

Title Page**Analysis of genotyping for predicting liver injury marker, Procollagen III in persons at risk of non-alcoholic fatty liver disease**

Jane I. Grove^{1,2}, Prarthana Thiagarajan^{1,2}, Stuart Astbury^{1,2}, Rebecca Harris^{1,2}, Toby Delahooke³, I. Neil Guha^{1,2}, Guruprasad P. Aithal^{1,2}.

¹ Nottingham Digestive Diseases Centre, University of Nottingham, Nottingham, UK.

² NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and University of Nottingham, Nottingham, UK.

³ Leicester Royal Infirmary, University Hospitals of Leicester NHS Trust, Leicester, UK.

Corresponding Author:

Dr Jane Grove,
Nottingham Digestive Diseases Centre,
Queens Medical Centre,
Nottingham
NG7 2UH
Tel: 01159709966
E-mail: Jane.Grove@nottingham.ac.uk

Word count: 4778

Figures: 1

Tables: 2

Abbreviations

ALT=alanine transaminase. AST=aspartate transaminase. CI=confidence interval. ELF=enhanced liver function. LDL=low-density lipoprotein. HCC=hepatocellular carcinoma. HDL=high-density lipoprotein. MAF=minor allele frequency. NAFLD=non-alcoholic fatty liver disease. NASH=non-alcoholic steatohepatitis. PCR=polymerase chain reaction. PIIINP=procollagen III amino-terminal peptide. SD=standard deviation.

Conflict of Interest:

There are no conflicts of interest to declare.

Financial Support:

Medical Research Council MR/N005953/1;
NIHR Nottingham Digestive Diseases Biomedical Research Unit;

Academic Health Science Network.

Abstract

Background & Aims

Chronic liver disease presents a major global public health challenge. Stratification of asymptomatic, at-risk patients in primary care using non-invasive methods has the potential to address this by identifying those likely to progress. We therefore evaluated variant alleles at loci associated with non-alcoholic fatty liver disease (NAFLD) as genetic determinants of substantial liver injury in patients with disease risk factors.

Methods

Levels of serum procollagen III (PIIINP), an established fibrosis and steatohepatitis marker, were determined in 467 people who had type 2 diabetes and/or BMI>27.3 (identified from registration with general practitioners) in this observational cross-sectional study. Patients were genotyped for characterised risk alleles in PNPLA3 (rs738409), GCKR (rs1260326) and TM6SF2 (rs58542926) and associations with PIIINP assessed.

Results

The risk alleles in PNPLA3, GCKR or TM6SF2 were not found to be individually associated with the presence of a disease risk factor and were not significantly more common in patients with raised serum PIIINP. The prevalence of possession of both PNPLA3 and GCKR variant alleles combined was significantly higher in at-risk patients with clinically significant liver disease indicated by serum PIIINP above 11 ng/ml ($P=0.014$).

Conclusions

Genotyping therefore has limited value for predicting severe liver disease in at-risk individuals identified in a community setting.

Abstract word count: 205

Key words: PNPLA3, GCKR, TM6SF2, NAFLD

Key points:

- We evaluated the potential clinical utility of genotyping at 3 loci in the stratification of asymptomatic individuals for chronic liver disease in the community.
- Risk alleles in PNPLA3, GCKR or TM6SF2 were not individually associated with risk factors for non-alcoholic fatty liver disease, type 2 diabetes, BMI>27.3) or significantly higher levels of the fibrosis biomarker, PIIINP.
- Cross-sectional analysis showed at-risk individuals with high levels of PIIINP were more likely to be carriers of both PNPLA3 and GCKR risk alleles in combination.
- We conclude that combined PNPLA3 and GCKR alleles may be considered for incorporating genetic risk into diagnostic or prognostic algorithms.

Introduction

Chronic liver disease presents a major global public health challenge: liver-related deaths are rising in the UK (1) with liver cirrhosis being the third most common cause of premature death in people aged below 55 years (2). Given that many culprits in the aetiopathogenesis of chronic liver disease (alcohol, obesity and insulin resistance) are potentially preventable, there exists an unmet need to risk-stratify individuals prior to the onset of established cirrhosis, in order to introduce effective intervention. Risk stratification is problematic in early stages of liver disease as the majority of affected patients are asymptomatic. The absence of a validated non-invasive biomarker to reliably diagnose or exclude liver disease has given rise to the use of a variety of strategies in order to optimise its detection (3); reviewed by (4).

Non-invasive methods for quantifying liver fibrosis, such as the enhanced liver fibrosis (ELF) serum biomarker panel, represent a major advance in availability of strategies with which to assess patients at risk of liver disease in the community (5, 6). In particular, measurement of amino-terminal peptide of procollagen III (PIIINP), a component of this panel, has been established for determining hepatic fibrosis in methotrexate therapy (7) and was identified as an accurate biomarker in discriminating between early-stage disease and more severe steatohepatitis (8).

Non-alcoholic fatty liver disease (NAFLD) comprises a spectrum of disease severity ranging from simple steatosis ('fatty liver') through steatohepatitis to cirrhosis. Genetic variants, which predispose to NAFLD have been identified from genome-wide association studies (reviewed by (9, 10). Three separate single nucleotide polymorphism functional variants, shown to confer increased susceptibility to hepatic

triglyceride accumulations and fibrogenesis, have been robustly validated in large independent cohorts and supporting evidence provided by mechanistic studies. The rs738409 (C>G) variant in adiponutrin (PNPLA3), has consistently been demonstrated to correlate with increased severity of NAFLD (10-15). An association has also been established between the rs1260326 (C>T) variant within the glucokinase regulator (GCKR) and steatosis and fibrosis in NAFLD (16, 17). More recently, variant rs58542926 (C>T) in transmembrane 6 superfamily 2 (TM6SF2) was found to be associated with NAFLD (18-20).

In this observational, cross-sectional study, we determined the genotype at these 3 loci in 467 asymptomatic adults with risk factors for NAFLD, identified in primary care. The observed allele frequencies in patients were compared to literature values and to the frequencies in a geographically similar cohort of biopsy-proven NAFLD patients. We then assessed the influence of common genetic variants in PNPLA3 (genotype GG), GCKR (genotype TT) and TM6SF2 (genotype CT or TT) on the degree of liver fibrosis as indicated by the serum biomarker PIIINP. Additionally, we analysed the effects of possession of two separate risk alleles in combination on the likelihood of clinically significant disease.

Patients and methods

Primary care cohort:

Asymptomatic adult patients were recruited from a general medical practice in Leicester, UK, with a total patient population of 4600 (4150 adults). Patients were identified from electronic primary care patient records if they possessed at least one of the following risk factors for NAFLD (1) Type 2 diabetes mellitus; (2) BMI >27.3

recorded in the past 5 years, and were invited to attend a risk stratification pathway for chronic disease previously described (21-23). The BMI cut-off was selected *a priori* based on the high prevalence of patients registered at the practice with Asian ethnicity. This is in accordance with recommendations by the World Health Organisation due to the higher risk of type 2 diabetes and cardiovascular disease at a lower BMI in Asians compared to other European groups (23, 24). Patients with type 2 diabetes were identified from primary care records; if no coding was given, it was assumed that this factor was not present. The study protocol was approved by the National Research Ethics Service Committee East Midlands-Leicester (Ref: 13/EM/0123); written informed consent was obtained from all participants.

Non-invasive analyses for at-risk patients:

Demographic data were recorded, and blood samples were taken for DNA extraction, clinical testing and biomarker analysis at the time of study visit. Serum was prepared from coagulated blood and stored at -80°C within 4 hours for ELF testing by iQur (London, UK). EDTA-anticoagulated whole blood stored at -80°C was used for DNA extraction.

NAFLD cohort:

DNA samples available from a cohort of biopsy-proven NAFLD patients were also included for cross-sectional analysis. These patients were recruited consecutively from hepatology clinics at the Nottingham University Hospitals NHS Trust between 2002 and 2010. NAFLD was diagnosed on the basis of the following criteria: a liver

biopsy showing histology consistent with NAFLD; a weekly ethanol consumption of <14 units for women and <21 units for men; appropriate exclusion of other causes of liver disease including alcohol, drugs, autoimmune or viral hepatitis, or cholestatic or metabolic/genetic liver disease. These were collected as part of a separate study approved by the Nottingham Research Ethics Committee 2 (Ref: GM010201) described previously (25, 26); all participants gave informed written consent.

Genotyping:

Genomic DNA was isolated using Flexigene Kit (Qiagen, UK). Restriction fragment length polymorphism methodology was used to determine the genotype at PNPLA3 rs738409, GCKR rs1260326 and TM6SF2 rs58542926 using primer pairs:

TGGAGAAAGCTTATGAAGGATCAG/ CAGCTGTGGCTACTCTGTCTG,
GGGTCTTAGGGTACCTGCTCAGAGG/ GGTAACCCATGACCTTGCCCAGC, and
CCTGGGCTCAAGCGATCCTCATGC /

CCAGGTATAATCCCAGCTACTCAGGAGG, respectively. The PCR amplified products were digested with FokI (PNPLA3), MspI (GCKR) and Hpy188I (TM6SF2), to identify wild type and variant alleles (New England Biolabs, UK).

Statistics:

Descriptive statistics were computed for all variables using arithmetic and geometric means (for non-normally distributed data) along with their respective standard deviations (SD) and 95% confidence intervals (CI) using SPSS (v. 22.0). The differences between subgroups were explored using chi-squared test for categorical

variables and t-test and Mann-Whitney test (for 2 groups) or Kruskal-Wallis test (for multiple groups) for continuous normally distributed and non-normally distributed data, respectively. Chi-square test was used to compare allele and genotype frequencies in patient sub-groups and to assess genotype distributions for Hardy-Weinberg equilibrium. Correlation coefficients (Spearman's rho) were used to determine the strength of the relationship of the variables. *P* values < 0.05 were considered statistically significant.

Results

Study population phenotype

The primary practice comprised 4150 adults of which 51.8% were male. The patients' ethnicity was poorly recorded within the electronic records but of the patients who attended the risk stratification pathway 69.9% were Caucasian and 23.2% were of Asian ethnicity which is similar to the proportions seen with the local census data records. Within the practice population 298 patients were identified with type 2 diabetes and 1081 patients had a BMI >27.3. BMI values were available for 87.7% of the patients. The type 2 diabetes prevalence of 7.2% in the practice is similar to the UK national prevalence of 6% (27). Obesity (BMI \geq 30) was less common in the cohort at 17.5% compared to 27% in the general UK population. Of the patients who attended the risk stratification pathway 467 with type 2 diabetes and/or BMI>27.3 provided a blood sample for genotyping DNA and determination of serum PIIINP. The characteristics of this cohort and the practice population are shown in Supplemental Table 1.

The baseline demographic and clinical data for the at-risk cohort is summarised in Table 1. In the genotyped cohort, a total of 61 patients (13%) had PIIINP levels above the 11 ng/ml threshold considered indicative of liver fibrosis grade \geq F3 (8). Serum PIIINP levels correlated significantly with the presence of systemic hypertension, aspartate aminotransferase (AST) and BMI (data not shown). There were no significant differences in serum PIIINP levels between the Caucasian and Asian sub-populations (Supplemental Table 2).

Genotypes and allele frequencies for PNPLA3, GCKR and TM6SF2 risk alleles

The genotype for PNPLA3 (rs738409), GCKR (rs1260326) and TM6SF2 (rs58542926) was determined in the at-risk primary care cohort and a biopsy-proven NAFLD patient cohort. Table 2 shows the minor allele frequency (MAF) for each variant in each cohort, sub-groups and the European Population MAF from the 1000 Genomes project (28). The genotype at the 3 loci, detailed in Supplemental Table 3, was in Hardy-Weinberg equilibrium for the tested cohorts (and in the sub groups analysed where >5 individuals with each genotype were included).

Within the community at-risk cohort, the Caucasian sub-group had similar allele frequencies (Table 2) to other European populations (19, 28, 29). There were no significant differences in the PNPLA3, GCKR and TM6SF2 variant frequencies between the Caucasian and Asian sub-groups however, possession of the GCKR T allele was less common in Asians as in the reference populations.

Higher frequencies of PNPLA3, GCKR and TM6SF2 risk alleles in biopsy-proven NAFLD patients than in the at-risk patient cohort

The individual PNPLA3, GCKR, TM6SF2 risk allele frequencies in the at-risk primary care cohort were compared to the frequencies in a cohort of 186 predominantly Caucasian, biopsy-proven NAFLD patients (age 58.0 (\pm 12.7), 61.5% male) also from central England (8, 26) (Table 2). All 3 of these alleles were significantly more prevalent in the biopsy-proven disease cohort than the at-risk cohort: 32% of NAFLD patient chromosomes had the PNPLA3 variant, compared to 22% of primary care individuals with NAFLD risk factors ($P=0.01$). This corresponds to risk allele carriage in 53% of NAFLD hospital patients and in 39% of at-risk community patients (Supplemental Table 3). In diseased patients, 49% of chromosomes possessed the GCKR variant and 15% had the TM6SF2 variant compared to 36% and 10% of chromosomes in the NAFLD at-risk cohort, respectively ($P<0.001$ and $P=0.04$, respectively). This is consistent with previous reports of associations of these so-called risk alleles with NAFLD progression to non-alcoholic steatohepatitis (NASH) (16, 18, 30). The observation that the at-risk cohort has lower risk allele frequencies than a diseased cohort, and has frequencies similar to that in the 1000 Genomes reference populations, demonstrates that the variants are not associated with possession of risk factors (being overweight or having diabetes), only with the development of liver disease.

Association of combined PNPLA3 plus GCKR variant alleles with high PIIINP levels

We then assessed the prevalence of risk alleles in at-risk patients sub-divided using a cut-off of PIIINP levels above 11 ng/ml, considered to indicate significant fibrosis \geq F3 and NASH diagnosis (8) as shown in Fig. 1. Individually, none of the risk alleles was significantly more prevalent in the 61 patients with elevated, clinically significant levels of PIIINP. As the PNPLA3 and GCKR variant alleles have been reported to interact to contribute to increased severity of NAFLD (17), we considered these two variants in combination (Fig. 1D). Analysis revealed a significantly higher proportion with the double variant PNPLA3 G* GCKR T* combination: 36% compared to 22% in those with PIIINP levels below this cut off ($P=0.014$). This double variant loci combination was significantly more prevalent in the patients with high serum PIIINP levels. In total, 110 (24%) at-risk patients had one or more risk alleles for both PNPLA3 and GCKR loci; 36 had PNPLA3 G* plus TM6SF2 T*. Since combinations including the TM6SF2 variant are less prevalent, numbers in our cohort were insufficient for analysis of this dual variant combination.

In the NAFLD risk community cohort there were no significant associations between each separate risk allele (at the 3 loci), and PIIINP levels indicative of extracellular matrix remodelling associated with fibrosis. We found ALT levels were significantly higher ($P<0.001$) in NAFLD risk patients carrying one or more copies of the PNPLA3 variant allele (29.6 (95% CI: 6-143) versus 24.8 (95% CI: 7-93) IU/l) as reported previously for an obese cohort (31). If only our 240 obese, NAFLD-risk patients were considered, we similarly found a significant association between possession of the PNPLA3 variant and serum PIIINP levels ($P=0.040$; 8.43 ng/ml (95% CI: 3-25) in carriers versus 7.61 ng/ml (95% CI: 2-24) in non-carriers). Levels of PIIINP were

also higher in PNPLA3 variant carriers with either BMI>27.3 ($n=406$) or diabetes ($n=120$), but this did not reach significance for either group.

Discussion

Heritable components interact with environmental influences, such as diet and alcohol intake, generating the observed individual disease susceptibility and inter-individual variability of complex traits (32). In this study we have assessed the value of genotyping at 3 separate loci previously implicated in the development of chronic liver disease associated with common, lifestyle-related risk factors in a community setting. We have also been able to consider an interaction between two prevalent risk alleles in combination (PNPLA3 G and GCKR T), in contributing to serum PIIINP levels, a surrogate biomarker for liver fibrosis.

Non-invasive tools for estimation of fibrosis in patients are now widely available and applicable for population stratification in primary care (21). PIIINP released into circulation is thought to reflect collagen synthesis associated with extracellular matrix turnover at the site of disease, indicative of fibrosis. Serum PIIINP levels are widely utilised for monitoring liver fibrosis (7, 33). Transient elastography using Fibroscan© similarly correlates with histological fibrosis but may not be informative for fibrosis staging in as many patients compared to the ELF biomarker panel (5). PIIINP levels were previously established for diagnosis of liver injury characteristic of NASH and were associated with steatosis, lobular inflammation and ballooning histology (8). In our study, a threshold of PIIINP 11 ng/ml was suggestive of significant liver injury or severe levels of fibrosis in 61 patients within the at-risk cohort of 467 tested.

We confirm that the PNPLA3 variant is significantly more prevalent in diagnosed, secondary care NAFLD patients, as previously reported in several populations and established as a risk factor for disease in meta-analyses (12, 14, 15, 30, 34, 35). Here we additionally report increased prevalence of GCKR and TM6SF2 variants in biopsy-proven NAFLD patients (Table 2). These findings are consistent with cohorts of patients identified in secondary care practice being more likely to have more advanced disease when compared to those in the primary care setting. In addition, in secondary care, patients are selected to liver biopsy based on the algorithms that identify those with more advanced liver disease.

Numerous studies have suggested PNPLA3 genotyping is of value in the risk stratification of NAFLD patients (reviewed by (20, 36). The PNPLA3 variant was shown to confer increased risk of NAFLD in 'healthy' Korean males (37). However, one study found PNPLA3 genotyping to be of little value in detecting the presence of NAFLD (38) but it was suggested to have clinical utility in detecting NASH (39) and may assist in stratification for HCC surveillance (40). We did not find an association of this variant with non-invasive measures of fibrosis in patients at risk of NAFLD. However, when only obese patients were evaluated, the variant was associated with serum PIIINP indicative of more advanced fibrosis consistent with previous research (31, 41, 42).

Similar to in general population studies (43), we found that serum triglyceride levels were elevated in at-risk individuals carrying the GCKR variant (but this did not reach statistical significance; data not shown), consistent with its effect increasing hepatic glucokinase activity (44) in at-risk individuals. A synergistic effect of PNPLA3 and

GCKR variants was also previously reported associated with susceptibility to NAFLD in an Indian cohort and the multiplicative mechanistic effect suggested to arise via overlapping functionality in lipid processing pathways (17). Our observed increased prevalence of this variant combination in at-risk patients with predicted clinically significant fibrosis (Figure 1) is consistent with these genetic factors acting together to promote the development of disease.

The main limitation of this study is that it cannot determine the key relationship between genetic background and liver disease progression over time as this would necessitate a longitudinal analysis rather than a cross-sectional study. Although our cohort is substantial, numbers are still restrictive for analysis of separate risk factors. The cohort includes people with diabetes or high BMI, or both and 61 people who also consume levels of alcohol above the National recommended limits (14 units per week) so there is heterogeneity in likely pathogenic processes which may distort our findings. There is also uncertain applicability to other cohorts with different ethnic make-up. Our cohort is reasonably representative of the UK population in terms of diabetes prevalence and alcohol consumption although obesity levels were lower than expected; it also included a substantial number of people with Asian ethnicity which have higher rates of NAFLD. However, the key strength is that the current study does overcome the selection bias of secondary care setting studies and uses the more relevant primary care patients which could be targeted for treatment.

Ever since genetic factors were established as contributing to disease progression, genotyping has been proposed as a potential route to stratify patients for treatments. This report highlights the complexity of the disease mechanisms and likely

interaction between individual genes and lifestyle factors, making it difficult to establish a feasible screening process. Although possession of risk alleles is thought to predispose to disease severity, this research reveals that there is no simple relationship with degree of fibrosis in asymptomatic individuals with NAFLD risk factors. Further work including replication in other cohorts or incorporation into marker algorithms, may reveal that targeted genotyping of specific risk groups could be of value to identify asymptomatic patients likely to have significant levels of liver fibrosis or steatohepatitis for targeting interventions.

Acknowledgements:

The views expressed are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research or the Department of Health. The study sponsors were not involved in this research or manuscript submission. We acknowledge Tim Card for his role in conceiving, designing and developing the community cohort study. We thank Melanie Lingaya and Yirga Falcone for excellent technical assistance and Digestive Diseases Research nurses for assisting with sample collection.

References:

1. BHALA N, AITHAL G, FERGUSON J. How to tackle rising rates of liver disease in the UK. *BMJ* 2013; 346: f807.
2. MURRAY C J, LOPEZ A D. Measuring the global burden of disease. *The New England journal of medicine* 2013; 369(5): 448-57.
3. ERGELEN R, AKYUZ U, AYDIN Y, EREN F, YILMAZ Y. Measurements of serum procollagen-III peptide and M30 do not improve the diagnostic accuracy of transient elastography for the detection of hepatic fibrosis in patients with nonalcoholic fatty liver disease. *European journal of gastroenterology & hepatology* 2015; 27(6): 667-71.
4. ARULANANDAN A, LOOMBA R. Non-invasive Testing for NASH and NASH with Advanced Fibrosis: Are We There Yet? *Current hepatology reports* 2015; 14(2): 109-18.
5. FRIEDRICH-RUST M, ROSENBERG W, PARKES J, HERRMANN E, ZEUZEM S, SARRAZIN C. Comparison of ELF, FibroTest and FibroScan for the non-invasive assessment of liver fibrosis. *BMC gastroenterology* 2010; 10: 103.
6. WAHL K, ROSENBERG W, VASKE B, et al. Biopsy-controlled liver fibrosis staging using the enhanced liver fibrosis (ELF) score compared to transient elastography. *PLoS one* 2012; 7(12): e51906.
7. KHAN S, SUBEDI D, CHOWDHURY M M. Use of amino terminal type III procollagen peptide (P3NP) assay in methotrexate therapy for psoriasis. *Postgraduate medical journal* 2006; 82(967): 353-4.
8. TANWAR S, TREMBLING P M, GUHA I N, et al. Validation of terminal peptide of procollagen III for the detection and assessment of nonalcoholic steatohepatitis in patients with nonalcoholic fatty liver disease. *Hepatology* 2013; 57(1): 103-11.
9. KAHALI B, HALLIGAN B, SPELIOTES E K. Insights from Genome-Wide Association Analyses of Nonalcoholic Fatty Liver Disease. *Seminars in liver disease* 2015; 35(4): 375-91.
10. ANSTEE Q M, SETH D, DAY C P. Genetic Factors That Affect Risk of Alcoholic and Nonalcoholic Fatty Liver Disease. *Gastroenterology* 2016; 150(8): 1728-44 e7.
11. ROMEO S, KOZLITINA J, XING C, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nature genetics* 2008; 40(12): 1461-5.
12. SOOKOIAN S, PIROLA C J. Meta-analysis of the influence of I148M variant of patatin-like phospholipase domain containing 3 gene (PNPLA3) on the susceptibility and histological severity of nonalcoholic fatty liver disease. *Hepatology* 2011; 53(6): 1883-94.
13. TREPO E, ROMEO S, ZUCMAN-ROSSI J, NAHON P. PNPLA3 gene in liver diseases. *Journal of hepatology* 2016; 65(2): 399-412.
14. VALENTI L, AL-SERRI A, DALY A K, et al. Homozygosity for the patatin-like phospholipase-3/adiponutrin I148M polymorphism influences liver fibrosis in patients with nonalcoholic fatty liver disease. *Hepatology* 2010; 51(4): 1209-17.
15. SHEN J H, LI Y L, LI D, WANG N N, JING L, HUANG Y H. The rs738409 (I148M) variant of the PNPLA3 gene and cirrhosis: a meta-analysis. *Journal of lipid research* 2015; 56(1): 167-75.
16. PETTA S, MIELE L, BUGIANESI E, et al. Glucokinase regulatory protein gene polymorphism affects liver fibrosis in non-alcoholic fatty liver disease. *PLoS one* 2014; 9(2): e87523.
17. TAN H L, ZAIN S M, MOHAMED R, et al. Association of glucokinase regulatory gene polymorphisms with risk and severity of non-alcoholic fatty liver disease: an interaction study with adiponutrin gene. *Journal of gastroenterology* 2014; 49(6): 1056-64.
18. LIU Y L, REEVES H L, BURT A D, et al. TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. *Nature communications* 2014; 5: 4309.

19. KOZLITINA J, SMAGRIS E, STENDER S, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nature genetics* 2014; 46(4): 352-6.
20. DONGIOVANNI P, ROMEO S, VALENTI L. Genetic Factors in the Pathogenesis of Nonalcoholic Fatty Liver and Steatohepatitis. *BioMed research international* 2015; 2015: 460190.
21. HARMAN D J, RYDER S D, JAMES M W, et al. Direct targeting of risk factors significantly increases the detection of liver cirrhosis in primary care: a cross-sectional diagnostic study utilising transient elastography. *BMJ open* 2015; 5(4): e007516.
22. TANAJEWSKI L, HARRIS R, HARMAN D, et al. Earlier Detection And Treatment of Non-Alcoholic Fatty Liver Disease: An Economic Evaluation to Appraise An Innovative Diagnostic Pathway to Detect And Intervene Where There Are Known Risk Factors. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research* 2015; 18(7): A627.
23. HARRIS R, DELAHOKE T, CARD T, GUHA I, AITHAL G. OC-032 Type 2 Diabetes and Obesity will Drive the Burden of Liver Cirrhosis in the Next Two Decades. *Gut* 2016; 65(Suppl 1): A19-A20.
24. CONSULTATION W H O E. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* 2004; 363(9403): 157-63.
25. AITHAL G P, THOMAS J A, KAYE P V, et al. Randomized, placebo-controlled trial of pioglitazone in nondiabetic subjects with nonalcoholic steatohepatitis. *Gastroenterology* 2008; 135(4): 1176-84.
26. GROVE J I, ANTOINE D J, KAYE P, et al. Do Serum Markers of Cell Injury and Death Have Potential to Become Mechanistic Markers in Non-Alcoholic Fatty Liver Disease (Nafld)? *Gut* 2014; 63: A245-A46.
27. GATINEAU M. H C, HOLMAN N., OUTHWAITE H., OLDRIDGE L. C A A E L. Adult obesity and type 2 diabetes. *Public Health England 2014(PHE publications gateway number: 2014211)*.
28. GENOMES PROJECT C, AUTON A, BROOKS L D, et al. A global reference for human genetic variation. *Nature* 2015; 526(7571): 68-74.
29. PALMER C N, MAGLIO C, PIRAZZI C, et al. Paradoxical lower serum triglyceride levels and higher type 2 diabetes mellitus susceptibility in obese individuals with the PNPLA3 148M variant. *PloS one* 2012; 7(6): e39362.
30. SPELIOTES E K, BUTLER J L, PALMER C D, et al. PNPLA3 variants specifically confer increased risk for histologic nonalcoholic fatty liver disease but not metabolic disease. *Hepatology* 2010; 52(3): 904-12.
31. ROMEO S, SENTINELLI F, DASH S, et al. Morbid obesity exposes the association between PNPLA3 I148M (rs738409) and indices of hepatic injury in individuals of European descent. *Int J Obes (Lond)* 2010; 34(1): 190-4.
32. LOOMBA R, SCHORK N, CHEN C H, et al. Heritability of Hepatic Fibrosis and Steatosis Based on a Prospective Twin Study. *Gastroenterology* 2015; 149(7): 1784-93.
33. XIE Q, ZHOU X, HUANG P, WEI J, WANG W, ZHENG S. The performance of enhanced liver fibrosis (ELF) test for the staging of liver fibrosis: a meta-analysis. *PloS one* 2014; 9(4): e92772.
34. SOOKOIAN S, CASTANO G O, BURGUENO A L, GIANOTTI T F, ROSSELLI M S, PIROLA C J. A nonsynonymous gene variant in the adiponutrin gene is associated with nonalcoholic fatty liver disease severity. *Journal of lipid research* 2009; 50(10): 2111-6.
35. ROTMAN Y, KOH C, ZMUDA J M, KLEINER D E, LIANG T J, NASH C R N. The association of genetic variability in patatin-like phospholipase domain-containing protein 3 (PNPLA3) with histological severity of nonalcoholic fatty liver disease. *Hepatology* 2010; 52(3): 894-903.
36. XU R, TAO A, ZHANG S, DENG Y, CHEN G. Association between patatin-like phospholipase domain containing 3 gene (PNPLA3) polymorphisms and nonalcoholic fatty liver disease: a HuGE review and meta-analysis. *Scientific reports* 2015; 5: 9284.

37. PARK J H, CHO B, KWON H, et al. I148M variant in PNPLA3 reduces central adiposity and metabolic disease risks while increasing nonalcoholic fatty liver disease. *Liver international : official journal of the International Association for the Study of the Liver* 2015; 35(12): 2537-46.
38. KOTRONEN A, PELTONEN M, HAKKARAINEN A, et al. Prediction of non-alcoholic fatty liver disease and liver fat using metabolic and genetic factors. *Gastroenterology* 2009; 137(3): 865-72.
39. HYYSALO J, MANNISTO V T, ZHOU Y, et al. A population-based study on the prevalence of NASH using scores validated against liver histology. *Journal of hepatology* 2014; 60(4): 839-46.
40. LIU Y L, PATMAN G L, LEATHART J B, et al. Carriage of the PNPLA3 rs738409 C >G polymorphism confers an increased risk of non-alcoholic fatty liver disease associated hepatocellular carcinoma. *Journal of hepatology* 2014; 61(1): 75-81.
41. PETAJA E M, YKI-JARVINEN H. Definitions of Normal Liver Fat and the Association of Insulin Sensitivity with Acquired and Genetic NAFLD-A Systematic Review. *International journal of molecular sciences* 2016; 17(5).
42. STENDER S, KOZLITINA J, NORDESTGAARD B G, TYBJAERG-HANSEN A, HOBBS H H, COHEN J C. Adiposity amplifies the genetic risk of fatty liver disease conferred by multiple loci. *Nature genetics* 2017; 49(6): 842-47.
43. SPARSO T, ANDERSEN G, NIELSEN T, et al. The GCKR rs780094 polymorphism is associated with elevated fasting serum triacylglycerol, reduced fasting and OGTT-related insulinaemia, and reduced risk of type 2 diabetes. *Diabetologia* 2008; 51(1): 70-5.
44. REES M G, WINCOVITCH S, SCHULTZ J, et al. Cellular characterisation of the GCKR P446L variant associated with type 2 diabetes risk. *Diabetologia* 2012; 55(1): 114-22.

Table 1. Demographic and clinical characteristics of genotyped community patients with NAFLD risk factors.

Variables ^a	Genotyped at-risk community cohort (<i>n</i> =467)
Age: mean (SD)	57.1 (± 14.1)
Males (%)	52.7 %
Caucasians (%)	71.0%
ALT (IU/L)	26.57 (6.24-113.0) <i>n</i> =454
Triglycerides (mmol/L)	1.49 (0.28-7.8) <i>n</i> =453
LDL cholesterol (mmol/L)	2.78 (± 1.0) <i>n</i> =415
HDL cholesterol (mmol/L)	1.32 (0.53-3.28) <i>n</i> =455
BMI (kg/m ²)	30.81 (19.5-48.8)
BMI >27.3 (%)	87.6 %
Obese (%)	51.8 %
Type 2 diabetes (%)	26.1 %
PIIINP (ng/ml)	7.57 (2.55-22.52)
PIIINP >11 ng/ml (likely fibrosis grade ≥3)	13.1 %

^a Variables are expressed as geometric mean ± 95% CI, mean ± SD, or frequency (%) and number of cases. ALT= alanine transaminase. AST= aspartate transaminase. LDL= low-density lipoprotein. HDL= high-density lipoprotein. PIIINP= pro-collagen III.

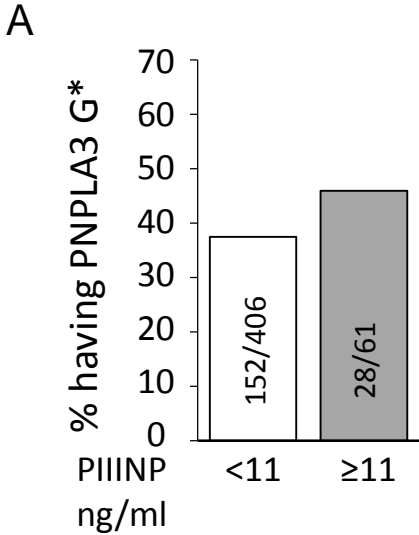
Table 2 Comparison of individual minor allele frequencies (MAF) in patient groups and populations.

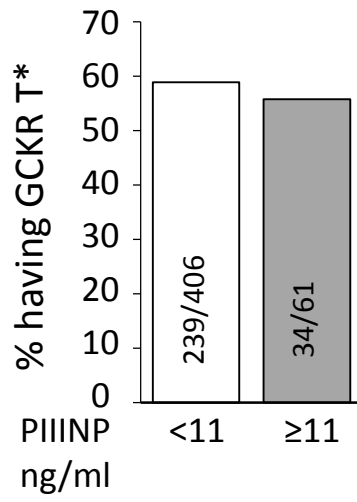
Cohort	PNPLA3 rs738409 MAF	GCKR rs1260326 MAF	TM6SF2 rs58542926 MAF
At-risk community cohort (<i>n</i> =467)	0.22	0.36	0.10
Biopsy-proven NAFLD hospital cohort (<i>n</i> =186)	0.32	0.49	0.15
Caucasian community patients (<i>n</i> =331)	0.22	0.38	0.10
Asian community patients (<i>n</i> =108)	0.22	0.32	0.10
1000 genomes European cohort ^a	0.23	0.41	0.07
1000 genomes South Asian cohort ^a	0.25	0.20	0.11

^a (28) <http://www.1000genomes.org/category/population>

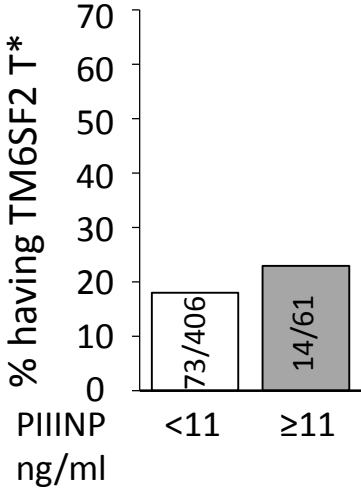
Figure Legend:

Fig. 1 Distribution of risk allele genotypes in at-risk community patients with clinically significant serum PIIINP levels compared to those with levels of PIIINP below 11ng/ml. (A) PNPLA3. (B) GCKR. (C) TM6SF2. (D) Combined PNPLA3 and GCKR variants.



B

C



D

