14.3) | 2,164 (6.5) | **<0.001** | **0.02** OR 3.76, 95% CI 1.28-11.04 |

ALP, alkaline phosphatase; ALT, alanine aminotransferases; AST, aspartate aminotransferase; EHR, electronic health record; GGT, gamma-glutamyl transferase; HbA1C, glycated hemoglobin; HCC, hepatocellular carcinoma; HDL, high-density lipoprotein; LDL, low-density lipoprotein; pLOF, putative loss-of-function; PMBB, Penn Medicine Biobank; SLD, steatotic liver disease; TC, total cholesterol; TG, triglycerides.

Continuous data is represented as median and interquartile range due to skewed distribution of data points, whereas categorical/quantitative measures are expressed as number of participants (n) and relative frequencies (%). A corresponding *p* value with univariate tests were done using a Chi-square test for categorical variables and a Kruskal-Wallis H test was performed for continuous variables. A robust logistics/linear model was fitted for multivariate *p* values adjusted for age, sex, BMI, and PC 1-10. Significant *p* values for categorical outcomes were reported with its corresponding odds ratios and confidence intervals of 95%. Level of significance: $p < 0.05$ (Chi square test for gender, age and univariate categorical outcomes, univariate t-test for age and BMI, binomial multivariate logistics regression model for categorical outcomes). Values below the level of significance are bolded.

rarer variants, such as P216L and W35X, we observed lower mean LDL and total cholesterol but these differences were not statistically significant (Tables S3 and S8).

We further interrogated data on 168 serum metabolites measured by nuclear magnetic resonance in a subset of patients from the UKB (n = 105,348). Of the patients with available metabolomics data, there were 1,118 homozygotes and 27,677 heterozygotes carrying the E167K variant, while 35 homozygotes and 5,136 heterozygotes carried L156P. E167K carriers exhibited a significantly lower number of LDL and VLDL particles, and lower levels of cholesterol, and phospholipids while the number of large-sized HDL particles was higher compared to non-carriers (all Bonferroni corrected $p < 0.001$, Fig. 3A). The majority of fatty acids (*i.e.* polyunsaturated, omega-3) and total choline's were also significantly lower. For the L156P variant, carriers had significantly lower polyunsaturated fatty acid, glycine, saturated fatty acid, and free cholesterol in small VLDL, compared to non-carriers (all Bonferroni corrected $p < 0.001$, Fig. 3B). The metabolomics analysis for P216L and W35X did not yield significant associations due to low power (Figs. S8–S9). However, the trends were consistent with the plasma lipid profile observed in carriers of E167K and L156P.

### Associations with cardiovascular disease

Cardiovascular ICD-10 codes were interrogated to determine if CVD risk was associated with E167K, L156P, P216L and pLOF carriers in the PMBB (Tables S14–S15). Intriguingly, ICD-diagnosed atherosclerosis was significantly associated with E167K and L156P heterozygosity, but the risk was increased compared to non-carriers (C/T *vs.* C/C: $p = 0.017$, OR 1.18, 95% CI 1.03-1.35; A/G *vs.* G/G: $p < 0.002$, OR 1.60, 95% CI 1.18-2.17). Additionally, E167K homozygotes were at a higher risk of hypertensive heart disease ($p = 0.005$, OR 3.07, 95% CI 1.41-6.67). Individuals carrying at least one copy of a rare *TM6SF2* pLOF variant exhibited an increased risk for chronic ischemic heart disease though underpowered ($p = 0.03$, OR 19.02, 95% CI 1.3-279.7).

### *In silico* predictions of steatosis-associated *TM6SF2* variants induce loss of function

We used AlphaFold-generated *in silico* prediction models to explore the structure of WT *TM6SF2* and to determine the potential structural consequences of the E167K and L156P substitutions. Consistent with previous reports,[26] the AlphaFold structure shows that *TM6SF2* has a helix-loop-helix structure, with 10 transmembrane segments (Fig. 4A and Fig. S10). The loops facing the luminal side of the endoplasmic reticulum (ER), collectively form the highly conserved EXPanded EBP superfamily (EXPERA) domain, characterized by a series of negatively charged amino acids.[26] L156P introduces a proline residue into the amino acid sequence of the 5$^{th}$ transmembrane α-helix (Fig. 4B). In the WT protein, the amide hydrogen of L156 hydrogen bonds to the carbonyl group

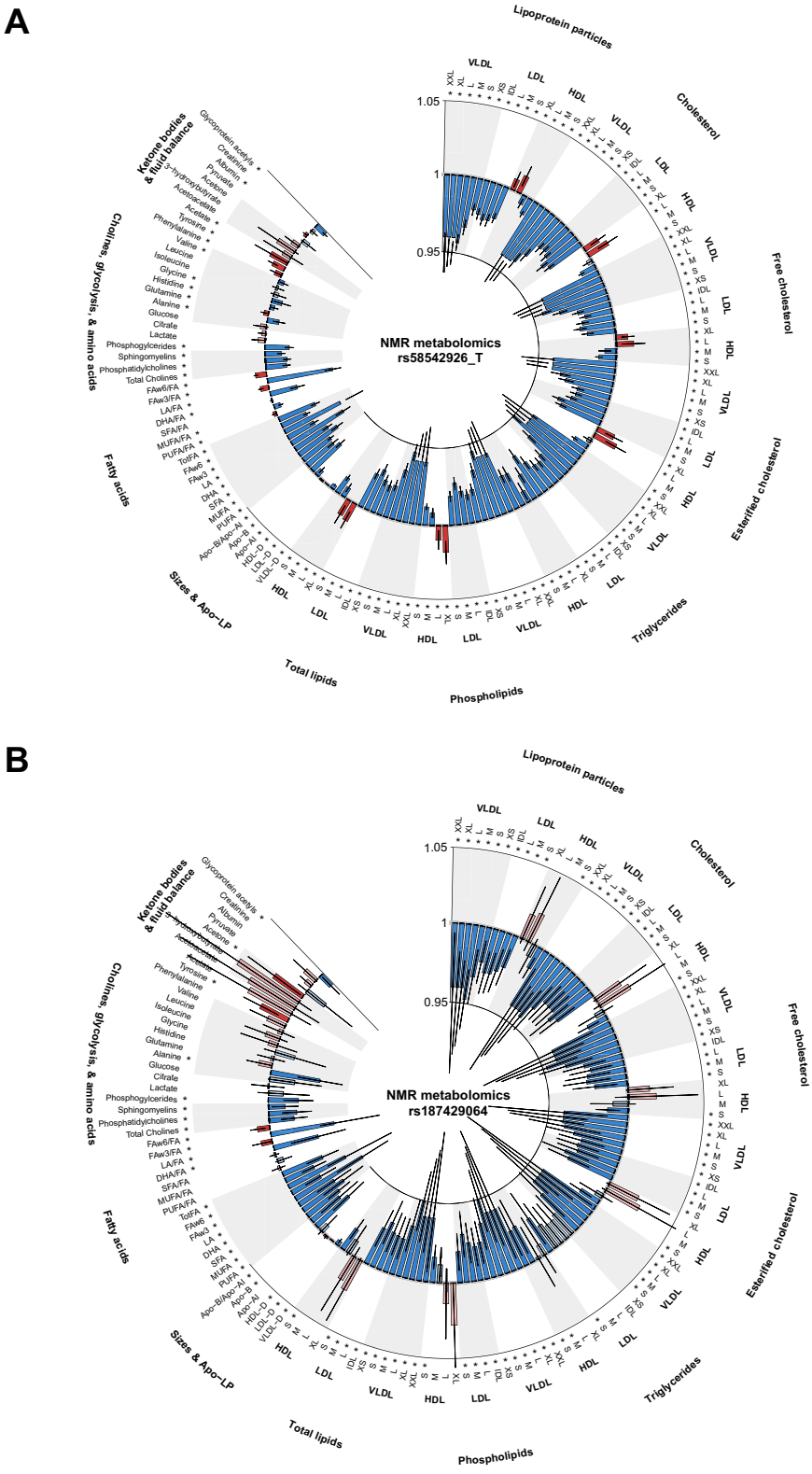



**Fig. 3. Circo plot of -log$_{10}$ (*p* value) metabolomics analysis in the UKB for E167K (rs58542926_T) and L156P (rs187429064).** 168 normalized lipidomic parameters (outer circle) were measured via NMR with Bonferroni-correction to account for multiple testing of major metabolic categofries ($p$ <0.05/168). Box and whisker plots were shown for each metabolite. Red boxes demonstrate increased levels (>1), while blue boxes demonstrate decreased levels (<1) per *TM6SF2* allele. Bars were bolded in color if the $p$ value for the association met Bonferroni-significance ($p$ <0.05/168 tests). Level of significance: $p$ <0.05/168 (binomial logistics regression model).

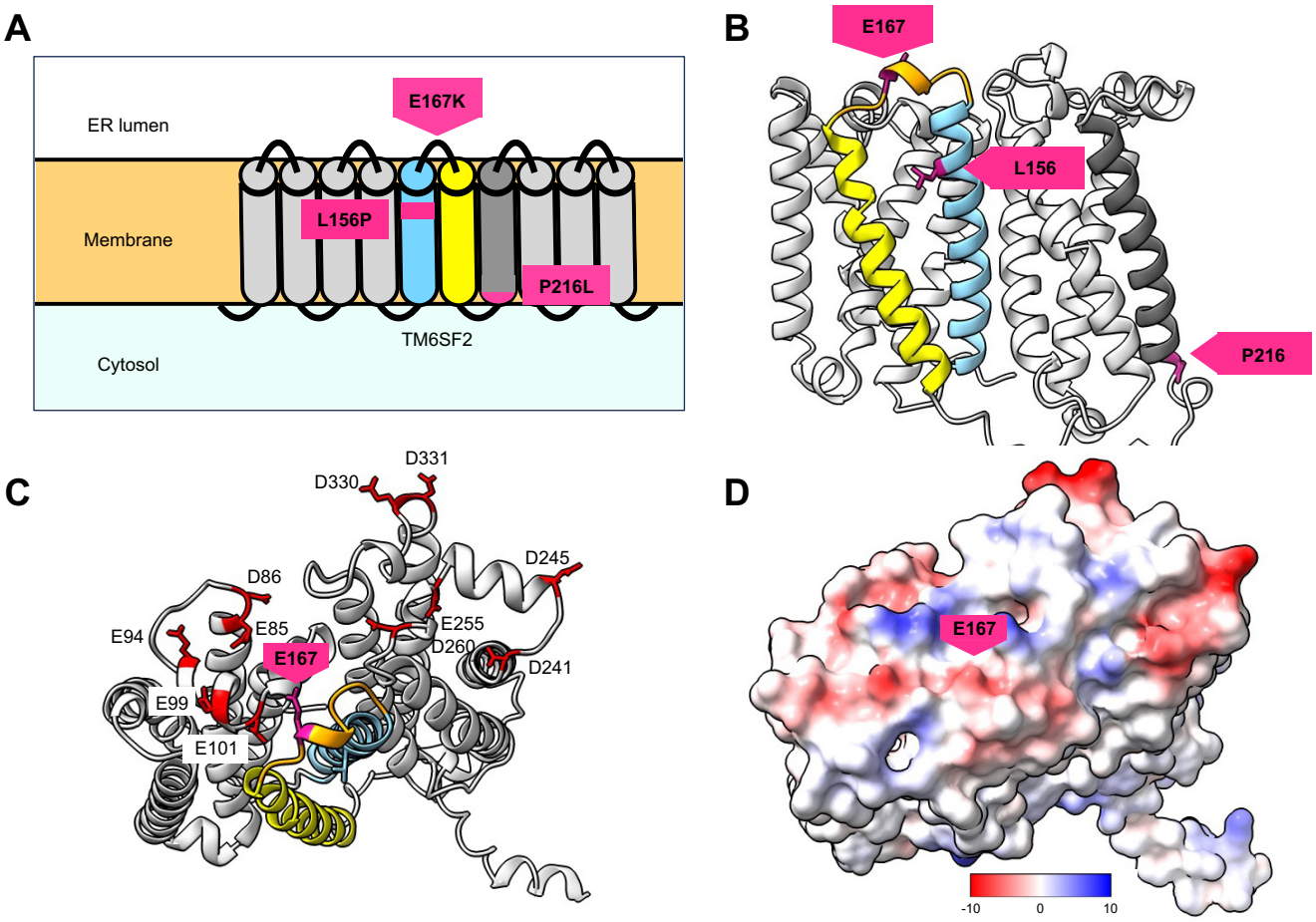



**Fig. 4. Mapping of mutated amino acid residues from non-synonymous missense variants onto WT TM6SF2 structure.** (A) Schematic representation of TM6SF2 structure, showing 10 transmembrane helical segments. Helix 5 is highlighted in light blue, helix 6 in yellow and helix 7 in dark grey. Location of E167K, L156P and P216L are indicated. (B-D) *In silico* predicted structures of TM6SF2 showing the locations of L156, E167 and P216. Predictions were generated using ColabFold Alphafold2 notebook with MMseqs2.[20–23] In all the structures, helix 5 is coloured in light blue, helix 6 in yellow, helix 7 in dark grey and the loop connecting helix 5-6 in orange. (C) Structure of the luminal interface of TM6SF2. Negatively charged amino acids located within the interhelical loops are coloured in red, E167 is coloured in magenta. (D) Surface map representation of coulombic charge for the structure shown in C. Red indicates negatively charged amino acids, blue positively charged and white neutral amino acids. Confidence plots for the TM6SF2 structure are reported in Fig. S10. WT, wild-type; ER, endoplasmic reticulumn.

of Ile152. Due to its lack of an amide hydrogen, P156 cannot create such a hydrogen bond and likely introduces a ~20° kink in the α-helix at V154 with the proline sidechain on the outside of the kink.[27] This structural change is expected to alter the position of the C-terminal residue in helix 5 (L161) and ultimately alter the topology of the residues G162-R169 that form the extracellular loop containing E167. Thus, both E167K and L156P likely impact the structural architecture of *TM6SF2* and disrupt the EXPERA domain (Fig. 4C,D).

For the rarer missense variants, we identified that P216L is located at the beginning of helix 7 (Fig. 4A,B). Prolines occur frequently at the first N-terminus turn of a helix, especially in transmembrane proteins.[28] In such positions, proline-mediated kinks often mark the intersection between loop sequences and transmembrane helical segments, thus promoting the proper protein packing. Therefore, we hypothesized that a lack of proline could potentially explain the LOF effects for P216L given its similar directionality with common missense variants in relation to steatosis.

## Discussion

Our findings represent one of the first genome-first analyses of *TM6SF2* variants and characterize metabolic risk factors in

168 normalized lipidomic parameters (outer circle) were measured via NMR with Bonferroni-correction to account for multiple testing of major metabolic categories ($p$ <0.05/168). Box and whisker plots were shown for each metabolite. Red boxes demonstrate increased levels (>1), while blue boxes demonstrate decreased levels (<1) per *TM6SF2* allele. Bars were bolded in color if the $p$ value for the association met Bonferroni-significance ($p$ <0.05/168 tests). Level of significance: $p$ <0.05/168 (binomial logistics regression model). ApoA-I, apolipoprotein A-I; ApoB, apolipoprotein B; DHA, docosahexaenoic acid; FA, fatty acids; HDL, high-density lipoprotein; LA, linoleic acid; LDL, low-density lipoprotein; L, large; M, medium; MUFA, monounsaturated fatty acids; NMR, nuclear magnetic resonance; PUFA, polyunsaturated fatty acids; S, small; SFA, saturated fatty acids; UKB, UK Biobank; VLDL, very-low-density lipoprotein; XL, very large; XS, extra small; XXL, extremely large. The original code was provided by by Diego J Aguilar-Ramirez and adjusted by Jan Clusmann.[74]

protein-altering variants, such as non-synonymous missense single nucleotide polymorphisms (SNPs) and pLOFs. Our findings present a holistic understanding of how *TM6SF2* works biologically and the clinical implications of variants across multiple phenotypic traits.

As previously reported, E167K is the most commonly described coding variant in *TM6SF2* associated with SLD and steatohepatitis in exome-wide clinical studies and hepatoma cell lines.[29] The variant is consistent with a loss of protein function, as the mutation E167K substitutes negatively charged glutamic acid at residue 167 with positively charged lysine, thereby disrupting the electrostatic properties of the EXPERA domain.[26] The associations with fibrosis/cirrhosis were most prominent in E167K carriers, consistent with one of the first studies reporting its influence in hepatic fibrosis progression amongst patients with SLD.[30–32] Subsequently, *TM6SF2* E167K T/C polymorphisms were previously observed to increase the risk of HCC as well as alcohol-related cirrhosis.[33–36] This is likely due to fat accumulation in the liver from *TM6SF2* deletion in combination with environmental stressors such as alcohol, which would increase oxidative stress and inflammation in hepatocytes leading to much more progressive phenotypes.[37] With the newly revised nomenclature defining SLD, we observe that the E167K variant of *TM6SF2* increased the risk of at-risk MASH by 3-fold compared to in non-carriers. The prevalence of this demographic is estimated to be approximately 0.6% in the population while MASLD is present in nearly 90% of those with hepatic steatosis.[19] Despite this, patients with at-risk MASH seemingly exhibit more prominent inflammatory and metabolic phenotypes that likely explain the propensity to exhibit higher fibrosis stages (F <2) and therefore are at a higher risk of morbidity.[38] Future studies may benefit from understanding the extent to which genetic risk would influence liver-related mortality in those with at-risk MASH to understand how therapeutic targets from recently approved clinical trials would benefit those with concurrent risk variants of *TM6SF2*.[39]

Although the development of SLD-HCC attributed to *TM6SF2* E167K remains disputed, in large part due to differences in studies reporting on alcohol-related and viral cirrhosis, our results support previous reports demonstrating its independent effects on HCC development.[40,41] In addition, our study strongly suggests that the L156P substitution exhibits similar directionality in increasing the risk of SLD-associated phenotypes, independent of hepatitis B/C and *PNPLA3* I148M, as demonstrated by imaging and clinical diagnoses. As well, the effect size of HCC was the highest in L156P mutation carriers (*TM6SF2*) compared to carriers of other common variants, such as *TM6SF2* E167K and *MARC1* rs2642438:A.[42] As such, one would expect to find strongly replicated HCC phenotypes across two biobanks in L156P carriers. However, we suggest that perhaps both E167K and L156P are associated with SLD-HCC progression in the absence of other causes of cirrhosis, as confirmed across two academic biobanks. Combined, this suggests that carriers may need to be followed-up for carcinogenic progression as well as investigating the impact of these variants on liver-related mortality.

Insights on plasma lipid phenotypes associated with *TM6SF2* E167K and L156P provide further clarification on the physiological mechanisms that cause steatosis.[43,44] *TM6SF2* is located on chromosome 19 and stimulates the biosynthesis of cholesterol for subsequent assembly into VLDL.[26,45,46] The gene contains a conserved catalytic site for 3-β-hydroxysteroid-8,7-isomerase, an enzyme which converts the double bond at the 8 position to 7 in sterols, which is necessary for the synthesis of cholesterol.[20] The consequence of liver damage due to *TM6SF2* knock-out may be explained by two mechanisms: abnormal triglyceride synthesis in the liver or a lack of VLDL secretion.[30,47] Early murine models found that loss of *TM6SF2* did not affect sterol responsive element binding proteins, decreasing the plausibility of increased synthesis of triglycerides as an underlying mechanism.[48] Rather, previous studies support the hypothesis that *TM6SF2* knock-out ameliorates circulating plasma cholesterol and triglycerides in murine models due to the inhibition of the VLDL secretion pathway by destabilizing apolipoprotein B100.[49] A novel study in a Finnish population additionally suggested that the *TM6SF2* SNP rs58542926 is associated with dose-dependent reductions in cholesterol and triglyceride content for VLDL particles.[50] Evidence of lower circulating lipids coupled with increased fat retention in the liver on imaging further supports the finding that hepatic retention is directly correlated with the inability to synthesize lipid particles due to protein LOF.[51,52] Such findings would suggest that *TM6SF2* LOF variants paradoxically reduce the risk of atherosclerotic cardiovascular disease though we could not independently confirm these associations with our study.[47,53–55] As individuals with SLD are often at an increased risk of atherosclerosis due to metabolic syndrome, it is likely that these factors limit the interpretation of any independent association of *TM6SF2* variants with cardiovascular disease risk, as demonstrated in our study.[56]

Alternatively, the relationship between intracellular stores of triglycerides and the secretion rate of VLDL disrupted by *TM6SF2* could be attributed to the availability of phospholipids. Kinetic studies have postulated that *TM6SF2* is located in the ER region adjacent to lipoprotein particle assembly and mediates the fusion of VLDL-sized precursor particles.[57,58] The availability of phospholipids such as phosphatidylcholine (PC) is crucial to facilitate the expansion in size of lipid droplets while passing through the smooth ER and is regulated by the rate-limiting enzyme CTP (phosphocholine cytidylyltransferases).[59,60] Earlier studies suggest this mechanism underpins the effect of *TM6SF2* E167K, wherein a deficiency in PC results from impaired synthesis of polyunsaturated fatty acids.[61] Our metabolomics data is in line with this mechanism, and further confirms that *TM6SF2* L156P carriers have lower polyunsaturated fatty acid and choline levels, which may be observed due to a similar functional mechanism described for E167K.

Based on our *in silico* models, we predict that the E167K and L156P substitutions alter the conformation and properties of the EXPERA domain, particularly the surface loop spanning G162-R169. This may, in turn, impair the ability of *TM6SF2* to interact with stabilizing protein partners, such as ERLIN1.[62] Previous studies suggested that *TM6SF2* interacts with ERLIN1 to generate functional complexes[49,63] and that such complexes may stabilize the presence of *TM6SF2* within the ER membrane.[63] These predictions are consistent with previous experimental data, which indicate that expression of E167K and L156P in cells is associated with markedly reduced *TM6SF2* protein levels and increasing cellular instability, ultimately disrupting the biological function of the *TM6SF2* protein involved in VLDL secretion.[63,64] This decrease was not

explained by reduced gene expression,[63,64] but rather by accelerated protein degradation.[63] Therefore it is plausible that the E167K and L156P amino acid substitutions may impair the stability of TM6SF2-complexes and promote premature post-translational degradation which results in a LOF phenotype.

In tandem with a deeper phenotyping of these well-established variants is the discovery of rarer variants that were associated with a low MAF across all ancestries with specific predominance in European lineages (Table S16). P216L is a rare non-synonymous missense variant (P216L) in *TM6SF2* with a REVEL (rare exome variant ensemble learner) score of over 0.5, thereby predicted to be deleterious.[65] Our *in silico* findings strengthen this hypothesis, wherein P216L substitution leads to protein misfolding potentially explained by the biological loss of proline-mediated kinks between helices and loop sequences. These effects are similar to E167K and L156P, whereby both SNPs cause structural disruptions in the folding of these segments and therefore disrupt the luminal domains to result in LOF. Though the sample size is relative underpowered, the effect size of P216L is the most striking, as the risk of imaging-proven steatosis and clinically diagnosed SLD and steatohepatitis is similar to that of an individual carrying two copies of the E167K variant. Serum parameters were consistent with E167K and L156P, whereby P216L variant carriers exhibited lower serum cholesterol and triglycerides though these findings were not significant. Our PheWAS additionally suggested that carriers were at a higher risk of secondary thrombocytopenia, a clinical presentation that is present in a quarter of patients with SLD.[50] It is thought that the degree of thrombocytopenia is related to the degree of fatty infiltration of hepatic tissue, likely to involve hypersplenism, thrombopoietin deficiency, or reduced peripheral blood cell survival in the setting of liver damage.[66,67]

To strengthen our hypothesis on the effects of protein-altering variants, we performed a gene-burden analysis of pLOF variants to demonstrate phenotypes associated with complete LOF. In other relevant studies, statistical aggregation of rare LOF variants for their cumulative effects have revealed associations with various subclinical pathologies.[11,68,69] Here, a liver-focused association test confirmed that pLOF variants in *TM6SF2* were associated with a higher risk of SLD in the same directionality as E167K, L156P, and P216L. A PheWAS-informed association with disruptions in amino acid metabolism phenotypes suggests that pLOF variant carriers may additionally exhibit secondary manifestations of liver dysfunction. In line with *in vitro* studies, sulphur-amino acid metabolism may be disrupted in hepatic steatosis which may explain marked deficiencies in S-adenosyl-L-methionine metabolism in cirrhotic samples.[70,71] The association with SLD in the gene burden test was largely driven by a premature stop-gain variant on exon 2, featuring a G-to-A substitution at nucleotide 105 (c.105G>A) which introduces a premature stop of translation at amino acid 35 (W35X), which is located within the second helix (Fig. S11). Therefore, this variant is associated with loss of over 90% of the protein sequence and can be confidently considered LOF. W35X carriage is associated with an increased risk of MASLD-MASH and fibrosis/cirrhosis, alongside evidence of insulin resistance. Elevated ALP amongst carriers could potentially signal liver damage but are not frequently used as a proxy for liver injury,[72,73] which suggests that the stop-gain codon exerts a separate functional role from E167K and L156P. This is further corroborated by the increase in circulating triglycerides in W35X carriers despite retention of hepatic fat on imaging. Due to its small sample size, these findings did not reach significance in the UKB but the metabolomics analysis suggests that the directionality of lipoproteins and cholesterol are similar to E167K and L156P. This is the first report of a new stop-gain codon in *TM6SF2* and confirms the utilty of a gene burden for rare pLOFs of unknown significance in a known disease-causing gene.

The identification of SLD relied on ICD-codes, which could suffer from a degree of underdiagnosis and lead to an underestimation of *TM6SF2* effects in SLD, steatohepatitis, or HCC. With the recent change in the classification of SLD, ICD-10 codes may not fully capture the actual prevalence of MASLD in carriers. However, we showed robust associations of hepatic steatosis in two independent cohorts and confirmed potentially missed diagnoses using imaging-based methods of quantifying SLD in both biobanks and additional biopsy data from the PMBB. Addressing confounding factors, including the common PNPLA3 I148M risk allele, in addition to adjustments with multiple testing for our metabolomics data and PheWAS strengthened these associations.

Our study shows value in utilizing a genome-first approach to variants in *TM6SF2*. With the support of metabolomics/lipid analyses and *in silico* structural predictions, we conduct deeper phenotyping of E167K and L156P substitutions in the *TM6SF2* exon and clarify its loss-of-function effects. We also identified a novel non-synonymous missense variant P216L, predicted to be deleterious, to be associated with an increased risk of MASH and NLP-derived imaging steatosis with a similar effect size as E167K homozygotes. Finally, we aggregated rare pLOF variants into a gene burden to demonstrate that complete loss of function SNPs in *TM6SF2* are associated with an increased risk of liver damage. We confirm that genetically reduced *TM6SF2* activity results in increased steatosis, MASLD, MASH, and HCC and reduced plasma lipids.

**Affiliations**

[1]*Department of Medicine, Division of Translational Medicine and Human Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA;* [2]*Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA;* [3]*Department of Biobehavioral Health Sciences, School of Nursing, University of Pennsylvania, Philadelphia, PA 19104, USA;* [4]*The Institute for Translational Medicine and Therapeutics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA;* [5]*NewYork-Presbyterian, Weill Cornell Medical Center, New York, NY 10065, USA;* [6]*Regeneron Genetics Center, Tarrytown, NY, USA;* [7]*Department of Medicine III, Gastroenterology, Metabolic diseases and Intensive Care, University Hospital RWTH Aachen, 52074 Aachen, Germany;* [8]*Medical Department 1, Technische Universität, Dresden, Germany*

**Abbreviations**

ER, endoplasmic reticulum; HCC, hepatocellular carcinoma; MAF, minor allele frequency; MASH, metabolic dysfunction-associated steatohepatitis; MASLD, metabolic dysfunction-associated steatotic liver disease; NLP, natural language processing; OR, odds ratio; PDFF, proton density fat fraction; PheWAS, phenome-wide association study; (p)LOF, (putative) loss of function; PMBB, Penn Medicine Biobank; SLD, steatotic liver disease; SNP, single nucleotide polymorphism; *TM6SF2*, transmembrane 6 superfamily member 2; UKB, UK Biobank;

VLDL, very-low density lipoprotein; WES, whole-exome sequencing; WT, wild-type.

## Financial support

During the duration of the project, there were no financial supports directly engaged in the study design, collection, analysis and interpretation of data, writing of report and the decision to submit the article for publication. C.V.S is supported by a grant from the Interdisciplinary Centre for Clinical Research within the faculty of Medicine at the RWTH Aachen University (PTD 1-13/IA 532313), the Junior Principal Investigator Fellowship program of RWTH Aachen Excellence strategy and the NRW Rueckkehr Programme of the Ministry of Culture and Science of the German State of North Rhine-Westphalia (MKW). K.M.S is supported by the Federal Ministry of Education and Research (BMBF) and the Ministry of Culture and Science of the German State of North Rhine-Westphalia (MKW) under the Excellence strategy of the federal government and the Laender as well as the NRW Rueckkehr Programme of the Ministry of Culture and Science of the German State of North Rhine-Westphalia (MKW). C.V.S and K.M.S are supported by the CRC 1382 project A11 and B09 funded by Deutsche Forschungsgesellschaft (DFG, German Research Foundation) – Project-ID 403224013-SFB 1382.

## Conflicts of interest

All authors report no conflict of interest for this study. CV is currently employed at Arrowhead pharmaceuticals. At the time of study, CV was employed at the University of Pennsylvania.

Please refer to the accompanying ICMJE disclosure forms for further details.

## Authors' contributions

The Penn Medicine Biobank data was analyzed by H.H with full access to the data, overseen by D.Z, C.V.S and D.J.R. C.V.S had unrestricted access to the United Kingdom Biobank. C.V. and M.C.P. performed protein structure function predictions and analyses. H.H and C.V.S drafted the first version of the manuscript and all authors critically reviewed and edited it. H.H and C.V.S take responsibility for the integrity of the data and the accuracy of the data analysis. All authors agreed to submit the manuscript, read, and approved the final draft and take full responsibility of its content, including the accuracy of the data and its statistical analysis.

## Data availability statement

The datasets used in the current study have not been deposited in a public repository, but are available after approval of a reasonable application at https://www.ukbiobank.ac.uk for UKB and for PMBB under https://www.itmat.upenn.edu/biobank/researchers.html.

## Statement of previous presentation

An early version of the abstract was presented in a parallel session at the American Association for the Study of Liver Disease (AASLD) 2023 Liver Meeting and a poster session at the 2023 International Conference for Healthcare and Medical Students (ICHAMS).

## Acknowledgments

This research has been conducted using the UK Biobank Resource under Application Number 71300. UK biobank data was accessed by C.V.S and K.M.S only. *Copyright © 2024, NHS England. Re-used with the permission of the NHS England and/or UK Biobank. All rights reserved. This work uses data provided by patients and collected by the NHS as part of their care and support*. The PMBB is funded by the Perelman School of Medicine at the University of Pennsylvania, a gift from the Smilow family, and the National Center for Advancing Translational Sciences of the National Institutes of Health under CTSA Award Number UL1TR001878. The circoplot code was provided by Diego J Aguilar-Ramirez and adjusted by Jan Clusmann.

## Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhepr.2024.101243.

## References

[1] Cotter TG, Rinella M. Nonalcoholic fatty liver disease 2020: the state of the disease. Gastroenterology 2020;158(7):1851–1864. https://doi.org/10.1053/j.gastro.2020.01.052.

[2] Younossi ZM, Koenig AB, Abdelatif D, et al. Global epidemiology of nonalcoholic fatty liver disease—meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology 2016;64(1):73–84. https://doi.org/10.1002/hep.28431.

[3] Margini C, Dufour JF. The story of HCC in NAFLD: from epidemiology, across pathogenesis, to prevention and treatment. Liver Int 2016;36(3):317–324. https://doi.org/10.1111/liv.13031.

[4] Rinella ME, Lazarus JV, Ratziu V, et al. A multi-society Delphi consensus statement on new fatty liver disease nomenclature. J Hepatol 2023. https://doi.org/10.1016/j.jhep.2023.06.003. Epub 20230620. PubMed PMID: 37364790.

[5] Jonas W, Schürmann A. Genetic and epigenetic factors determining NAFLD risk. Mol Metab 2021;50:101111. https://doi.org/10.1016/j.molmet.2020.101111.

[6] Speliotes EK, Yerges-Armstrong LM, Wu J, et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. PLoS Genet 2011;7(3):e1001324. https://doi.org/10.1371/journal.pgen.1001324. Epub 20110310. PubMed PMID: 21423719; PMCID: PMC3053321.

[7] Anstee QM, Darlay R, Cockell S, et al. Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort*. J Hepatol 2020;73(3):505–515. https://doi.org/10.1016/j.jhep.2020.04.003.

[8] Mahdessian H, Taxiarchis A, Popov S, et al. TM6SF2 is a regulator of liver fat metabolism influencing triglyceride secretion and hepatic lipid droplet content. Proc Natl Acad Sci U S A 2014;111(24):8913–8918. https://doi.org/10.1073/pnas.1323785111. Epub 20140604. PubMed PMID: 24927523; PMCID: PMC4066487.

[9] Kozlitina J, Smagris E, Stender S, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 2014;46(4):352–356. https://doi.org/10.1038/ng.2901.

[10] Wilczewski CM, Obasohan J, Paschall JE, et al. Genotype first: clinical genomics research through a reverse phenotyping approach. Am J Hum Genet 2023;110(1):3–12. https://doi.org/10.1016/j.ajhg.2022.12.004. PubMed PMID: 36608682; PMCID: PMC9892776.

[11] Mirshahi UL, Kim J, Best AF, et al. A genome-first approach to characterize DICER1 pathogenic variant prevalence, penetrance, and phenotype. JAMA Netw Open 2021;4(2):e210112-e. https://doi.org/10.1001/jamanetworkopen.2021.0112.

[12] Park J, Levin MG, Haggerty CM, et al. A genome-first approach to aggregating rare genetic variants in LMNA for association with electronic health record phenotypes. Genet Med 2020;22(1):102–111. https://doi.org/10.1038/s41436-019-0625-8.

[13] MacLean MT, Jehangir Q, Vujkovic M, et al. Quantification of abdominal fat from computed tomography using deep learning and its association with electronic health records in an academic biobank. J Am Med Inform Assoc 2021;28(6):1178–1187. https://doi.org/10.1093/jamia/ocaa342. Epub 2021/02/13. PubMed PMID: 33576413; PMCID: PMC8661423.

[14] Schneider CV, Li T, Zhang D, et al. Large-scale identification of undiagnosed hepatic steatosis using natural language processing. EClinicalMedicine 2023;62:102149. https://doi.org/10.1016/j.eclinm.2023.102149. Epub 20230809. PubMed PMID: 37599905; PMCID: PMC10432816.

[15] Haas ME, Pirruccello JP, Friedman SN, et al. Machine learning enables new insights into genetic contributions to liver fat accumulation. Cell Genom 2021;1(3). https://doi.org/10.1016/j.xgen.2021.100066. PubMed PMID: 34957434; PMCID: PMC8699145.

[16] Szczepaniak LS, Nurenberg P, Leonard D, et al. Magnetic resonance spectroscopy to measure hepatic triglyceride content: prevalence of hepatic steatosis in the general population. Am J Physiol Endocrinol Metab 2005;288(2):E462–E468. https://doi.org/10.1152/ajpendo.00064.2004. Epub 20040831. PubMed PMID: 15339742.

[17] Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. Hepatology 2023;77(5):1797–1835. https://doi.org/10.1097/HEP.0000000000000323. Epub 20230317. PubMed PMID: 36727674; PMCID: PMC10735173.

[18] Hsu CL, Loomba R. From NAFLD to MASLD: implications of the new nomenclature for preclinical and clinical research. Nat Metab

2024;6(4):600–602. https://doi.org/10.1038/s42255-024-00985-1. PubMed PMID: 38383845.
[19] Schneider CV, Schneider KM, Raptis A, et al. Prevalence of at-risk MASH, MetALD and alcohol-associated steatotic liver disease in the general population. Aliment Pharmacol Ther 2024;59(10):1271–1281. https://doi.org/10.1111/apt.17958. Epub 20240319. PubMed PMID: 38500443.
[20] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596(7873):583–589. https://doi.org/10.1038/s41586-021-03819-2. Epub 20210715. PubMed PMID: 34265844; PMCID: PMC8371605.
[21] Mirdita M, Schütze K, Moriwaki Y, et al. ColabFold: making protein folding accessible to all. Nat Methods 2022;19(6):679–682. https://doi.org/10.1038/s41592-022-01488-1. Epub 20220530. PubMed PMID: 35637307; PMCID: PMC9184281.
[22] Mirdita M, Steinegger M, Söding J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. Bioinformatics 2019;35(16):2856–2858. https://doi.org/10.1093/bioinformatics/bty1057. PubMed PMID: 30615063; PMCID: PMC6691333.
[23] Mirdita M, von den Driesch L, Galiez C, et al. Uniclust databases of clustered and deeply annotated protein sequences and alignments. Nucleic Acids Res 2017;45(D1):D170–D176. https://doi.org/10.1093/nar/gkw1081. Epub 20161128. PubMed PMID: 27899574; PMCID: PMC5614098.
[24] Consortium U. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 2021;49(D1):D480–D489. https://doi.org/10.1093/nar/gkaa1100. PubMed PMID: 33237286; PMCID: PMC7778908.
[25] Karczewski KJ, Solomonson M, Chao KR, et al. Systematic single-variant and gene-based association testing of thousands of phenotypes in 394, 841 UK Biobank exomes. Cell Genom 2022;2(9):100168. https://doi.org/10.1016/j.xgen.2022.100168. Epub 20220815. PubMed PMID: 36778668; PMCID: PMC9903662.
[26] Sanchez-Pulido L, Ponting CP. TM6SF2 and MAC30, new enzyme homologs in sterol metabolism and common metabolic disease. Front Genet 2014;5:439. https://doi.org/10.3389/fgene.2014.00439. Epub 20141211. PubMed PMID: 25566323; PMCID: PMC4263179.
[27] Wilman HR, Shi J, Deane CM. Helix kinks are equally prevalent in soluble and membrane proteins. Proteins 2014;82(9):1960–1970. https://doi.org/10.1002/prot.24550. Epub 20140416. PubMed PMID: 24638929; PMCID: PMC4285789.
[28] Schmidt T, Situ AJ, Ulmer TS. Structural and thermodynamic basis of proline-induced transmembrane complex stabilization. Sci Rep 2016;6:29809. https://doi.org/10.1038/srep29809. Epub 20160720. PubMed PMID: 27436065; PMCID: PMC4951694.
[29] Liu J, Ginsberg HN, Reyes-Soffer G. Basic and translational evidence supporting the role of TM6SF2 in VLDL metabolism. Curr Opin Lipidol 2024. https://doi.org/10.1097/MOL.0000000000000930. Epub 20240311. PubMed PMID: 38465912.
[30] Liu YL, Reeves HL, Burt AD, et al. TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. Nat Commun 2014;5:4309. https://doi.org/10.1038/ncomms5309. Epub 20140630. PubMed PMID: 24978903; PMCID: PMC4279183.
[31] Koo BK, Joo SK, Kim D, et al. Additive effects of PNPLA3 and TM6SF2 on the histological severity of non-alcoholic fatty liver disease. J Gastroenterol Hepatol 2018;33(6):1277–1285. https://doi.org/10.1111/jgh.14056.
[32] Dongiovanni P, Petta S, Maglio C, et al. Transmembrane 6 superfamily member 2 gene variant disentangles nonalcoholic steatohepatitis from cardiovascular disease. Hepatology 2015;61(2):506–514. https://doi.org/10.1002/hep.27490. PubMed PMID: 25251399.
[33] Du S, Lu L, Miao Y, et al. E167K polymorphism of TM6SF2 gene affects cell cycle of hepatocellular carcinoma cell HEPA 1-6. Lipids Health Dis 2017;16(1):76. https://doi.org/10.1186/s12944-017-0468-8.
[34] Tang S, Zhang J, Mei TT, et al. Association of TM6SF2 rs58542926 T/C gene polymorphism with hepatocellular carcinoma: a meta-analysis. BMC Cancer 2019;19(1):1128. https://doi.org/10.1186/s12885-019-6173-4. Epub 20191121. PubMed PMID: 31752753; PMCID: PMC6868855.
[35] Anstee QM, Seth D, Day CP. Genetic factors that affect risk of alcoholic and nonalcoholic fatty liver disease. Gastroenterology 2016;150(8):1728–17244 e7. https://doi.org/10.1053/j.gastro.2016.01.037. Epub 20160210. PubMed PMID: 26873399.
[36] Buch S, Stickel F, Trepo E, et al. A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. Nat Genet 2015;47(12):1443–1448. https://doi.org/10.1038/ng.3417. Epub 20151019. PubMed PMID: 26482880.
[37] Longo M, Meroni M, Paolini E, et al. TM6SF2/PNPLA3/MBOAT7 loss-of-function genetic variants impact on NAFLD development and progression both in patients and in in vitro models. Cell Mol Gastroenterol Hepatol 2022;13(3):759–788. https://doi.org/10.1016/j.jcmgh.2021.11.007. Epub 20211123. PubMed PMID: 34823063; PMCID: PMC8783129.
[38] Noureddin M, Truong E, Mayo R, et al. Serum identification of at-risk MASH: the metabolomics-advanced steatohepatitis fibrosis score (MASEF). Hepatology 2024;79(1):135–148. https://doi.org/10.1097/HEP.0000000000000542. Epub 20230724. PubMed PMID: 37505221; PMCID: PMC10718221.
[39] Noureddin M, Muthiah MD, Sanyal AJ. Drug discovery and treatment paradigms in nonalcoholic steatohepatitis. Endocrinol Diabetes Metab 2020;3(4):e00105. https://doi.org/10.1002/edm2.105. Epub 20191210. PubMed PMID: 33102791; PMCID: PMC7576222.
[40] Falleti E, Cussigh A, Cmet S, et al. PNPLA3 rs738409 and TM6SF2 rs58542926 variants increase the risk of hepatocellular carcinoma in alcoholic cirrhosis. Dig Liver Dis 2016;48(1):69–75. https://doi.org/10.1016/j.dld.2015.09.009. Epub 20150928. PubMed PMID: 26493626.
[41] Donati B, Dongiovanni P, Romeo S, et al. MBOAT7 rs641738 variant and hepatocellular carcinoma in non-cirrhotic individuals. Sci Rep 2017;7(1):4492. https://doi.org/10.1038/s41598-017-04991-0. Epub 20170703. PubMed PMID: 28674415; PMCID: PMC5495751.
[42] Innes H, Nischalke HD, Guha IN, et al. The rs429358 locus in apolipoprotein E is associated with hepatocellular carcinoma in patients with cirrhosis. Hepatol Commun 2022;6(5):1213–1226. https://doi.org/10.1002/hep4.1886. Epub 20211227. PubMed PMID: 34958182; PMCID: PMC9035556.
[43] Pant A, Chen Y, Kuppa A, et al. Perturbation of TM6SF2 expression alters lipid metabolism in a human liver cell line. Int J Mol Sci 2021;22(18). https://doi.org/10.3390/ijms22189758. Epub 20210909. PubMed PMID: 34575933; PMCID: PMC8471112.
[44] Sveinbjornsson G, Ulfarsson MO, Thorolfsdottir RB, et al. Multiomics study of nonalcoholic fatty liver disease. Nat Genet 2022;54(11):1652–1663. https://doi.org/10.1038/s41588-022-01199-5. Epub 20221024. PubMed PMID: 36280732; PMCID: PMC9649432.
[45] Long T, Hassan A, Thompson BM, et al. Structural basis for human sterol isomerase in cholesterol biosynthesis and multidrug recognition. Nat Commun 2019;10(1):2452. https://doi.org/10.1038/s41467-019-10279-w.
[46] Fan Y, Lu H, Guo Y, et al. Hepatic transmembrane 6 superfamily member 2 regulates cholesterol metabolism in mice. Gastroenterology 2016;150(5):1208–1218. https://doi.org/10.1053/j.gastro.2016.01.005. Epub 20160113. PubMed PMID: 26774178; PMCID: PMC4842105.
[47] Luo F, Oldoni F, Das A. TM6SF2: a novel genetic player in nonalcoholic fatty liver and cardiovascular disease. Hepatol Commun 2022;6(3):448–460. https://doi.org/10.1002/hep4.1822. Epub 20210916. PubMed PMID: 34532996; PMCID: PMC8870032.
[48] Smagris E, Gilyard S, BasuRay S, et al. Inactivation of Tm6sf2, a gene defective in fatty liver disease, impairs lipidation but not secretion of very low density lipoproteins. J Biol Chem 2016;291(20):10659–10676. https://doi.org/10.1074/jbc.M116.719955. Epub 20160324. PubMed PMID: 27013658; PMCID: PMC4865914.
[49] Li BT, Sun M, Li YF, et al. Disruption of the ERLIN-TM6SF2-APOB complex destabilizes APOB and contributes to non-alcoholic fatty liver disease. PLoS Genet 2020;16(8):e1008955. https://doi.org/10.1371/journal.pgen.1008955. Epub 20200810. PubMed PMID: 32776921; PMCID: PMC7462549.
[50] Kim DS, Jackson AU, Li YK, et al. Novel association of TM6SF2 rs58542926 genotype with increased serum tyrosine levels and decreased apoB-100 particles in Finns. J Lipid Res 2017;58(7):1471–1481. https://doi.org/10.1194/jlr.P076034. Epub 20170524. PubMed PMID: 28539357; PMCID: PMC5496043.
[51] Haas ME, Pirruccello JP, Friedman SN, et al. Machine learning enables new insights into genetic contributions to liver fat accumulation. Cell Genomics 2021;1(3):100066. https://doi.org/10.1016/j.xgen.2021.100066.
[52] Park J, MacLean MT, Lucas AM, et al. Exome-wide association analysis of CT imaging-derived hepatic fat in a medical biobank. Cell Rep Med 2022;3(12):100855. https://doi.org/10.1016/j.xcrm.2022.100855.
[53] Surakka I, Horikoshi M, Magi R, et al. The impact of low-frequency and rare variants on lipid levels. Nat Genet 2015;47(6):589–597. https://doi.org/10.1038/ng.3300. Epub 20150511. PubMed PMID: 25961943; PMCID: PMC4757735.
[54] Prill S, Caddeo A, Baselli G, et al. The TM6SF2 E167K genetic variant induces lipid biosynthesis and reduces apolipoprotein B secretion in human hepatic 3D spheroids. Sci Rep 2019;9(1):11585. https://doi.org/10.1038/s41598-019-47737-w. Epub 20190812. PubMed PMID: 31406127; PMCID: PMC6690969.
[55] Brouwers M, Simons N, Stehouwer CDA, et al. Non-alcoholic fatty liver disease and cardiovascular disease: assessing the evidence for causality. Diabetologia 2020;63(2):253–260. https://doi.org/10.1007/s00125-019-05024-3. Epub 20191111. PubMed PMID: 31713012; PMCID: PMC6946734.

[56] Kahali B, Liu YL, Daly AK, et al. TM6SF2: catch-22 in the fight against nonalcoholic fatty liver disease and cardiovascular disease? Gastroenterology 2015;148(4):679–684. https://doi.org/10.1053/j.gastro.2015.01.038. Epub 20150129. PubMed PMID: 25639710.
[57] Reyes-soffer G, Thomas T, Matveyenko A, et al. Abstract 219: Effects of Homozygosity for the P.e167k Variant in *Tm6sf2* on Lipid and Lipoprotein Metabolism in Humans. Arteriosclerosis, Thromb Vasc Biol 2020;40(Suppl_1):A219–A. https://doi.org/10.1161/atvb.40.suppl_1.219.
[58] O'Hare EA, Yang R, Yerges-Armstrong LM, et al. TM6SF2 rs58542926 impacts lipid processing in liver and small intestine. Hepatology 2017;65(5).
[59] Fei W, Shui G, Zhang Y, et al. A role for phosphatidic acid in the formation of "supersized" lipid droplets. PLOS Genet 2011;7(7):e1002201. https://doi.org/10.1371/journal.pgen.1002201.
[60] Krahmer N, Guo Y, Wilfling F, et al. Phosphatidylcholine synthesis for lipid droplet expansion is mediated by localized activation of CTP:phosphocholine cytidylyltransferase. Cell Metab 2011;14(4):504–515. https://doi.org/10.1016/j.cmet.2011.07.013.
[61] Luukkonen PK, Zhou Y, Nidhina Haridas PA, et al. Impaired hepatic lipid synthesis from polyunsaturated fatty acids in TM6SF2 E167K variant carriers with NAFLD. J Hepatol 2017;67(1):128–136. https://doi.org/10.1016/j.jhep.2017.02.014.
[62] Rendel MD, Vitali C, Creasy KT, et al. The common p.Ile291Val variant of ERLIN1 enhances TM6SF2 function and is associated with protection against MASLD. Med 2024. https://doi.org/10.1016/j.medj.2024.04.010. Epub 20240516. PubMed PMID: 38776916.
[63] Ehrhardt N, Doche ME, Chen S, et al. Hepatic Tm6sf2 overexpression affects cellular ApoB-trafficking, plasma lipid levels, hepatic steatosis and atherosclerosis. Hum Mol Genet 2017;26(14):2719–2731. https://doi.org/10.1093/hmg/ddx159. PubMed PMID: 28449094; PMCID: PMC5886214.
[64] Kozlitina J, Smagris E, Stender S, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 2014;46(4):352–356. https://doi.org/10.1038/ng.2901. Epub 20140216. PubMed PMID: 24531328; PMCID: PMC3969786.
[65] Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet 2016;99(4):877–885. https://doi.org/10.1016/j.ajhg.2016.08.016. Epub 20160922. PubMed PMID: 27666373; PMCID: PMC5065685.
[66] Lopez-Trujillo MA, Olivares-Gazca JM, Cantero-Fortiz Y, et al. Nonalcoholic fatty liver disease and thrombocytopenia III: its association with insulin resistance. Clin Appl Thromb Hemost 2019;25:1076029619888694. https://doi.org/10.1177/1076029619888694. PubMed PMID: 31840531; PMCID: PMC7019400.
[67] Rivera-Alvarez M, Cordova-Ramirez AC, Elias-De-La-Cruz GD, et al. Nonalcoholic fatty liver disease and thrombocytopenia IV: its association with granulocytopenia. Hematol Transfus Cell Ther 2022;44(4):491–496. https://doi.org/10.1016/j.htct.2021.06.004. Epub 20210721. PubMed PMID: 34312112; PMCID: PMC9605888.
[68] Ahmadmehrabi S, Li B, Park J, et al. Genome-first approach to rare EYA4 variants and cardio-auditory phenotypes in adults. Hum Genet 2021;140(6):957–967. https://doi.org/10.1007/s00439-021-02263-6. Epub 20210321. PubMed PMID: 33745059.
[69] Park J, Levin MG, Haggerty CM, et al. A genome-first approach to aggregating rare genetic variants in LMNA for association with electronic health record phenotypes. Genet Med 2020;22(1):102–111. https://doi.org/10.1038/s41436-019-0625-8. Epub 20190806. PubMed PMID: 31383942; PMCID: PMC7719049.
[70] Duce AM, Ortiz P, Cabrero C, et al. S-adenosyl-L-methionine synthetase and phospholipid methyltransferase are inhibited in human cirrhosis. Hepatology 1988;8(1):65–68. https://doi.org/10.1002/hep.1840080113. PubMed PMID: 3338721.
[71] Kwon DY, Jung YS, Kim SJ, et al. Impaired sulfur-amino acid metabolism and oxidative stress in nonalcoholic fatty liver are alleviated by betaine supplementation in rats. J Nutr 2009;139(1):63–68. https://doi.org/10.3945/jn.108.094771. Epub 20081203. PubMed PMID: 19056644.
[72] Siddique A, Kowdley KV. Approach to a patient with elevated serum alkaline phosphatase. Clin Liver Dis 2012;16(2):199–229. https://doi.org/10.1016/j.cld.2012.03.012. Epub 20120406. PubMed PMID: 22541695; PMCID: PMC3341633.
[73] Pantsari MW, Harrison SA. Nonalcoholic fatty liver disease presenting with an isolated elevated alkaline phosphatase. J Clin Gastroenterol 2006;40(7):633–635. https://doi.org/10.1097/00004836-200608000-00015. PubMed PMID: 16917408.
[74] Bragg F, Trichia E, Aguilar-Ramirez D, et al. Predictive value of circulating NMR metabolic biomarkers for type 2 diabetes risk in the UK Biobank study. BMC Med 2022;20(1):159. https://doi.org/10.1186/s12916-022-02354-9. Epub 20220503. PubMed PMID: 35501852; PMCID: PMC9063288.

Keywords: *TM6SF2*; MASLD; MASH; HCC; Lipid metabolism; Genome-first; Pathogenic variants.
*Received 9 July 2024; received in revised form 28 September 2024; accepted 7 October 2024; Available online 11 October 2024*