Metabolic effects of risk alleles in PNPLA3, TM6SF2, GCKR and LYPLAL1 inform about heterogeneity of non-alcoholic fatty liver disease

Eeva Sliz1,2,3, eeva.sliz@oulu.fi; Sylvain Sebert1,3,4, sylvain.sebert@oulu.fi; Peter Würtz5,6, peter.wurtz@helsinki.fi; Antti K Kangas6, antti.kangas@nightingalehealth.com; Pasi Soininen6,7, pasi.soininen@nightingalehealth.com; Terho Lehtimäki8, terho.lehtimaki@uta.fi; Mika Kähönen9, mika.kahonen@uta.fi; Jorma Viikari10, jorvii@utu.fi; Minna Männikkö1,2, minna.mannikko@oulu.fi; Mika Ala-Korpela1,3,7,11,12, mika.ala-korpela@computationalmedicine.fi; Olli T Raitakari13,14, olli.raitakari@utu.fi; and Johannes Kettunen1,3,11,12 *, johannes.kettunen@computationalmedicine.fi

1 Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland
2 Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland
3 Biocenter Oulu, University of Oulu, Oulu, Finland
4 Department of Genomics of complex diseases, School of Public Health, Imperial College London, UK
5 Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland
6 Nightingale Health Ltd, Helsinki, Finland
7 NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland
8 Department of Clinical Chemistry, Fimlab Laboratories, Finnish Cardiovascular Research Center Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland.
9 Department of Clinical Physiology, Tampere University Hospital and Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland
10 Department of Medicine, University of Turku and Division of Medicine, Turku University Hospital, Turku, Finland
11 Computational Medicine, Faculty of Medicine, University of Oulu, Oulu, Finland
12 Medical Research Council Integrative Epidemiology Unit at the University of Bristol, Bristol, UK
13 Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland
14 Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland.

Key words
Genetic association, NMR metabolomics, Steatosis, Steatohepatitis
Contact Information
Johannes Kettunen
Address: Faculty of Medicine, Computational Medicine, Center For Life Course Health Research, Aapistie 5A, P.O.Box 5000, 90014 University of Oulu, Oulu, Finland
Telephone: + 358 50 562 9718
E-mail: johannes.kettunen@computationalmedicine.fi

List of Abbreviations
NAFLD, non-alcoholic fatty liver disease; PNPLA3, Patatin-like phospholipase domain containing 3; TM6SF2, Transmembrane 6 superfamily member 2; LYPLAL1, Lysophospholipase-like 1; GCKR, Glucokinase regulator; GWAS, genome-wide association study; YFS, Young Finns Study; VLDL, very low-density lipoprotein; IDL, intermediate-density lipoprotein; LDL, low-density lipoprotein; NMR, nuclear magnetic resonance; SNP, single nucleotide polymorphism; FA, fatty acid; TG, triglyceride.

Financial Support
ES was funded by University of Oulu Graduate School (UniOGS). SS is thankful to the academy of Finland, Biocenter Oulu and the European commission (grant DynaHEALTH - H2020 – 633595). JK was funded through Academy of Finland (grant numbers 283045, 297338 and 307247) and Novo Nordisk Foundation (NNF17OC0026062). PW: AoF 312476 & 312477, and Novo Nordisk Foundation. MAK was supported by the Sigrid Juselius Foundation. MAK works in a Unit that is supported by the University of Bristol and UK Medical Research Council (MC_UU_12013/1). The Young Finns Study has been financially supported by the Academy of Finland: grants 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117787 (Gendi), and 41071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001 for T.L.); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research; Finnish Cultural Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; and Diabetes Research Foundation of Finnish Diabetes Association.

Disclosures
PW, AJK, and PS are employees and shareholders of Nightingale Health Ltd, a company offering NMR-based biomarker profiling.
Abstract

Fatty liver has been associated with unfavourable metabolic changes in circulation and is considered as a risk factor for cardiometabolic complications such as type 2 diabetes and cardiovascular disease. We aimed to provide insights in fatty liver related metabolic deviations by studying the resemblance between the metabolic profile associated with fatty liver observationally and metabolic profiles of non-alcoholic fatty liver disease (NAFLD) risk increasing genotypes. We determined cross-sectional associations of ultrasound-ascertained fatty liver status with 123 metabolic traits in 1,810 individuals aged 34-49 years from The Cardiovascular Risk in Young Finns Study. The cross-sectional associations were compared with the association profiles of NAFLD risk alleles in PNPLA3, TM6SF2, GCKR, and LYPLAL1 with the corresponding metabolic traits obtained from a publicly available genome-wide association study including up to 24,925 European individuals. The analysis revealed substantially different metabolic effects of the risk alleles. PNPLA3 rs738409-G, the strongest genetic risk factor to NAFLD, did not associate with metabolic changes. GCKR rs1260326-T resulted in an association profile similar to the observational fatty liver associations. Metabolic effects of LYPLAL1 rs12137855-C were similar, but statistically less robust, to the effects of GCKR rs1260326-T. In contrast, NAFLD risk allele TM6SF2 rs58542926-T displayed opposite metabolic associations when compared with the observational association pattern. Conclusion: The divergent effects of the risk alleles on circulating lipids and metabolites underline involvement of several metabolic pathways in NAFLD and suggest that there are pathogenically different subtypes of NAFLD with alternate metabolic consequences. NAFLD risk alleles may have neutral or even cardioprotective effect on circulating lipids and metabolites providing evidence that hepatic lipid accumulation by itself would not necessarily cause the metabolic deviations associated observationally with fatty liver.
Non-alcoholic fatty liver disease (NAFLD) covers a range of liver disorders that originate from excessive lipid, mainly triglyceride, accumulation in the liver (1). Fatty liver has been proposed to result in dyslipidemia due to increased secretion of very low-density lipoproteins (VLDL) and impaired clearance of intermediate and low-density lipoproteins (IDL, LDL) from circulation (2, 3). Studies using detailed metabolic profiling have shown that fatty liver associates with a wide range of metabolic aberrations in circulation (4-6). Kaikkonen et al. assessed cross-sectional and prospective associations of fatty liver with 68 circulating metabolic measures in the population-based Young Finns Study (YFS): the most prominent associations were seen with triglycerides in the largest VLDL particles as well as with VLDL particle concentrations, but fatty liver was robustly associated also with many non-lipid traits, such as branched-chain amino acids leucine and isoleucine (4). Observations from animal models and human studies have shown that liver fat content correlates with levels of blood lipids and glucose (7-9). However, fatty liver does not always seem to relate to dyslipidemia or changes in glycemic traits and thus its effects on metabolic changes in circulation has remained unclear (10, 11).

Genetic factors contribute to the pathogenesis of fatty liver. Four DNA sequence variants, Patatin-like phospholipase domain containing 3 (PNPLA3) rs738409-G, Glucokinase regulator (GCKR) rs780094-T, Neurocan (NCAN) rs2228603-T, and Lysophospholipase-like 1 (LYPLAL1) rs12137855-C, were associated with computed tomography defined steatosis and biopsy-proven NAFLD involving lobular inflammation and fibrosis in a large-scale genome-wide association study (GWAS) (12). Variants in these four loci have also been used to determine a genetic risk score for NAFLD (13). Other studies have shown that GCKR rs1260326-T and Transmembrane 6 superfamily member 2 (TM6SF2) rs58542926-T are the functional variants at the GCKR and NCAN loci, respectively (14-16). PNPLA3 rs738409-G is the strongest genetic risk factor for NAFLD (17) having an odds ratio of 3.24 for histologic NAFLD (12). A meta-analysis on 2,937 individuals with biopsy-diagnosed NAFLD showed that the GG genotype increases hepatic triglyceride content by 73% in comparison to reference CC genotype (18). GCKR rs1260326-T and TM6SF2 rs58542926-T are also recognized as important determinants of inter-individual variation in liver fat (14, 19, 20). Among the aforementioned variants, the function of LYPLAL1 rs12137855-C is the least known.
In the present study we perform detailed metabolic profiling of fatty liver in young and middle-aged adults with ultrasound ascertained fatty liver. To add insights into molecular mechanisms of fatty liver related metabolic aberrations, we compare how the observational fatty liver associations match with metabolic association profiles of the aforementioned NAFLD risk alleles. We utilize publicly available summary statistics from a metabolomics GWAS to assess the detailed metabolic effects of the risk variants (21). Understanding the relation between fatty liver and changes in circulating lipids and metabolites is helpful in acquiring more opportunities for treatment and prevention of this complex condition and related cardiometabolic complications.

Methods

Cross-sectional fatty liver associations with circulating metabolites

The Cardiovascular risk in Young Finns Study (YFS) is a population based follow-up study started in 1980. In 2011, 2,046 individuals aged 34 to 49 years participated to ultrasound imaging (Acuson Sequoia 512, Acuson, Mountain View, CA, USA) of the liver. A trained sonographer assessed the presence of fatty liver using 4.0 MHz adult abdominal transducers, and the participants were categorised into two groups: fatty liver and no fatty liver. In total, fatty liver was diagnosed in 18.6% (N=372) of the participants. The population and ultrasound imaging of the liver is described in more detail in (4). After exclusion of pregnant women and individuals using lipid lowering medication or oral contraceptives, the total number of individuals included to determine the associations between fatty liver and metabolic phenotypes was 1,810 (N_{fatty liver}=338). The study was approved by the local Ethics Committee and written informed consent was obtained from all the participants.

NAFLD risk allele associations with circulating metabolites

Effect estimates of PNPLA3 rs738409-G, GCKR rs1260326-T, GCKR rs780094-T, TM6SF2 rs58542926-T, NCAN rs2228603-T, and LYPLAL1 rs12137855-C on the metabolic traits were acquired from a published GWAS performed using 14 European cohorts from Finland, Germany, the Netherlands and Estonia adding
up to 24,925 individuals (21). The mean age and BMI of the participants per cohort ranged from 23.9 to 61.3 years and from 23.1 to 28.2 kg/m² with the whole sample means being 46.3 years and 26.0 kg/m².

Risk allele frequencies and odds ratios to NAFLD for the studied loci are described in Table 1. To facilitate comparison of the risk allele associated metabolic changes relative to NAFLD risk increase, the risk allele effects and corresponding standard errors were scaled with respect to the log(odds ratio) on histologic NAFLD associated with the corresponding locus in a large-scale GWAS (12).

**Metabolic profiling**

The metabolic profiles of all samples included were assessed using a nuclear magnetic resonance (NMR) metabolomics platform (Nightingale Health, Helsinki, Finland) described in (21, 22) and reviewed in more detail in (23, 24). The platform covers multiple distinct metabolic traits including lipoprotein subclasses and their lipids, fatty acids, amino acids and glycolysis precursors. To determine the cross-sectional metabolic aberrations associated with fatty liver, we utilized 123 metabolic measures representing a broad molecular signature of systemic metabolism (Supplementary Table 1).

**Statistical analyses**

All analyses were done using R version 3.2.2. Statistical significance was considered at P<0.002 (0.05/22), where 22 is the number of principal components explaining 95% of the variation in the NMR metabolomics data (21).

**Cross-sectional associations**

Linear regression models were fitted to determine the cross-sectional associations of fatty liver with each of the metabolic measures in the YFS population. To facilitate the comparison of observed and genetic effect estimates, the data processing and analysis model were done correspondingly to Kettunen et al. (21): the metabolic phenotypes were adjusted for age, sex, and ten first genetic principal components preceding the analysis, and the resulting residuals were transformed to normal distribution by inverse rank-based normal transformation. The adjusted and transformed metabolic phenotypes were used as outcomes in the equations and fatty liver served as a categorical variable (fatty liver vs. no fatty liver).
Comparisons of the metabolic association profiles

The overall resemblance between the risk alleles and observational fatty liver effects on metabolites was determined by a linear fit of each of the risk allele association profile versus cross-sectional fatty liver association profile (25, 26). In addition, the linear fit of all the pairs of the risk allele association profiles were determined in order to identify alleles with similar metabolic effects informing about gene products functioning in related biological pathways. This method was also used to ensure that the metabolic profiles of the NAFLD risk alleles GCKR rs780094-T and NCAN rs2228603-T correspond to ones of GCKR rs1260326-T and TM6SF2 rs58542926-T that have been identified as the functional variants in GCKR and NCAN/TM6SF2 loci, respectively (14-16).

Results

Fatty liver and circulating metabolites

Characteristics of the YFS study population are shown in Table 2. Fatty liver was associated with 88 metabolic phenotypes (P<0.002) in the YFS population when adjusted with age, sex, and ten genetic principal components (Panel 1 in Figure 1; Supplementary Figure 1; Supplementary Table 2). Lipoprotein subclass concentrations showed a trend where the largest VLDL particles displayed the most pronounced associations that got weaker while the particle size decreased, and again stronger for medium and small LDL particles (Panel 1 in Figure 1). Concentration of very large and large HDL particles showed negative association with fatty liver, while small HDL particles displayed a strong positive association. Fatty liver was also associated with increased VLDL diameter while the associations with LDL and HDL particle diameters were negative. Fatty liver associated negatively with apolipoprotein A-I and positively with apolipoprotein B. Concentrations of serum triglycerides and triglycerides in all lipoprotein subclasses were increased. Fatty liver was also associated with increased level of saturation of circulating fatty acids along with increased concentrations of circulating fatty acids, while it showed a negative association with fatty acid length. In addition, fatty liver displayed positive association with glucose, lactate, pyruvate, and glycerol, as well as
with amino acids alanine, isoleucine, leucine, valine, phenylalanine, and tyrosine. Association between fatty liver and amino acid glutamine was negative.

**NAFLD risk alleles in GCKR and LYPLAL1 tend to increase concentrations of circulating lipids**

The studied NAFLD risk increasing alleles showed different association profiles on metabolic measures.

*GCKR* rs1260326-T associated with increased particle concentrations of all VLDL, IDL, and LDL subclasses as well as concentrations of medium and small HDL particles (Panel 2 in Figure 1; Supplementary Figure 2; Supplementary Table 3). The amount of triglycerides was increased in all the lipoprotein subclasses. *GCKR* rs1260326-T was associated with increased diameter of VLDL particles, while the association with LDL and HDL particle diameters were negative. The variant showed positive association with concentrations of apolipoproteins A-I and B, as well as with all the studied fatty acids, and displayed negative association with fatty acid length. *GCKR* rs1260326-T associated also with increased fatty acid saturation. *GCKR* rs1260326-T was associated positively with glycolysis related metabolites lactate, pyruvate, and glycerol, while association with citrate was negative. This risk allele associated positively with amino acids alanine, isoleucine, leucine, and valine, and negatively with glutamine. The results of *GCKR* rs780094-T, a variant originally identified as a NAFLD risk allele in the GCKR locus (12) being in linkage disequilibrium with the functional *GCKR* rs1260326-T (14), were highly similar to the results of *GCKR* rs1260326-T and can be found in the supplementary material (Supplementary Figure 1; Supplementary Figure 2; Supplementary Table 3).

The metabolic association profile of *LYPLAL* rs12137855-C was highly similar to the association profile of *GCKR* rs1260326-T in terms of highly correlated point estimates of the two risk alleles (Supplementary Figure 3A), but the effects of *LYPLAL* rs12137855-C were statistically less robust and the effect magnitudes were weaker than for the *GCKR* variant (Panel 3 in Figure 1; Supplementary Figure 2; Supplementary Table 3).

**PNPLA3 rs738409-G does not show association with circulating metabolic traits**

*PNPLA3* rs738409-G, the strongest genetic contributor to the hepatic fat content (17), displayed metabolic association profile close to null (Panel 4 in Figure 1; Supplementary Figure 2; Supplementary Table 3). When
compared with the cross-sectional fatty liver effects, the NAFLD-scaled metabolic effects of PNPLA3 rs738409-G were much closer to zero, and the confidence intervals for effect estimates of cross-sectional fatty liver and PNPLA3 rs738409-G were clearly separated.

**TM6SF2 rs58542926-T associates with lower-risk metabolic profile**

TM6SF2 rs58542926-T displayed strong association profile throughout the studied metabolites (Panel 5 in Figure 1; Supplementary Figure 2; Supplementary Table 3). However, the associations were negative indicating reduced concentrations of the lipids and metabolites in relation to higher NAFLD risk. The TM6SF2 variant was observed to decrease concentrations of all the VLDL, IDL and LDL particle subclasses and all the lipid species in these subclasses. In addition, the variant associated inversely with serum total triglycerides and triglycerides in all the lipoprotein subclasses including the HDL subclasses. The diameter of the VLDL particles was reduced, while the variant did not contribute to the LDL nor HDL particle diameters. The TM6SF2 rs58542926-T associated also with decreased concentration of the apolipoprotein B but did not influence concentration of apolipoprotein A-I. It associated inversely also with all the fatty acid concentrations, while the qualitative measures such as fatty acid length or saturation measures were not influenced by this variant. In addition, concentrations of histidine, isoleucine and glycerol were reduced. The results for NCAN rs2228603-T in the same NAFLD risk locus can be found in the supplement (Supplementary Figure 1; Supplementary Figure 2; Supplementary Table 3).

**Resemblance of the metabolic effects**

The correspondence between the metabolic effects of the risk alleles and observational fatty liver was the highest between the GCKR rs1260326-T and fatty liver ($R^2 = 0.77$; Figure 2C). The remaining coefficients of determination were $R^2 = 0.67$ for LYPLAL1 rs12137855-C versus fatty liver (Figure 2D), $R^2 = 0.45$ for TM6SF2 rs58542926-T versus fatty liver (Figure 2B), and $R^2 = 0.30$ for PNPLA3 rs738409 G versus fatty liver (Figure 2A).

Pairwise comparisons of the risk allele association profiles indicated that the effects of GCKR rs1260326-T and LYPLAL1 rs12137855-C on circulating metabolites were highly similar ($R^2 = 0.71$; Supplementary Figure
3A). The overall pattern of metabolic effects of TM6SF2 rs58542926-T correlated inversely with effects of both GCKR rs1260326-T and LYPLAL1 rs12137855-C ($R^2 = 0.66$, Supplementary Figure 3C, and $R^2 = 0.50$, Supplementary Figure 3E, respectively). The remaining correlations were weaker ($0.19 \leq R^2 \leq 0.24$; Supplementary Figure 3).

**Discussion**

We assessed fatty liver related metabolic changes in 1,810 young and middle-aged adults from a Finnish population cohort using 123 circulating metabolic measures covering a wide range of metabolic pathways, and further compared the cross-sectional observations with metabolic association profiles of known NAFLD risk alleles obtained from a publicly available metabolomics GWAS including up to 24,925 individuals (21). The studied NAFLD risk alleles resulted in divergent metabolic association profiles. Despite PNPLA3 rs738409-G being the strongest genetic risk factor for NAFLD (17), it showed a null effect on circulating lipids and metabolites. Association profile of GCKR rs1260326-T showed similarities to the cross-sectional fatty liver associations whereas the TM6SF2 rs58542926-T provided strong statistical evidence to the opposite direction. The present results provide molecular evidence supportive to the recent findings about worsened metabolic features seen in association obesity linked NAFLD but not with “genetic NAFLD”, as defined by NAFLD arising due to risk alleles in PNPLA3 and TM6SF2 (19, 27-29).

The differing metabolic effects reflect the biological functions of the risk alleles and suggest that fatty liver can arise from at least three distinct molecular pathways which have divergent consequences on circulating lipids and metabolites. These pathways are summarized in Figure 3. **Pathway I, excessive hepatic glucose levels and amplified lipogenesis:** GCKR rs1260326-T reduces GCKR ability to inhibit glucokinase resulting in enhanced hepatic glucose uptake, reduced fatty acid oxidation and increased lipogenesis (30). Hepatic fatty acids can be converted to triglycerides and distributed to downstream pathways to be stored in hepatic lipid droplets or to be secreted in VLDL particles (31), where they can contribute respectively to development of steatosis or to levels of circulating lipids. In agreement with this, GCKR rs1260326-T increases risk of fatty liver (12, 14) and raises concentrations of all the apolipoprotein B containing...
lipoprotein particles and amount of triglycerides within all the examined lipoprotein subclasses (Panel 2 in Figure 1). GCKR rs1260326-T associates also with elevated levels of glycolysis related metabolites and circulating fatty acids, as well as increased fatty acid saturation (Panel 2 in Figure 1) compatible with the enhanced glycolytic and lipogenic activities promoted by this variant (30, 32, 33). Further studies are warranted to understand the association between the GCKR variant and aberrations in circulating amino acid concentrations. The LYPLAL1 rs12137855-C variant has a metabolically similar but statistically less robust effect than the GCKR rs1260326-T, as seen in highly correlated effect estimates of the two (Supplementary Figure 3A). This provides supportive evidence for LYPLAL1 functioning in hepatic glucose metabolism as proposed in a study by Ahn et al. where they showed that LYPLAL1 inhibition leads to increase in glucose production in human, rat, and mouse hepatocytes (34).

Pathway II, reduced VLDL secretion: The TM6SF2 protein contributes to VLDL secretion from the liver (35, 36). In mice, knockdown of Tm6sf2 causes threefold increase in hepatic triglycerides content while plasma triglycerides and cholesterol are reduced (15). Conversely, overexpression of human TM6SF2 in mice increases serum lipids while it does not affect liver phenotype (16). The NAFLD risk allele TM6SF2 rs58542926-T is a loss-of-function variant resulting in a misfolded protein undergoing accelerated degradation (15). Supportive to the previous findings, we observed that this allele associates with reduced concentrations of multiple circulating lipid species while it does not seem to influence on qualitative lipid measures, such as fatty acid saturation (Panel 5 in Figure 1). The inverse associations between TM6SF2 rs58542926-T and all the circulating VLDL particle subclass concentrations, lipid species within the lipoprotein subclasses, as well as apolipoprotein B concentration suggest that in humans TM6SF2 rs58542926-T disturbs both lipidation and secretion of VLDL particles. This differs from the observation of a study with Tm6sf2−/− mice, which provided evidence that mouse Tm6sf2 is required for lipidation of VLDL particles, but lack of it does not influence VLDL secretion (37). The present results are compatible with a study that associated rs58542926 T with a favourable plasma lipid profile (19) and its suggested cardioprotective effect (38).
Pathway III, impairment of triglyceride mobilisation from hepatic lipid storage: PNPLA3 is localized in the endoplasmic reticulum and lipid droplet membranes in human liver cells (39) and is suggested to be a multifunctional enzyme having both acyltransferase and hydrolase activities on glycerolipids (40-42). NAFLD risk allele rs738409-G enhances triglyceride retention in hepatic lipid droplets by inhibiting their hydrolysis (39, 40, 43) and seemingly has only minimal, if any, contribution to the metabolic traits in circulation (Panel 4 in Figure 1). Our results are in line with other studies showing that PNPLA3 rs738409-G does not show observable associations with alterations in plasma triglycerides, total cholesterol, HDL-C, LDL-C, nor glucose homeostasis (10, 11), and the present findings extend the same perception to fatty acids, amino acids, and detailed lipoprotein subclass measures. Moreover, our findings are compatible with a mouse model where Pnpla3^{148M/M} knock-in mice show no differences in levels of circulating lipids and glucose in comparison to wild-type mice regardless the increase in liver triglycerides (44). These findings advocate that hepatic lipid accumulation can be neutral for circulatory changes and that the strong cross-sectional associations are likely due to lifestyle-related aspects such as dietary factors, excess energy intake or sedentary lifestyle. Dietary lipids and sugars supply the fatty acid pool upstream from the lipid storage and secretion pathways (31), and consequently overnutrition could contribute to both fatty liver development as well as altered metabolic profile.

This study has some limitations. We examine only a limited number of NAFLD risk alleles while there are multiple other genetic pathways contributing to pathogenesis of NAFLD (17, 45). However, the variants studied here are important determinants of liver fat content (13-15, 17, 18) and therefore the current study setting covers some of the fundamental pathways involved in NAFLD pathogenesis. Regarding the cross-sectional associations, ultrasound lacks sensitivity to detect mild steatosis (46) which reduces our power to determine observational associations thus leading to conservative association magnitudes. In addition, ultrasound cannot distinguish simple steatosis from steatohepatitis and therefore we are unable to characterize the liver phenotype in more detail. Alternative techniques for metabolic profiling, such as a more detailed lipidomics platform (47), may reveal other systemic metabolic biomarker changes that could be associated with steatosis induced by the NAFLD risk alleles showing no association to the metabolic
traits studied here. However, the metabolic measures captured with the current panel show strong associations with fatty liver, and the studied NAFLD risk alleles show divergent association profiles on the corresponding lipid and metabolite levels informing about the heterogeneous nature of fatty liver.

The present study demonstrates that detailed metabolic profiling can provide extensive information on the biological functions of disease-associated genetic variants, and illustrates how omics data can be utilized in evaluation of molecular mechanisms complex traits, such as NAFLD. The divergence in the direction of the genetic association profiles emphasises that NAFLD is a heterogeneous condition and that its impact on circulating metabolic traits varies depending on the pathogenic mechanism. We highlight the barely discernible metabolic consequence of the strongest genetic determinant of NAFLD, PNPLA3 rs738409-G, and the cardioprotective metabolic association profile of the TM6SF2 rs58542926-T. Our findings suggest that hepatic triglyceride accumulation by itself does not necessarily cause metabolic changes increasing the risk of cardiometabolic complications.
References


Author names in bold designate shared co-first authorship.
Figure Legends

Figure 1. Cross-sectional associations of fatty liver with lipoprotein particle subfraction concentrations, lipoprotein particle diameter, apolipoproteins, triglycerides, fatty acids, fatty acid saturation, beta-oxidation, glycolysis and amino acid related metabolites, and the corresponding associations with four NAFLD risk alleles. Cross-sectional associations were determined in 1,810 adults aged 34-49 years of whom 338 were diagnosed with ultrasound-based fatty liver. The metabolic phenotypes were adjusted for age, sex, and ten first genetic principal components prior to analysis. Genetic effects of the NAFLD risk alleles GCKR rs1260326-T, LYPLAL1 rs12137855-C, PNPLA3 rs738409-G and TM6SF2 rs58542926-T were acquired from a metabolomics GWAS including up to 24,925 Europeans (21). Genetic effect estimates were scaled with respect to the NAFLD risk associated with the corresponding locus (12).

Figure 2. The overall match between the metabolic effects of the NAFLD risk alleles and fatty liver. The dashed line shows the linear fit between metabolic changes associated with fatty liver and PNPLA3 rs738409-G (A), TM6SF2 rs58542926-T (B), GCKR rs1260326-T (C), and LYPLAL1 rs12137855-C (D). The grey area indicates the 95% confidence interval for the line. $R^2$ is a measure of goodness of fit.

Figure 3. Relation of the studied NAFLD risk alleles to the main pathways in hepatic triglyceride partitioning. Liver converts carbohydrates to lipids in de novo lipogenesis. Newly synthesized fatty acids enter to the hepatic fatty acid pool which is also supplied by dietary fats and circulating free fatty acids derived mostly from adipose tissue lipolysis or lipoprotein lipase spillover. Fatty acids can be partitioned to oxidative pathway or esterified to triglycerides that can be stored in hepatic lipid droplets or used for VLDL production to be secreted from the liver. Now, GCKR rs1260326-T enhances the lipogenic pathway by providing more substrates for lipogenesis; LYPLAL1 may be functioning on the same hepatic glucose metabolism and lipogenesis related pathway, as the metabolic effects of GCKR rs1260326-T and LYPLAL1 rs12137855-C are highly similar. Because the hepatic fatty acid pool is located upstream from the storage
and secretion pathways, abundance in hepatic fatty acids can contribute to both development of fatty liver and increased production of VLDL. On the contrary, \textit{TM6SF2} rs58542926-T impairs the secretory pathway leading to lipid accumulation into the liver and reduction in levels of circulating lipids and lipoproteins. \textit{PNPLA3} rs738409-G, in turn, enhances triglyceride accumulation to the storage pool by diminishing triglyceride hydrolysis to fatty acids, but does not directly contribute to VLDL secretion, and thus conveys no consequences to circulation.
Supporting Information

Supplementary Figure 1. Correspondence of the metabolic effects of $GCKR$ rs1260326-T with $GCKR$ rs780094-T, and $TM6SF2$ rs58542926-T with $NCAN$ rs2228603-T. The dashed line shows the highly matching metabolic effects between the genotypes at the two loci associated with NAFLD risk.

Supplementary Figure 2. Cross-sectional associations of fatty liver with all the studied 123 metabolic traits, and the corresponding associations of the studied NAFLD risk alleles. The metabolic traits were adjusted for age, sex, and ten first genetic principal components prior to analysis. Effect estimates of $GCKR$ rs1260326-T, $GCKR$ rs780094-T, $LYPLAL1$ rs12137855-C, $PNPLA3$ rs738409-G, $TM6SF2$ rs58542926-T, and $NCAN$ rs2228603-T on metabolic traits were acquired from a publicly available metabolomics GWAS (17). Biomarker abbreviations as in Supplementary Table 1.

Supplementary Figure 3. The overall match between the metabolic effects of the NAFLD risk alleles $GCKR$ rs1260326-T, $LYPLAL1$ rs12137855-C, $PNPLA3$ rs738409-G, and $TM6SF2$ rs58542926-T. The dashed line shows the linear fit between the studied risk alleles, and the grey area indicates the 95% confidence interval for the line.

Supplementary Table 1. NMR based metabolic traits.

Supplementary Table 2. Fatty liver associations with circulating metabolic traits.

Supplementary Table 3. NAFLD risk allele associations with circulating metabolic traits.
# Figures

## Figure 1.

<table>
<thead>
<tr>
<th>Particle concentration</th>
<th>Fatty liver</th>
<th>GCKR rs1260326-T</th>
<th>LYPLAL1 rs12137855-C</th>
<th>PNPLA3 rs736409-G</th>
<th>TM6SF2 rs58542926-T</th>
</tr>
</thead>
<tbody>
<tr>
<td>XXL VLDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XL VLDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L VLDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S VLDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XS VLDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M LDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S LDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XL HDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L HDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S HDL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Particles diameter**

<table>
<thead>
<tr>
<th>VLDL</th>
<th>LDL</th>
<th>HDL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Apolipoproteins**

<table>
<thead>
<tr>
<th>ApoA-I</th>
<th>ApoA-II</th>
<th>ApoB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Triglycerides in**

<table>
<thead>
<tr>
<th>XXL VLDL</th>
<th>XL VLDL</th>
<th>L VLDL</th>
<th>S VLDL</th>
<th>XS VLDL</th>
<th>IDL</th>
<th>LDL</th>
<th>M LDL</th>
<th>S LDL</th>
<th>XL HDL</th>
<th>L HDL</th>
<th>S HDL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fatty acids**

<table>
<thead>
<tr>
<th>Total</th>
<th>Omega-3</th>
<th>Docosahexaenoic acid</th>
<th>Omega-6</th>
<th>Linoleic acid</th>
<th>MUFA</th>
<th>Omega-7, -9 and saturated length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Saturation**

<table>
<thead>
<tr>
<th>Bis-allylic bonds: double bonds</th>
<th>Bis-allylic bonds: total fatty acids</th>
<th>Double bonds in fatty acids</th>
<th>CH₂ groups in fatty acids</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Beta-oxidation**

<table>
<thead>
<tr>
<th>Acetoacetate</th>
<th>Acetate</th>
<th>Beta-hydroxybutyrate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Glycolysis**

<table>
<thead>
<tr>
<th>Glucose</th>
<th>Lactate</th>
<th>Pyruvate</th>
<th>Citrate</th>
<th>Glycerol</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Amino acids**

<table>
<thead>
<tr>
<th>Alanine</th>
<th>Glutamine</th>
<th>Glycine</th>
<th>Histidine</th>
<th>Isoleucine</th>
<th>Leucine</th>
<th>Valine</th>
<th>Phenylalanine</th>
<th>Tyrosine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Effect estimates in SD-units (95% CI)**

- $P \geq 0.002$
- $P < 0.002$

Genetic effects are scaled with the locus associated risk on histologic NAFLD.
Figure 2.

A  Intercept = 0.006, Slope = 0.014, $R^2 = 0.296$

B  Intercept = -0.094, Slope = -0.149, $R^2 = 0.453$

C  Intercept = 0.147, Slope = 0.572, $R^2 = 0.769$

D  Intercept = 0.03, Slope = 0.128, $R^2 = 0.672$
Figure 3.
Tables

Table 1. Description of NAFLD risk increasing genotypes extracted from the open access data. DNA sequence variants studied in the present study were associated with computed tomography characterized steatosis and biopsy-proven NAFLD involving liver inflammation and fibrosis by Speliotes et al. (12). The functional variants explaining the NAFLD associations of NCAN rs2228603 and GCKR rs780094 are denoted separately (*). To achieve coherency, the NCAN locus is referred as TM6SF2 throughout the paper.

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>SNP consequence</th>
<th>EA</th>
<th>Locus OR for histologic</th>
<th>EAF NAFLD (GWAS)</th>
<th>EAF metabolomics (GWAS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PNPLA3</td>
<td>rs738409</td>
<td>I148M</td>
<td>G</td>
<td>3.24 (2.83-3.72)</td>
<td>0.23</td>
<td>0.23</td>
</tr>
<tr>
<td>NCAN</td>
<td>rs2228603</td>
<td>P92S</td>
<td>T</td>
<td>1.90 (1.55-2.34)</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>* TM6SF2</td>
<td>rs58542926</td>
<td>E167K</td>
<td>T</td>
<td>NA</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>LYLPL1</td>
<td>rs12137855</td>
<td>Intergenic</td>
<td>C</td>
<td>1.21 (1.02-1.43)</td>
<td>0.79</td>
<td>0.74</td>
</tr>
<tr>
<td>GCKR</td>
<td>rs780094</td>
<td>Intronic</td>
<td>T</td>
<td>1.18 (1.05-1.34)</td>
<td>0.39</td>
<td>0.37</td>
</tr>
<tr>
<td>* GCKR</td>
<td>rs1260326</td>
<td>P446L</td>
<td>T</td>
<td>NA</td>
<td>0.36</td>
<td></td>
</tr>
</tbody>
</table>

PNPLA3, Patatin-like phospholipase domain containing 3; NCAN, Neurocan; TM6SF2, Transmembrane 6 superfamily member 2; LYLPL1, Lysophospholipase-like 1; GCKR, Glucokinase regulator; SNP, single nucleotide polymorphism; EA, effect allele; OR, odds ratio; NAFLD, non-alcoholic fatty liver disease; EAF, effect allele frequency; NA, not available.
Table 2. Study population: YFS.

<table>
<thead>
<tr>
<th></th>
<th>Fatty liver</th>
<th>No fatty liver</th>
<th>Total population</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>338</td>
<td>1,472</td>
<td>1,810</td>
</tr>
<tr>
<td>Male (%)</td>
<td>237 (70.1)</td>
<td>630 (42.8)</td>
<td>867 (47.9)</td>
</tr>
<tr>
<td>BMI, kg/m²</td>
<td>30.5 ± 5.3</td>
<td>25.5 ± 4.3</td>
<td>26.4 ± 4.9</td>
</tr>
<tr>
<td>Age in years</td>
<td>43.0 ± 4.7</td>
<td>41.6 ± 5.0</td>
<td>41.9 ± 5.0</td>
</tr>
<tr>
<td>Smokers (%)</td>
<td>55 (16.3)</td>
<td>198 (13.5)</td>
<td>253 (14.0)</td>
</tr>
<tr>
<td>Alcohol consumption per day*</td>
<td>2.56 ± 1.7</td>
<td>2.09 ± 1.5</td>
<td>2.18 ± 1.5</td>
</tr>
</tbody>
</table>

* One unit equals to 12 g of alcohol.

Values are mean ± SD, or absolute N count and corresponding percentage.

YFS, Young Finns Study.