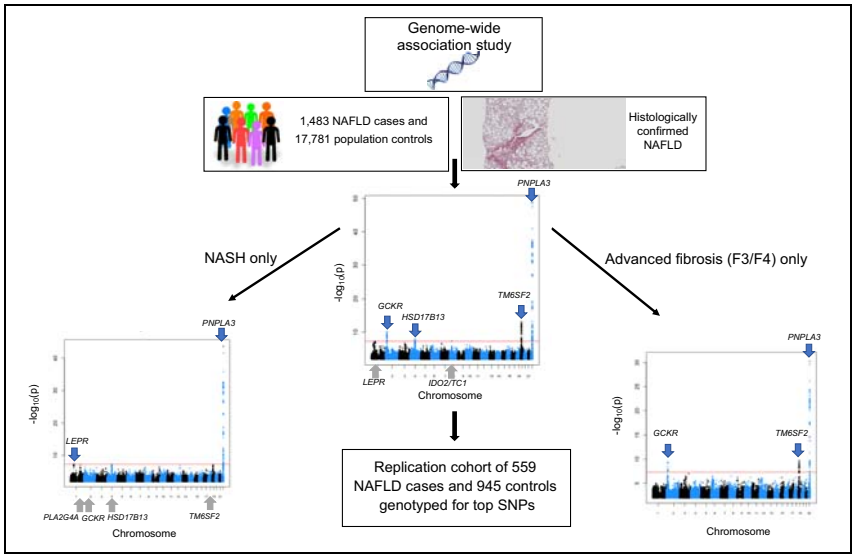
Research Article
NAFLD and Alcohol-Related Liver Diseases

JOURNAL OF HEPATOLOGY

Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort

Graphical abstract



Highlights

- Genome-wide association study involved 1,483 biopsied NAFLD cases and 17,781 controls.
- Main analysis shows genome-wide significance for *PNPLA3*, *TM6SF2*, *HSD17B13* and *GCKR*.
- Sub-analyses show significance near *LEPR* for NASH and near *PYGO1* for steatosis.
- Except for *GCKR*, the genome-wide significant signals were replicated.

Authors

Quentin M. Anstee, Rebecca Darlay, Simon Cockell, …, Christopher P. Day, Heather J. Cordell, Ann K. Daly

Correspondence

a.k.daly@ncl.ac.uk (A.K. Daly), quentin.anstee@ncl.ac.uk (Q.M. Anstee).

Lay summary

Non-alcoholic fatty liver disease is a common disease where excessive fat accumulates in the liver and may result in cirrhosis. To understand who is at risk of developing this disease and suffering liver damage, we undertook a genetic study to compare the genetic profiles of people suffering from fatty liver disease with genetic profiles seen in the general population. We found that particular sequences in 4 different areas of the human genome were seen at different frequencies in the fatty liver disease cases. These sequences may help predict an individual's risk of developing advanced disease. Some genes where these sequences are located may also be good targets for future drug treatments.

source: https://doi.org/10.7892/boris.146469 | downloaded: 12.10.2024

https://doi.org/10.1016/j.jhep.2020.04.003
© 2020 European Association for the Study of the Liver. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/). J. Hepatol. 2020, 73, 505–515

Research Article
NAFLD and Alcohol-Related Liver Diseases

JOURNAL OF HEPATOLOGY

Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort☆

Quentin M. Anstee[1,2,*], Rebecca Darlay[3], Simon Cockell[4], Marica Meroni[5], Olivier Govaere[1], Dina Tiniakos[1,6], Alastair D. Burt[1,7], Pierre Bedossa[1], Jeremy Palmer[1], Yang-Lin Liu[1], Guruprasad P. Aithal[8], Michael Allison[9], Hannele Yki-Järvinen[10], Michele Vacca[9,11], Jean-Francois Dufour[12], Pietro Invernizzi[13,21], Daniele Prati[5], Mattias Ekstedt[14], Stergios Kechagias[14], Sven Francque[15], Salvatore Petta[16], Elisabetta Bugianesi[17], Karine Clement[18], Vlad Ratziu[19], Jörn M. Schattenberg[20], Luca Valenti[5], Christopher P. Day[1], Heather J. Cordell[3], Ann K. Daly[1,*], on behalf of the EPoS Consortium Investigators[#]

[1]Translational & Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom; [2]Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, United Kingdom; [3]Population & Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom; [4]Bioinformatics Support Unit, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom; [5]Department of Pathophysiology and Transplantation, University of Milan, Translational Medicine - Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy; [6]Department of Pathology, Aretaieio Hospital, National & Kapodistrian University of Athens, Greece; [7]Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide, Australia; [8]NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and University of Nottingham, Nottingham, UK; [9]Liver Unit, Department of Medicine, Cambridge Biomedical Research Centre, Cambridge University NHS Foundation Trust, United Kingdom; [10]Department of Medicine, University of Helsinki, Helsinki, Finland & Helsinki University Hospital, Helsinki, Finland; [11]Department of Biochemistry and Wellcome Trust/MRC Institute of Metabolic Science, MRC Metabolic Diseases Unit, Metabolic Research Laboratories, University of Cambridge, UK; [12]University Clinic for Visceral Surgery and Medicine, University of Berne, Freiburgstrasse, Berne 3010, Switzerland; [13]Division of Gastroenterology and Center for Autoimmune Liver Diseases, Department of Medicine and Surgery, University of Milano - Bicocca, Monza, Italy; [14]Division of Gastroenterology and Hepatology, Department of Medicine and Health Sciences, Linköping University, Linkoping, Sweden; [15]Department of Gastroenterology and Hepatology, Antwerp University Hospital, Antwerp, Belgium; [16]Sezione di Gastroenterologia, Dipartimento Promozione della Salute, Materno-Infantile, di Medicina Interna e Specialistica di Eccellenza "G. D'Alessandro", Università di Palermo, Palermo, Italy; [17]Department of Medical Sciences, Division of Gastro-Hepatology, A.O. Città della Salute e della Scienza di Torino, University of Turin, Turin, Italy; [18]Sorbonne University, Inserm, Nutrition and obesity: Systemic approaches, Nutrition department, Pitié-Salpêtrière hospital, Assistance Publique-Hôpitaux de Paris, 75013 Paris, France; [19]Sorbonne Université, Assistance Publique-Hôpitaux de Paris, Hôpital Pitié-Salpêtrière, Institute of Cardiometabolism and Nutrition (ICAN), Paris, France; [20]NAFLD Research Center, Department of Medicine, University Medical Center of the Johannes Gutenberg University, Mainz, Germany; [21]European Reference Network on Hepatological Diseases (ERN RARE-LIVER), San Gerardo Hospital, Monza, Italy

Background & Aims: Genetic factors associated with non-alcoholic fatty liver disease (NAFLD) remain incompletely understood. To date, most genome-wide association studies (GWASs) have adopted radiologically assessed hepatic triglyceride content as the reference phenotype and so cannot address steatohepatitis or fibrosis. We describe a GWAS encompassing the full spectrum of histologically characterised NAFLD.

Methods: The GWAS involved 1,483 European NAFLD cases and 17,781 genetically matched controls. A replication cohort of 559 NAFLD cases and 945 controls was genotyped to confirm signals showing genome-wide or close to genome-wide significance.

Results: Case-control analysis identified signals showing p values $\leq 5 \times 10^{-8}$ at 4 locations (chromosome [chr] 2 *GCKR/C2ORF16*; chr4 *HSD17B13*; chr19 *TM6SF2*; chr22 *PNPLA3*) together with 2 other signals with $p < 1 \times 10^{-7}$ (chr1 near *LEPR* and chr8 near *IDO2/TC1*). Case-only analysis of quantitative traits showed that the *PNPLA3* signal (rs738409) had genome-wide significance for steatosis, fibrosis and NAFLD activity score and a new signal (*PYGO1* rs62021874) had close to genome-wide significance for steatosis ($p = 8.2 \times 10^{-8}$). Subgroup case-control analysis for NASH confirmed the *PNPLA3* signal. The chr1 *LEPR* single nucleotide polymorphism also showed genome-wide significance for this phenotype. Considering the subgroup with advanced fibrosis (≥F3), the signals on chr2, chr19 and chr22 maintained their genome-wide significance. Except for *GCKR/C2ORF16*, the genome-wide significance signals were replicated.

Conclusions: This study confirms *PNPLA3* as a risk factor for the full histological spectrum of NAFLD at genome-wide significance levels, with important contributions from *TM6SF2* and *HSD17B13*. *PYGO1* is a novel steatosis modifier, suggesting that Wnt signalling pathways may be relevant in NAFLD pathogenesis.

Lay summary: Non-alcoholic fatty liver disease is a common disease where excessive fat accumulates in the liver and may

Keywords: NAFLD; NASH; Fibrosis; GWAS; PNPLA3; TM6SF2; GCKR; HSD17B13; SNP.
Received 3 October 2019; received in revised form 20 March 2020; accepted 2 April 2020; available online 13 April 2020

* Corresponding authors. Addresses: Translational & Clinical Research Institute, The Medical School, Newcastle University, 4th Floor, William Leech Building, Framlington Place, Newcastle upon Tyne, NE2 4HH, United Kingdom. Tel.: + 44 (0) 191 208 7031 (A.K. Daly), or Tel.: + 44 (0) 191 208 7012 (Q.M. Anstee).
E-mail addresses: a.k.daly@ncl.ac.uk (A.K. Daly), quentin.anstee@ncl.ac.uk (Q.M. Anstee).

☆ Guest Editor: Michael Manns.

EPoS Consortium Investigators listed at the end of the manuscript.

https://doi.org/10.1016/j.jhep.2020.04.003

result in cirrhosis. To understand who is at risk of developing this disease and suffering liver damage, we undertook a genetic study to compare the genetic profiles of people suffering from fatty liver disease with genetic profiles seen in the general population. We found that particular sequences in 4 different areas of the human genome were seen at different frequencies in the fatty liver disease cases. These sequences may help predict an individual's risk of developing advanced disease. Some genes where these sequences are located may also be good targets for future drug treatments.
© 2020 European Association for the Study of the Liver. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Non-alcoholic fatty liver disease (NAFLD) represents a spectrum of progressive liver disease characterised by increased hepatic triglyceride content (HTGC) in the absence of excess alcohol consumption.[1] NAFLD encompasses steatosis (non-alcoholic fatty liver [NAFL]), steatohepatitis (non-alcoholic steatohepatitis [NASH]), fibrosis and ultimately cirrhosis. It is strongly associated with features of the metabolic syndrome (obesity, type 2 diabetes mellitus [T2DM] and dyslipidaemia).[1] Although common, affecting approximately 25% of the global adult population, only a minority of patients with NAFL develop NASH, progress to significant fibrosis or experience associated morbidity.[1,2] NAFLD is best considered a complex trait where disease phenotype results from environmental exposures acting on a susceptible polygenic background comprising multiple independent modifiers.[3]

Genome-wide association studies (GWASs) have contributed greatly to our understanding of the genetic contribution to NAFLD pathogenesis and variability of prognosis.[3] Amongst the loci identified, the non-synonymous single nucleotide polymorphism (SNP) in *PNPLA3* (phospholipase domain-containing 3) (rs738409),[4,5] and more recently, a non-synonymous SNP in *TM6SF2* (transmembrane 6 superfamily member 2) (rs58542926), originally ascribed to the neighbouring *NCAN* gene,[5] have been associated with ^{1}H-MRS quantified HTGC.[6] Both genetic associations have been replicated in further studies where they have been associated not only with steatosis, but also with clinically relevant factors including grade of steatohepatitis and stage of hepatic fibrosis/cirrhosis[7,8] and, in the case of *PNPLA3*, with the development of NAFLD-related hepatocellular carcinoma.[9,10] A number of other associations, with *LYPLAL1, GCKR*, and *PPP1R3B,* have been reported by GWAS comprising relatively few histologically characterised cases and are currently less robustly replicated.[3,5] A recent study using exome sequencing[11] confirmed a previously reported association of raised alanine aminotransferase (ALT) with a *HSD17B13* SNP (rs6834314)[12] in a general patient population and then demonstrated that this polymorphism was associated with NAFLD. Two further studies broadly confirmed this association.[13,14]

To date, most adequately powered GWASs relevant to NAFLD have addressed either radiologically determined HTGC[4,6,12] or clinical biochemistry parameters such as ALT.[12,15] They have therefore been unable to address the more clinically relevant phenotypes of steatohepatitis grade or fibrosis stage (reviewed[3]). One GWAS has assessed a large number of histologically characterised patients, reporting associations with both *PNPLA3* and with chromosome 19 close to *TM6SF2*.[16] These patients, however, were recruited from bariatric surgery programmes with dietary restrictions prior to surgery and wedge biopsy collection which may affect liver histology; in addition such patients tend to be younger and have a higher average BMI than NAFLD cases more generally.[17] The current study aims to identify genetic modifiers of steatohepatitis and fibrosis, attaining genome-wide levels of statistical significance by using a large internationally derived cohort of patients (with histologically characterised NAFLD and representing all stages of the disease). We now report the largest histology-based NAFLD GWAS to date in a cohort of 1,483 European patients exhibiting the full spectrum of biopsy-proven NAFLD.

Materials and methods

NAFLD cases

For the main GWAS study, patients were recruited from clinics at several leading European tertiary liver centres (see supplementary methods). Additional cases for replication were recruited at Foundation IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. The study had the necessary ethical approvals from the relevant national/institutional review boards (see supplementary methods) and all participants provided informed consent. All cases were unrelated patients that had undergone a liver biopsy as part of the routine diagnostic workup for presumed NAFLD having originally been identified due to abnormal biochemical tests (ALT and/or gamma-glutamyltransferase) and/or an ultrasonographically detected bright liver, associated with features of the metabolic syndrome; or having abnormal biochemical tests (ALT and/or gamma-glutamyltransferase) and macroscopic appearances of a steatotic liver at the time of bariatric surgery. Full details of inclusion/exclusion criteria are provided in the supplementary methods.

Controls

We used general population samples with existing genome-wide genotype data as study controls. For the GWAS, we selected European ancestry controls (n = 17,781) from multiple sources as described in the supplementary methods. To replicate GWAS associations, we used an Italian control cohort (n = 945) consisting of controls described previously[18] with some newly collected individuals. Any that were found to match the Hypergenes controls already used in our discovery GWAS were excluded.

Histology

Liver biopsy specimens (at least 1.6 cm length and ~1 mm diameter) were formalin-fixed and paraffin-embedded. Tissue sections (5 μm-thick) were routinely stained with haematoxylin and eosin and trichrome stain to visualise collagen. All cases were recruited at tertiary centres where liver biopsies were routinely assessed according to accepted criteria by experienced liver pathologists and scored using the well validated NIDDK NASH-CRN system.[19] To ensure optimum data quality, biopsies were retrieved from archival storage where possible (78% of cases) and scored centrally by an expert liver pathologist from the FLIP/EPoS central pathology team (DT, ADB, PB), as described in detail previously.[20] Where archival samples were unavailable for central reading, the local liver pathologist's scores were used. To maximise insights into the specific pathophysiological processes that occur as NAFLD progresses, 6 phenotypes of interest were studied: degree of steatosis (S0-3); degree of ballooning

(B0-2); degree of lobular inflammation (I0-3); severity of NASH activity (calculated as 'disease activity' = hepatocyte ballooning (B0-2) + lobular inflammation (I0-3) and also an overall NAFLD activity score 'NAS' combining all 3 parameters (NAS0-8)); and stage of fibrosis (F0-4).

Genotyping

DNA was prepared from blood samples collected with EDTA as described previously.[21] GWAS genotyping was carried out in 2 phases. For phase I, genotyping was performed initially using the Illumina OmniExpress BeadChip by Edinburgh Clinical Research Centre. To obtain data for additional exomic SNPs, further genotyping of these samples was performed using the Illumina HumanCoreExome BeadChip (Aros, Denmark). Genome-wide genotyping of the phase II cases was performed using the Illumina OmniExpressExome BeadChip by the Edinburgh Clinical Research Centre. A total of 721,078 markers shared across the batches passed quality control (see supplementary methods). SNP imputation was performed as described in detail in the supplementary methods.

The top associated SNPs were further confirmed in replication cases using TaqMan® SNP genotyping assays (ThermoFisher Scientific, Waltham, MA) in accordance with the manufacturer's recommendations. If an assay could not be designed for the SNP showing the strongest signal for the region, a suitable proxy SNP was chosen (https://ldlink.nci.nih.gov/?tab = home).

RNA sequencing and *in vitro* studies

RNA sequencing

RNA sequencing data on samples from 206 liver biopsies from patients with NAFLD, as described elsewhere (Govaere *et al.*, submitted), was used to further investigate the functional significance of *HSD17B13* variants.

Bioluminescent retinol dehydrogenase assays for HSD17B13

Retinol (75 μM; Sigma-Aldrich, St. Louis, Missouri, USA) was incubated with recombinant HSD17B13 (TP313132; Origene, Maryland, USA) for 1 h at room temperature in the presence of 0.5 mM NAD in 200 mM Tris-HCl, pH7.5. As a control, the known HSD17B13 substrate β-estradiol (75 μM) was incubated in parallel assays. NADH production was measured by Bioluminescent NAD/NADH-Glo™ Assay (Promega, Wisconsin, USA) according to manufacturer's guidelines.

Statistical analysis

We used principal component analysis (PCA) of the genome-wide genotype data to investigate the ancestry of the cases and controls; this showed the expected north/south variation commonly seen across Europe[22] but, importantly, suggested adequate matching between cases and controls (Fig S1A and Fig S1B). Case/control analysis and quantitative trait analysis of GWAS data was performed as described in detail in the supplementary methods, using a linear mixed modelling approach with the incorporation of the top 5 principal components as covariates to adjust for any population stratification. Examination of the resulting genome-wide QQ plots and genomic control inflation factors (λ)[23] (see Results) indicated that this adjustment adequately corrected for any population differences.

Significance of findings in the replication cohort was assessed by calculation of odds ratios, 95% confidence intervals and *p* values by univariate analysis and multiple logistic regression using PLINK.[24]

Results

Clinical characteristics of the cases

Clinical details of the NAFLD cases included in the main GWAS are summarised in Table 1. The replication cohort details are shown in Table S1. All cases in both cohorts were of white European ethnicity. The percentage with advanced fibrosis (stage F3 or F4) was similar in both cohorts (p >0.05) but other parameters including age, BMI, T2DM, sex and incidence of NASH were different.

Overall NAFLD case-control analysis

The overall NAFLD case-control analysis is presented as a Manhattan plot (Fig 1). PCA scattergrams for cases and controls are shown in Fig S1 and the QQ plot of the association results in Fig S2. As summarised in Table 2, 4 different regions (on

Table 1. Characteristics of the cohort (n = 1,483).

Patient demographic and clinical characteristics	
Age (years) (mean ± SD)	50.1 ± 13.0
Sex (% female)	47.30%
BMI median, kg/m^2 (IQR)	35.19 (29.1–39.7)
T2DM, n (%)	593(40.0)*
Histologic characteristics	
Steatosis, n (%)	
0	53 (3.6)
1	483 (32.6)
2	541 (36.5)
3	390 (26.3)
Missing	16 (1.1)
NAS score, n (%),	
0	19 (1.3)
1	138 (9.3)
2	225 (15.2)
3	258 (17.4)
4	271 (18.3)
5	283 (19.1)
6	178 (12.0)
7	80 (5.4)
8	15 (1.0)
Missing	16 (1.1)
Disease activity score, n (%)	
0	255 (17.2)
1	285 (19.2)
2	418 (28.2)
3	308 (20.8)
4	166 (11.2)
5	35 (2.4)
Missing	16 (1.1)
NASH, n (%)	
Yes	836 (56.4)
No	631 (42.5)
Missing	16 (1.1)
Fibrosis, n (%)	
0	432 (29.1)
1	350 (23.6)
2	312 (21.0)
3	240 (16.2)
4	147 (9.9)
Missing	2 (0.13)

NAS, non-alcoholic fatty liver disease activity score; NASH, non-alcoholic steatohepatitis; T2DM, type 2 diabetes mellitus.
*For T2DM, 5 (0.33%) missing.

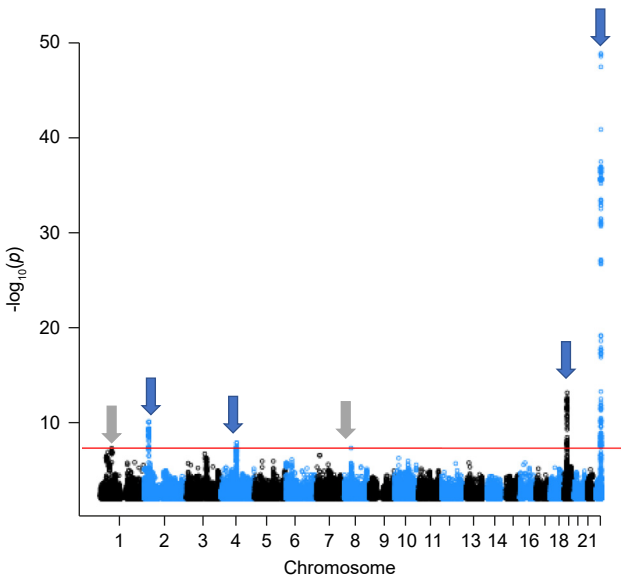


Fig. 1. Manhattan plot from imputed GWAS case-control analysis. Included 1,483 NAFLD cases and 17,781 controls. Threshold for genome-wide significance was taken to be 5×10^{-8}. The first 5 principal components were included as covariates. Genome-wide significant signals are indicated by blue arrows with those showing p in the range 1×10^{-7} to 5×10^{-8} shown by grey arrows. GWAS, genome-wide association study; NAFLD, non-alcoholic fatty liver disease. (This figure appears in color on the web.)

chromosomes 2, 4, 19 and 22) passed conventional genome-wide significance ($p < 5 \times 10^{-8}$) with 2 other regions (on chromosomes 1 and 8) showing p values $< 1 \times 10^{-7}$ (for LocusZoom plots see Fig S3). Data presented in Fig. 1 were obtained from imputation analysis. Primary case-control analysis without imputation showed similar signals in chromosomes 2, 4, 19 and 22 only but no additional signals at $p < 1 \times 10^{-7}$ (Fig. S4 and Table S2). Correction of the imputed data for sex in addition to the first 5 principal components used in the main analyses did not result in large changes in p value (Table S3). Together, these results point to *PNPLA3, TM6SF2, HSD17B13* and the *GCKR/C2ORF16* region being the major risk factors for disease susceptibility with borderline signals for chromosome 1 near *LEPR* and for chromosome 8 adjacent to *IDO2* and *TC1(C8orf4)*. In view of the well-established strong association of *PNPLA3* rs738409 with NAFLD, additional analysis using a model conditioning on this SNP was performed. This analysis gave broadly similar findings to those summarised in Table 2 with no new signals (data not shown).

Quantitative trait analysis of NAFLD phenotypes

Case-only analyses assessing relevance of genotype to grade of steatosis (assessed as predefined 'disease activity' and 'NAS') and stage of fibrosis were also performed using the imputed data. Results of these analyses are shown in Fig. 2 with the most significant signals summarised in Table 3 (for QQ and LocusZoom plots see Figs. S5 and S6). The primary data without imputation are summarised in Fig. S7 and Table S4. For steatosis, NAS and fibrosis as quantitative traits, signals with $p < 10^{-10}$ were detected for *PNPLA3* rs738409 and other SNPs in this region of chromosome 22. For steatosis, a signal with $p = 8.2 \times 10^{-8}$ on chromosome 15 (rs62021874 in *PYGO1*) was also detected (Table 3). This variant is in complete linkage disequilibrium with a missense variant rs11858624 which also showed a signal close to significance ($p = 1.7 \times 10^{-7}$). No signals reached conventional genome-wide significance ($p < 5 \times 10^{-8}$) for disease activity score alone or when ballooning or inflammation were considered as individual traits (Fig. S8). The effect of correction of the imputed data for clinical covariates was also assessed for each trait (Table S5), giving results very similar to those obtained originally.

To further assess the relevance of genotype to particular NAFLD phenotypes, the contribution to NAFLD progression of the 4 major genetic risk factors identified in the case-control GWAS was assessed by calculating a combined genetic risk score based on summing the allele count (with no weighting by effect size) for *PNPLA3* rs738409, *TM6SF2* rs58542926, *GCKR* rs1260326 and *HSD17B13* rs9992651 and relating the resulting score to grade of steatosis, NAS and fibrosis stage (Fig. S9). Trend tests by linear regression showed that there was a statistically significant relationship between the value of the semi-quantitative steatosis/NAS/fibrosis scores and the value of the genetic risk score for all 3 phenotypes, with the most significant relationship ($p = 4.68 \times 10^{-13}$) detected for fibrosis stage (Fig S9). Those with a risk score of 2 (n = 216) had a mean fibrosis score of 1.27 (SE 0.08) compared with 1.94 (SE 0.09) for a risk score of 5 (n = 260).

Additional subgroup case-control analysis

Since both steatohepatitis and advanced fibrosis are clinically important phenotypes in NAFLD,[25] additional case-control analyses were undertaken including cases with NASH only (n = 836)

Table 2. Summary of top findings in the NAFLD case-control analysis.

SNP	Chromosome	A1	Gene	*p* value	OR (95% CI)
rs12077210*	1	T	*LEPR*	$5.62E^{-08}$	1.484 (1.287–1.711)
rs1260326*	2	T	*GCKR*	$1.06E^{-10}$	1.278 (1.186–1.377)
rs1919127*	2	C	*C2orf16*	$5.61E^{-10}$	1.290 (1.190–1.398)
rs2068834	2	C	*ZNF512*	$8.49E^{-11}$	1.302 (1.202–1.410)
rs9992651	4	A	*HSD17B13*	$2.78E^{-08}$	0.744 (0.671–0.826)
rs13118664	4	T	*HSD17B13*	$1.41E^{-08}$	0.740 (0.667–0.821)
rs139648192	8	T	-	$5.20E^{-08}$	1.538 (1.317–1.796)
rs58542926*	19	T	*TM6SF2*	$2.05E^{-11}$	1.609 (1.400–1.849)
rs8107974	19	T	*SUGP1*	$2.58E^{-12}$	1.632 (1.423–1.872)
rs17216588	19	T	-	$7.25E^{-14}$	1.612 (1.423–1.827)
rs10500212	19	T	*PBX4*	$3.40E^{-12}$	1.549 (1.369–1.752)
rs738409*	22	G	*PNPLA3*	$1.45E^{-49}$	1.827 (1.687–1.979)

7,412,561 imputed SNPs included; total number of cases and controls = 19,264. ORs were obtained from logistic regression in PLINK and confidence intervals were calculated from back-transformation of FaST-LMM p-values and PLINK ORs.
OR, odds ratio; SNP, single nucleotide polymorphism.
*Denotes validated SNP following imputation. The first 5 principal components were included as covariates.

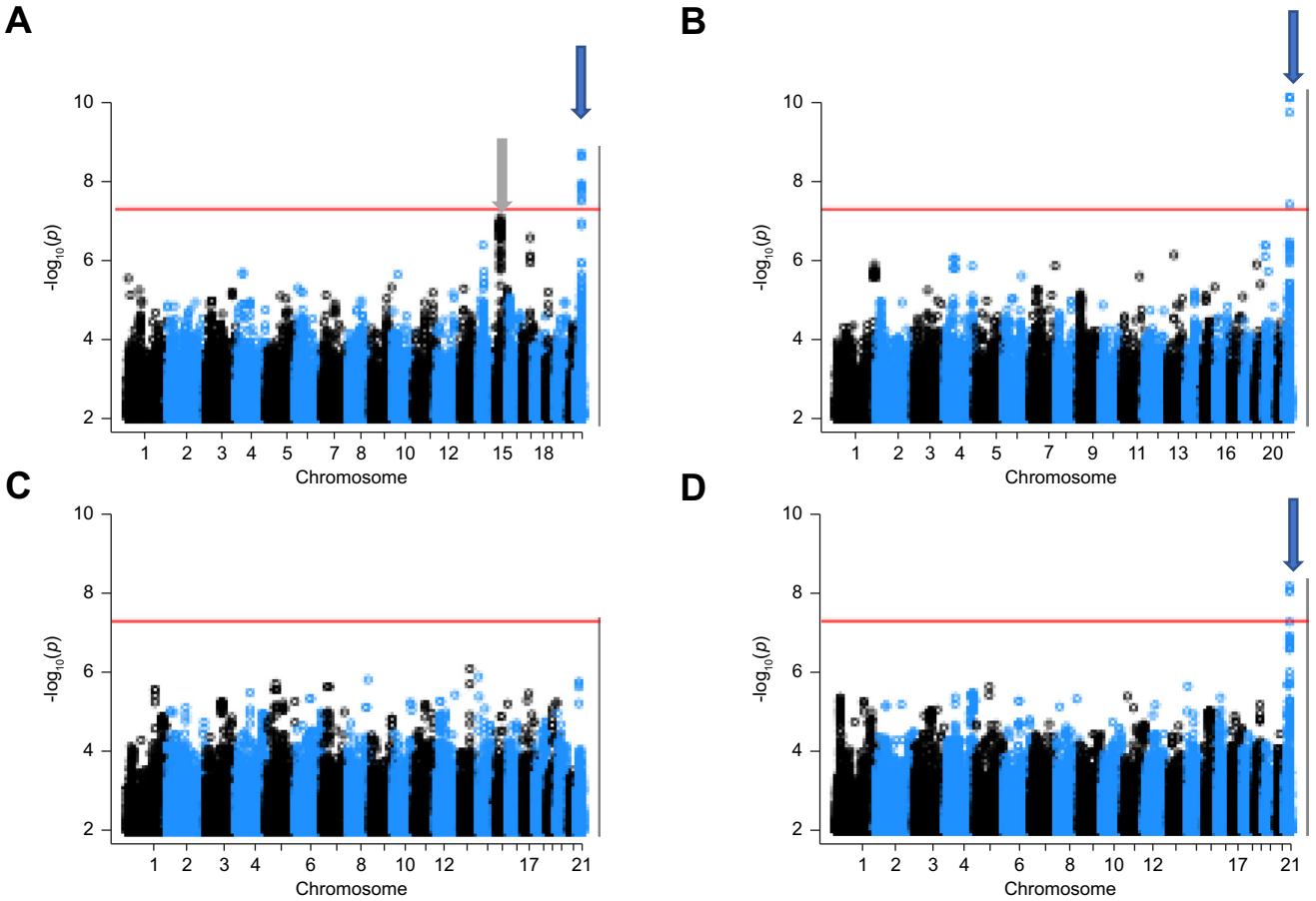


Fig. 2. Manhattan plots from imputed GWAS analysis on the basis of quantitative traits. Included 1,483 NAFLD cases. Threshold for genome-wide significance was taken to be 5×10^{-8} but signals showing $p < 1 \times 10^{-7}$ are also indicated. Panel A shows data for steatosis, B for fibrosis, C for disease activity score and D for NAS score. The first 5 principal components were included as covariates. Genome-wide significant signals are indicated by blue arrows with those showing p in the range 1×10^{-7} to 5×10^{-8} shown by grey arrows. GWAS, genome-wide association study; NAFLD, non-alcoholic fatty liver disease; NAS, NAFLD activity score. (This figure appears in color on the web.)

and fibrosis stage F3 and F4 only (n = 386). The findings for both phenotypes are summarised in Fig. 3 and Table 4 (for QQ and LocusZoom plots see Fig. S10 and S11). For NASH, signals showing p values of $<5 \times 10^{-8}$ were detected for chromosome 1 (*LEPR*) and chromosome 22 (*PNPLA3*) (Table 4). For *LEPR* rs12077210, the p value of 4.4×10^{-9} was lower for NASH than for NAFLD overall (Table 2). A second novel chromosome 1 signal (rs80084600) with $p = 7.1 \times 10^{-8}$ located in an intergenic region downstream of phospholipase A2 group IVA (*PLA2G4A*) was also detected. The SNPs in chromosomes 2, 4 and 19 that were significant in the main case-control analysis showed p values in the region of 2×10^{-7} so came close to significance for NASH. For fibrosis stages F3 and F4, chromosome 2, 19 and 22 signals showing p values of $<5 \times 10^{-8}$ were detected but the signals from the main case-control analysis detected previously for chromosomes 1, 8 and 4 showed p values $>1 \times 10^{-7}$. For *HSD17B13* rs9992651 (chromosome 4), the p value was 1.16×10^{-5}.

Replication of GWAS signals and investigation of additional possible NAFLD risk factors

A replication cohort of 559 Italian NAFLD cases was assembled from a different centre to the discovery cohort. Allele frequencies for selected SNPs in these cases were compared with those for Italian controls. Findings for 8 separate loci giving signals with $p < 1 \times 10^{-7}$ in either the main GWAS or the quantitative trait studies are summarised in Table 5. The *PNPLA3*, *TM6SF2* and *HSD17B13* signals seen in the main GWAS replicated ($p < 0.05$) but we found only borderline effects or no significance for 4 other loci. However, the *PYGO1* signal, which was associated with steatosis by quantitative trait analysis, showed a significant association in the analysis in the same protective direction as observed for steatosis. The *GCKR/C2Orf16* signal did not replicate either in the main replication cohort (Table 5) or in a subgroup of replication cases (n = 134) with fibrosis stage 3 or 4. Due to the relatively low number of NASH cases in the replication cohort, we did not seek to replicate the novel rs80084600 signal seen for this phenotype. Multiple logistic regression analysis with adjustment for *PNPLA3* rs738409 and *TM6SF2* rs58542926 (Table 5) generated similar findings to the univariate analysis, apart from small decreases in p values for the *HSD17B13* and *PYGO1* signals.

Results for selected variants reported recently by others as risk factors for NAFLD but which had not shown p values of $<1 \times 10^{-7}$ in the current GWAS were also extracted from the main case-control analysis. Only rs2642438 in *MARC1* (mitochondrial amidoxime-reducing component 1) and rs28929474 in *AAT* (alpha1-antitrypsin) showed p values <0.05 (Table S6). For rs2642438, the p value was 6×10^{-6} with a protective odds ratio of 0.816, in line with that reported previously.[26]

Table 3. Summary of top findings in quantitative trait analysis.

SNP	Chromosome	A1	Gene	Phenotype	n	*p* value (no clinical covariates)	Beta (95% CI)
rs738409*	22	G	*PNPLA3*	Steatosis	1,469	$2.37E^{-09}$	0.183 (0.123–0.243)
rs62021874	15	T	*PYGO1*	Steatosis	1,469	$8.16E^{-08}$	−0.303 (−0.414 to −0.192)
rs11858624*	15	T	*PYGO1*	Steatosis	1,469	$1.64E^{-07}$	−0.295 (−0.406 to −0.185)
rs738409*	22	G	*PNPLA3*	Fibrosis	1,481	$7.58E^{-11}$	0.318 (0.222–0.414)
rs738409*	22	G	*PNPLA3*	NAS	1,467	$8.78E^{-09}$	0.364 (0.240–0.488)

Results for 7,900,223 imputed SNPs. First 5 principal components were included as covariates. ORs were obtained from logistic regression in PLINK and confidence intervals were calculated from back-transformation of FaST-LMM p-values and PLINK ORs.
SNP, single nucleotide polymorphism.
*Validated directly by genotyping.

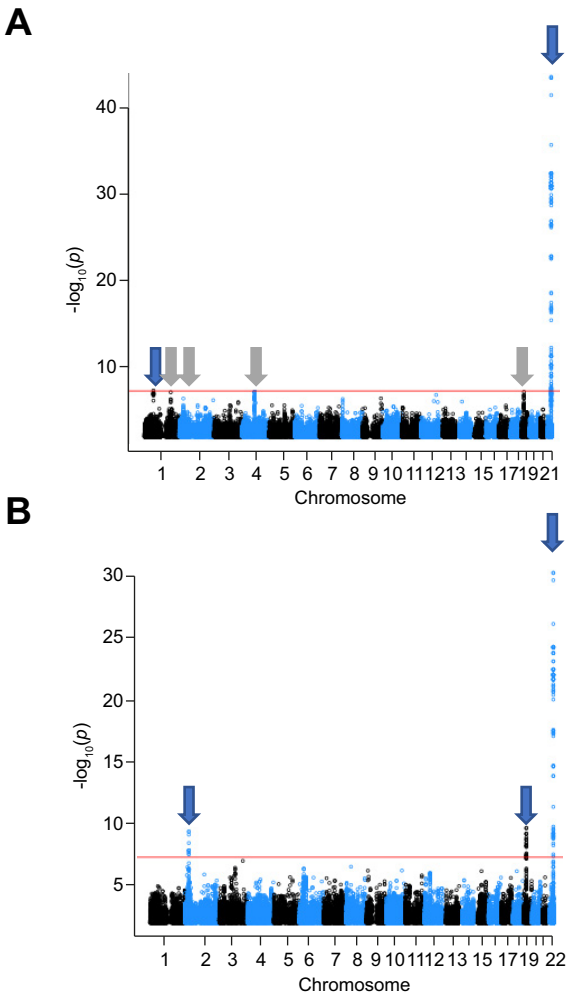


Fig. 3. Manhattan plots from imputed GWAS case-control analysis of NASH and severe fibrosis (F3/F4). Threshold for genome-wide significance was taken to be 5×10^{-8}. The first 5 principal components were included as covariates. Panel A. NASH analysis. 836 cases and 17,781 controls. Panel B. F3/F4 analysis. 386 cases and 17,781 controls. Genome-wide significant signals are indicated by blue arrows with those showing *p* in the range 1×10^{-7} to 5×10^{-8} shown by grey arrows. GWAS, genome-wide association study; NASH, non-alcoholic steatohepatitis. (This figure appears in color on the web.)

EQTL analysis and studies on expression of GWAS signals in liver biopsies from different NAFLD stages

While the signals seen for NAFLD relating to *PNPLA3*, *TM6SF2* and *GCKR* are already well-established risk factors for this disease from population studies[4–6,27] and studies on functional significance,[28–30] evidence for functional significance for the other signals is limited. The relationship of rs9992651 and rs72613567 in *HSD17B13* with gene expression was evaluated by sequencing RNA samples from liver biopsies. Three different *HSD17B13* transcripts were detected (Fig. S12), including a full-length transcript with all 7 exons, a variant with exon 2 deleted and a variant without exon 6. Based on genotype for rs9992651 from the RNA sequencing data, the variant without exon 6 was generally not detectable in homozygotes for the reference G allele but was expressed at a higher level in homozygotes for the minor A allele and also heterozygotes. The ability of recombinant HSD17B13 to oxidise retinol[13] was also confirmed (Fig. S13).

Other loci showing associations in the case-control studies including rs12077210 in *LEPR* (intronic), rs139648192 on chromosome 8 and rs80084600 on chromosome 1 could not be investigated by RNA sequencing due to their locations. The borderline significant rs11858624 in *PYGO1* (Table 3) is a missense variant (P299H). Analysis with data obtained from GTEx (https://gtexportal.org/home/) indicated no difference in RNA expression between rs11858624 homozygous wild-types and heterozygotes in liver tissue (Fig. S14).

Discussion

This study is the largest GWAS to date on histologically characterised NAFLD enrolled in a hepatology setting that addresses the full disease spectrum from steatosis to cirrhosis. This contrasts with the only previous GWAS involving more than 1,000 histologically characterised cases, which was in a predominantly female bariatric cohort with extreme obesity but relatively mild NAFLD.[16] Furthermore, that study only considered grade of steatosis, not the more clinically relevant phenotypes of steatohepatitis or fibrosis.[16] The current study confirms the well-established signals in *PNPLA3*, *TM6SF2* and *GCKR*, together with the more recently reported *HSD17B13* signal.[11] The findings for *GCKR* are in line with several candidate gene studies on NAFLD however, this is the first GWAS study reporting this 4 gene combination as NAFLD risk modifiers.

HSD17B13 has been reported to be relevant to NAFLD with several variants associated with decreased risk.[11,13] The current study found a protective effect against NAFLD generally, with the strongest effect related to the SNPs rs9992651 and rs13118664. These SNPs are in non-coding regions of *HSD17B13* but are in strong linkage disequilibrium with rs72613567, which is associated with a single base-pair insertion that has been suggested to be of functional significance in relation to RNA splicing.[11] The current study confirms that an *HSD17B13* isoform lacking exon 6 is associated with rs9992651 and a protective effect against NAFLD; consistent with a report

Table 4. Summary of top findings from case-control analysis for NAFLD cases with NASH or with fibrosis scores F3 and F4 only.

SNP	Chromosome	Gene	*p* value (no clinical covariates)	OR (95% CI)
NASH				
rs12077210	1	*LEPR*	$4.42E^{-09}$	1.671 (1.390–2.008)
rs80084600	1	–	$7.08E^{-08}$	1.977 (1.543–2.533)
rs1260326	2	*GCKR*	$3.78E^{-07}$	1.302 (1.176–1.442)
rs9992651	4	*HSD17B13*	$2.92E^{-07}$	0.718 (0.633–0.815)
rs13118664	4	*HSD17B13*	$2.37E^{-07}$	0.716 (0.631–0.813)
rs58542926	19	*TM6SF2*	$1.90E^{-07}$	1.606 (1.344–1.919)
rs8107974	19	*SUGP1*	$1.36E^{-07}$	1.609 (1.348–1.920)
rs738409	22	*PNPLA3*	$2.58E^{-44}$	2.053 (1.856–2.271)
Fibrosis F3/F4				
rs1260326	2	*GCKR*	$4.07E^{-10}$	1.678 (1.427–1.974)
rs56255430	19	–	$2.11E^{-10}$	1.863 (1.538–2.257)
rs738409	22	*PNPLA3*	$5.66E^{-31}$	2.374 (2.051–2.748)

N = 18,167 (Cases = 386, Controls = 17,781), covariate model includes first 5 principal components. ORs were obtained from logistic regression in PLINK and confidence intervals were calculated from back-transformation of FaST-LMM p-values and PLINK ORs.
NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; OR, odds ratio; SNP, single nucleotide polymorphism.

Table 5. Genotype frequencies in replication cohort.

				Univariate analysis		Multiple logistic regression adjusting for *PNPLA3* rs738409 and *TM6SF2* rs58542926	
Gene	**SNP**	**Case frequency**	**Control frequency**	**Odds ratio**	***p* value**	**Odds ratio**	***p* value**
LEPR	rs12077210	0.05877	0.05983	0.98 (0.71–1.35)	0.91	0.96 (0.69–1.34)	0.81
GCKR	rs1260326	0.5407	0.5305	1.04 (0.90–1.21)	0.59	1.08 (0.92–1.27)	0.36
C2ORF16	rs1919127	0.382	0.3566	1.12 (0.96–1.30)	0.16	1.1 (0.94–1.29)	0.25
HSD17B13	rs72613567	0.2101	0.2462	0.81 (0.68–0.97)	0.025	0.78 (0.64–0.95)	0.013
IDO2 /TC1(C8orf4)	rs79137099	0.03789	0.03891	0.97 (0.66–1.44)	0.89	1.05 (0.6–1.59)	0.83
PYGO1	rs11852624	0.05144	0.0709	0.71 (0.52–0.98)	0.035	0.67 (0.48–0.96)	0.027
TM6SF2	rs58542926	0.08813	0.05027	1.83 (1.36–2.45)	$4.63E^{-05}$	n.a.	n.a.
PNPLA3	rs738409	0.4436	0.2754	2.10 (1.80–2.45)	$6.60E^{-21}$	n.a.	n.a.

Significance of findings was assessed by calculation of odds ratios, 95% confidence intervals and *p* values by univariate analysis (chi-square test) and multiple logistic regression using PLINK.
SNP, single nucleotide polymorphism.

showing a similar splicing pattern with the SNPs rs6834314 and rs72613567[13] but differing from that described in the original report.[11] Consistent with that recent study,[13] we also show the *HSD17B13* gene product possesses retinol dehydrogenase activity. Retinol metabolism is a complex multistep process involving a number of different enzymes.[31] While it remains unclear whether loss of HSD17B13 retinol dehydrogenase activity can explain the protective effect of the variant, it is likely that enzyme activity in the reverse direction involving retinal reduction to retinol could also be impaired since these enzymes operate in both oxidising and reducing directions.[31] Thus, increased levels of retinal and the biologically active retinoic acid isomers could occur in those carrying *HSD17B13* variants. This effect might protect against NAFLD development, in line with recent evidence that 13-*cis* and all-*trans* retinoic acid are found at significantly decreased levels in human livers with NAFLD.[32] A clear trend towards a protective effect against advanced hepatic fibrosis was observed, although this did not reach genome-wide significance levels (*p* value approx. 10^{-5}). Given that the strength of association with NASH was stronger (*p* values approx. 2×10^{-7}), it may be that the protective effect of *HSD17B13* is more relevant to development of steatohepatitis than progression of fibrosis.

The *GCKR* signal in both the main GWAS and advanced fibrosis-only analysis identified rs1260326 as the most significant SNP within this region, with T-variant carriage increasing NAFLD risk. This common missense variant has been studied widely both as a risk factor for T2DM and for NAFLD. An upstream SNP, rs780094, in strong linkage disequilibrium with rs1260326, has also been shown to be a NAFLD risk factor in candidate gene studies.[33] The relationship between both SNPs and susceptibility to NAFLD and T2DM is complex. Rs1260326 is well established to have a protective effect against T2DM, probably due to the *GCKR* variant showing weaker interaction with glucokinase compared with the wild-type.[34] This promotes hepatic glucose metabolism, decreasing plasma glucose levels, and is associated with an increased risk of NAFLD.[33] The underlying mechanism is unclear but rs1260326 is associated with higher levels of circulating lactate,[35] presumably due to increased glucose metabolism via glycolysis. The inability to replicate the *GCKR* association was slightly surprising but may reflect the overall lower severity of NAFLD in the replication cohort. There are a relatively large number of reports of a significant increased risk for *GCKR* variants in NAFLD generally, especially for paediatric cases.[27,36,37]

A further interesting finding relates to a signal on chromosome 15 (rs11858624) that was close to genome-wide significance for steatosis and was validated in the replication study. The gene involved is *PYGO1*, which encodes a transcription factor that contributes to the Wnt signalling pathway.[38] The exact impact of *PYGO1* in Wnt signalling remains unclear, though a homologue *PYGO2* appears to contribute to several physiological pathways

including increased adiposity and impaired glucose tolerance in mice lacking this protein.[39]

Signals on chromosomes 1 and 8 were detected in the case-control analysis, however these just failed to meet genome-wide significance and did not replicate. The chromosome 1 SNP was genome-wide significant in the NASH-only case-control analysis and lies in the region encoding *LEPROT* and *LEPR*; both genes share the same promoter and first 2 exons but encode separate proteins. This association is notable given that *db/db* mice, carrying a spontaneous loss of function mutation in the OB-Rb leptin receptor, have been widely used to model NAFLD.[40] There are also some previous reports from candidate gene studies that *LEPR* variants are risk factors for NAFLD but the current variant lies considerably upstream of these previously studied variants.[41,42] The signal on chromosome 8 relates to an area between *IDO2* and *TC1*. Of potential relevance to NAFLD, both genes have roles in modulating inflammation with IDO2 inducible by lipopolysaccharide and contributing to immune function[43] while TC1 modulates NF-κB signalling. Further investigation of these variants is needed. The subgroup analysis on NASH grade showed a second novel chromosome 1 signal separate from *LEPR*. The *p* value for NASH, though not genome-wide significant at 7×10^{-8}, was considerably lower than that seen for this variant in the main case-control study (0.0049). The variant is in an intergenic region but is downstream of *PLA2G4A*, which shows elevated expression in adipose tissue in obesity and may contribute to T2DM susceptibility.[44]

The most significant associations in this study were obtained for NAFLD in the binary case-control design. The quantitative trait analyses has shown a clear association for *PNPLA3* rs738409 with steatosis, NAS score and fibrosis, which is generally in line with previous reports in NAFLD and alcohol-related liver disease.[45] However, there were no significant associations of any genotype with disease activity when considered separately from steatosis. The failure to see more specific associations for *TM6SF2* and *HSD17B13* with other histological traits similar to those reported previously in candidate gene studies may reflect the complex nature of the histological disease phenotype[8,11] and also limited statistical power. In contrast to quantification of HTGC by imaging techniques, which provides a highly reproducible quantitative measure of a single biochemical entity, the histological scoring systems used to evaluate steatohepatitis and fibrosis provide only non-linear, semi-quantitative or categorical assessments of disease and are subject to intra- and inter-observer variation. Indeed, clear diagnostic consensus regarding the presence or absence of steatohepatitis among pathologists is not always feasible.[19,20,46] Thus, the conduct of a histology-based GWAS, whilst addressing the most clinically relevant phenotypic characteristics, is technically more challenging. We have addressed this challenge by using expert liver pathologists to provide histological diagnosis and scoring. The reduced statistical power due to the limited number of cases in particular histological categories, may limit the number of variants that attain the genome-wide significance threshold to only the most strongly associated, such as the *PNPLA3* variant. Despite these limitations, disease severity was correlated with genetic risk score based on the most significant case-control GWAS signals, statistically significant relationships for association of the risk score with increasing degree of steatosis, grade of steatohepatitis and fibrosis stage were found, which suggests that a risk score approach may be of value prognostically although further studies on this are needed.

Despite a fairly extensive supporting literature, we and others[47] have not found *MBOAT7* to be a risk factor for NAFLD. Notably, no NAFLD focussed GWAS to date has reported a significant association with *MBOAT7*. Other signals for NAFLD reported by others previously including in *PPP1R3B*,[5] *AAT*[48] and interferon lambda 4[49] also failed to show genome-wide significance in the case-control analysis. This is not surprising in the case of *AAT* as patients known to have this condition were specifically excluded from the cohort, limiting the minor allele frequency substantially. However, the gene *MARC1*, where a non-synonymous variant has been reported to protect against both "all cause" cirrhosis and fatty liver disease,[26] showed a similar protective effect against NAFLD with a low *p* value, though this did not attain genome-wide significance. This gene encodes the mitochondrial amidoxime-reducing component enzyme which can reduce trimethylamine N-oxide (TMAO) generated by oxidation of trimethylamine. Elevated plasma TMAO has been suggested to be a risk factor for cardiovascular disease and T2DM so could also be relevant to NAFLD.[50]

There are several limitations to our study. NAFLD is a common phenotype in the general population, affecting up to 25% of individuals in Europe.[51] Our population controls cannot therefore be considered to be entirely free of NAFLD and there is no way of investigating this further. Our use of large numbers of controls with genetic matching helps mitigate the risk that this will lead to an underestimate of genuine genetic risk factors but does not eliminate it entirely. We undertook some "case only" studies, which included a small group of patients with biochemical evidence of NAFLD but liver biopsies showing steatosis below the normal disease definition, to further mitigate this. It is generally accepted that histological interpretation of liver biopsies is subject to some inter-observer variation, even amongst experienced hepatopathologists.[19,52] This is therefore inherent to a histopathological phenotype. However, all data used in the analysis were generated by highly experienced liver pathologists based in tertiary centres and, to further mitigate against this issue, the majority of liver biopsies were scored by a member of the project's central pathology team. Finally, our replication cohort was not perfectly matched with our discovery cohort in terms of disease severity and factors such as sex, T2DM and BMI. This is due, at least in part, to this being from a single centre from Southern Europe where NAFLD risk factors such as diet may be different to those further north in the continent, resulting in lower obesity rates within the NAFLD population.[53] We were unfortunately not able to identify another suitable European replication cohort involving patients who had undergone liver biopsy following referral to a hepatology clinic.

In conclusion, this relatively large GWAS of histologically characterised NAFLD cases has confirmed previously reported associations and provided evidence for 4 novel signals. Much larger meta analyses may be helpful in investigating the relevance of these novel signals.

Abbreviations

ALT, alanine aminotransferase; GWAS, genome-wide association study; HTGC, hepatic triglyceride content; NAFLD, non-alcoholic

fatty liver disease; NAS, NAFLD activity score; NASH, non-alcoholic steatohepatitis; OR, odds ratio; PCA, principal component analysis; SAF, steatosis, activity, and fibrosis; SNP, single nucleotide polymorphism; T2DM, type 2 diabetes mellitus; TMAO, trimethylamine N-oxide.

Financial support

This study has been supported by the EPoS (Elucidating Pathways of Steatohepatitis) consortium funded by the Horizon 2020 Framework Program of the European Union under Grant Agreement 634413, the FLIP consortium (European Union FP7 grant agreement 241762) and the Newcastle NIHR Biomedical Research Centre.

Conflict of interest

Quentin Anstee reports grants from European Commission during the conduct of the study; other from Acuitas Medical, grants, personal fees and other from Allergan/Tobira, other from E3Bio, other from Eli Lilly & Company Ltd, other from Galmed, grants, personal fees and other from Genfit SA, personal fees and other from Gilead, other from Grunthal, other from Imperial Innovations, grants and other from Intercept Pharma Europe Ltd, other from Inventiva, other from Janssen, personal fees from Kenes, other from MedImmune, other from NewGene, grants and other from Pfizer Ltd, other from Raptor Pharma, grants from GlaxoSmithKline, grants and other from Novartis Pharma AG, grants from AbbVie, personal fees from BMS, grants from GSK, other from NGMBio, other from Madrigal, other from Servier, outside the submitted work; Dina Tiniakos reports consultation fees from Intercept Pharmaceuticals Inc, Allergan, Cirius Therapeutics and an educational grant from Histoindex Pte Ltd; Guruprasad *P.* Aithal reports institutional consultancy income outside the scope of this study from GSK and Pfizer; Michael Allison reports consultancy/advisory with MedImmune/Astra Zeneca, E3Bio, honoraria from Intercept, Grant support from GSK, Takeda; Jean-Francois Dufour reports advisory committees with AbbVie, Bayer, BMS, Falk, Genfit, Genkyotex, Gilead Science, HepaRegenix, Intercept, Lilly, Merck, Novartis and speaking and teaching with Bayer, BMS, Intercept, Genfit, Gilead Science, Novartis; Pietro Invernizzi reports grants from Intercept, Gilead and Bruschettini; Mattias Ekstedt reports personal fees from AbbVie, AstraZeneca, Albireo, Diapharma, Gilead and non-financial support from Echosens (through LITMUS IMI project); Karine Clement has no personal honoraria but has consultancy and scientific collaboration activity for LNC therapeutics, Confotherapeutics and Danone Research; Jörn M. Schattenberg reports grants from Gilead and Boehringer Ingelheim and fees from Gilead, Boehringer Ingelheim, Galmed, Genfit, Intercept, Novartis, Pfizer and AbbVie outside the submitted work. All other authors report no conflicts of interest.

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

Authors' contributions

Study concept and design: QMA, CPD, AKD; acquisition of data: QMA, CPD, LV, MM, DT, ADB, PB, OG, JP, YL-L, GPA, MA, HY-J, MV, J-FD, PI, DP, ME, SK, SF, SP, EB, KC, VR, JMS; analysis and interpretation of data: HJC, RD, QMA, SC, AKD; drafting of the manuscript: AKD, RD, HJC, QMA; critical revision of the manuscript for important intellectual content: all; statistical analysis: RD, HJC; obtained funding: QMA, CPD, AKD; administrative, technical, or material support: OG, JP, YLL; study supervision: QMA, HJC, CPD, LV, AKD.

Acknowledgements

We are grateful to Julian Leathart for technical help, Lee Murphy and colleagues (Edinburgh CRF) for their assistance with GWAS provision, Elsbeth Henderson, the liver theme of the Newcastle NIHR Biomedical Research Centre (BRC), the gastrointestinal and liver disorder theme of Nottingham NIHR BRC (reference no BRC-1215-20003) and the Assistance Publique Hôpitaux de Paris for help with patient recruitment, Kristy Wonders for study management, Anna Fracanzani for helpful discussions, Michael Lowe for contributing to statistical genetics analysis and Daniele Cusi (Hypergenes) for provision of control data.

The EPoS Consortium Investigators

Newcastle University, UK
Quentin M. Anstee, Simon Cockell, Heather J. Cordell, Ann K. Daly, Rebecca Darlay, Christopher *P.* Day, Olivier Govaere, Katherine Johnson, Yang-Lin Liu, Fiona Oakley, Jeremy Palmer, Helen Reeves, Dina Tiniakos, Kristy Wonders

University of Turin, Italy
Elisabetta Bugianesi, Fabio Marra, Maurizio Parola, Chiara Rosso, Ramy Younes

University of Cambridge, UK
Michael Allison, Sergio Rodriguez Cuenca, Vanessa Pellegrinelli, Michele Vacca, Antonio Vidal-Puig

ICAN, France
Karine Clement, Raluca Pais, Vlad Ratziu, Timothy Schwartz

University of Mainz, Germany
Jörn M. Schattenberg, Detlef Schuppan

Consiglio Nazionale Delle Ricerche, Italy
Amalia Gastaldelli

University of Orebro, Sweden
Tuulia Hyötyläinen, Matej Orešič

University of Helsinki, Finland
Hannele Yki-Järvinen, Panu K. Luukkonen

Nordic Biosciences, Denmark
Morten Karsdal, Diana Leeming, Mette Juul Nielsen

University Hospital of Zurich, Switzerland
Felix Stickel

IXSCIENT, UK
Dave Wenn

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2020.04.003.

References

[1] Anstee QM, Targher G, Day CP. Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. Nat Rev Gastroenterol Hepatol 2013;10:330–344.
[2] McPherson S, Hardy T, Henderson E, Burt AD, Day CP, Anstee QM. Evidence of NAFLD progression from steatosis to fibrosing-steatohepatitis using paired biopsies: implications for prognosis and clinical management. J Hepatol 2015;62:1148–1155.
[3] Anstee QM, Day CP. The genetics of NAFLD. Nat Rev Gastroenterol Hepatol 2013;10:645–655.
[4] Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 2008;40:1461–1465.
[5] Speliotes EK, Yerges-Armstrong LM, Wu J, Hernaez R, Kim LJ, Palmer CD, 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:e1001324.
[6] Kozlitina J, Smagris E, Stender S, Nordestgaard BG, Zhou HH, Tybjaerg-Hansen A, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 2014;46:352–356.
[7] Valenti L, Al-Serri A, Daly AK, Galmozzi E, Rametta R, Dongiovanni P, et al. Homozygosity for the patatin-like phospholipase-3/adiponutrin I148M polymorphism influences liver fibrosis in patients with nonalcoholic fatty liver disease. Hepatology 2010;51:1209–1217.
[8] Liu YL, Reeves HL, Burt AD, Tiniakos D, McPherson S, Leathart JB, et al. TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. Nat Commun 2014;5:4309.
[9] Liu YL, Patman GL, Leathart JB, Piguet AC, Burt AD, Dufour JF, et al. Carriage of the PNPLA3 rs738409 C>G polymorphism confers an increased risk of non-alcoholic fatty liver disease associated hepatocellular carcinoma. J Hepatol 2014;61:75–81.
[10] Grimaudo S, Pipitone RM, Pennisi G, Celsa C, Camma C, Di Marco V, et al. Association between PNPLA3 rs738409 C>G variant and liver-related outcomes in patients with non-alcoholic fatty liver disease. Clin Gastroenterol Hepatol 2020;18:935–944.e3.
[11] Abul-Husn NS, Cheng X, Li AH, Xin Y, Schurmann C, Stevis P, et al. A protein-truncating HSD17B13 variant and protection from chronic liver disease. N Engl J Med 2018;378:1096–1106.
[12] Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, Van der Harst P, et al. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. Nat Genet 2011;43:1131–1138.
[13] Ma Y, Belyaeva OV, Brown PM, Fujita K, Valles K, Karki S, et al. 17-Beta hydroxysteroid dehydrogenase 13 is a hepatic retinol dehydrogenase associated with histological features of nonalcoholic fatty liver disease. Hepatology 2019;69:1504–1519.
[14] Pirola CJ, Garaycoechea M, Flichman D, Arrese M, San Martino J, Gazzi C, et al. Splice variant rs72613567 prevents worst histologic outcomes in patients with nonalcoholic fatty liver disease. J Lipid Res 2019;60:176–185.
[15] Namjou B, Lingren T, Huang Y, Parameswaran S, Cobb BL, Stanaway IB, et al. GWAS and enrichment analyses of non-alcoholic fatty liver disease identify new trait-associated genes and pathways across eMERGE network. BMC Med 2019;17:135.
[16] DiStefano JK, Kingsley C, Craig Wood G, Chu X, Argyropoulos G, Still CD, et al. Genome-wide analysis of hepatic lipid content in extreme obesity. Acta Diabetol 2015;52:373–382.
[17] Vargas V, Allende H, Lecube A, Salcedo MT, Baena-Fustegueras JA, Fort JM, et al. Surgically induced weight loss by gastric bypass improves non alcoholic fatty liver disease in morbid obese patients. World J Hepatol 2012;4:382–388.
[18] Cordell HJ, Han Y, Mells GF, Li Y, Hirschfield GM, Greene CS, et al. International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. Nat Commun 2015;6:8019.
[19] Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005;41:1313–1321.
[20] Bedossa P. Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease. Hepatology 2014;60:565–575.
[21] Daly AK, King BP, Leathart JB. Genotyping for cytochrome P450 polymorphisms. Methods Mol Biol 2006;320:193–207.
[22] Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, et al. Genes mirror geography within Europe. Nature 2008;456:98–101.
[23] Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999;55:997–1004.
[24] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575.
[25] Dulai PS, Singh S, Patel J, Soni M, Prokop LJ, Younossi Z, et al. Increased risk of mortality by fibrosis stage in nonalcoholic fatty liver disease: systematic review and meta-analysis. Hepatology 2017;65:1557–1565.
[26] Emdin CA, Haas M, Khera AV, Aragam K, Chaffin M, Klarin D, et al. A missense variant in Mitochondrial Amidoxime Reducing Component 1 gene and protection against liver disease. PLoS Genet 2020;16(4):e1008629.
[27] Santoro N, Zhang CK, Zhao H, Pakstis AJ, Kim G, Kursawe R, et al. Variant in the glucokinase regulatory protein (GCKR) gene is associated with fatty liver in obese children and adolescents. Hepatology 2012;55:781–789.
[28] Rees MG, Wincovitch S, Schultz J, Waterstradt R, Beer NL, Baltrusch S, et al. Cellular characterisation of the GCKR P446L variant associated with type 2 diabetes risk. Diabetologia 2012;55:114–122.
[29] He S, McPhaul C, Li JZ, Garuti R, Kinch L, Grishin NV, et al. A sequence variation (I148M) in PNPLA3 associated with nonalcoholic fatty liver disease disrupts triglyceride hydrolysis. J Biol Chem 2010;285:6706–6715.
[30] Smagris E, Gilyard S, BasuRay S, Cohen JC, Hobbs HH. Inactivation of Tm6sf2, a gene defective in fatty liver disease, impairs lipidation but not secretion of very low density lipoproteins. J Biol Chem 2016;291:10659–10676.
[31] Kedishvili NY. Retinoic acid synthesis and degradation. Subcell Biochem 2016;81:127–161.
[32] Zhong G, Kirkwood J, Won KJ, Tjota N, Jeong H, Isoherranen N. Characterization of vitamin A metabolome in human livers with and without nonalcoholic fatty liver disease. J Pharmacol Exp Ther 2019;370:92–103.
[33] Petta S, Miele L, Bugianesi E, Camma C, Rosso C, Boccia S, et al. Glucokinase regulatory protein gene polymorphism affects liver fibrosis in non-alcoholic fatty liver disease. PLoS One 2014;9:e87523.
[34] Brouwers M, Jacobs C, Bast A, Stehouwer CDA, Schaper NC. Modulation of glucokinase regulatory protein: a double-edged sword? Trends Mol Med 2015;21:583–594.
[35] Tin A, Balakrishnan P, Beaty TH, Boerwinkle E, Hoogeveen RC, Young JH, et al. GCKR and PPP1R3B identified as genome-wide significant loci for plasma lactate: the Atherosclerosis Risk in Communities (ARIC) study. Diabet Med 2016;33:968–975.
[36] Hudert CA, Selinski S, Rudolph B, Blaker H, Loddenkemper C, Thielhorn R, et al. Genetic determinants of steatosis and fibrosis progression in paediatric non-alcoholic fatty liver disease. Liver Int 2019;39:540–556.
[37] Di Costanzo A, Belardinilli F, Bailetti D, Sponziello M, D'Erasmo L, Polimeni L, et al. Evaluation of polygenic determinants of non-alcoholic fatty liver disease (NAFLD) by a candidate genes resequencing strategy. Sci Rep 2018;8:3702.
[38] Thompson B, Townsley F, Rosin-Arbesfeld R, Musisi H, Bienz M. A new nuclear component of the Wnt signalling pathway. Nat Cell Biol 2002;4:367–373.
[39] Xie YY, Mo CL, Cai YH, Wang WJ, Hong XX, Zhang KK, et al. Pygo2 regulates adiposity and glucose homeostasis via beta-Catenin-Axin2-GSK3beta signaling pathway. Diabetes 2018;67:2569–2584.
[40] Anstee QM, Goldin RD. Mouse models in non-alcoholic fatty liver disease and steatohepatitis research. Int J Exp Pathol 2006;87:1–16.
[41] Lu H, Sun J, Sun L, Shu X, Xu Y, Xie D. Polymorphism of human leptin receptor gene is associated with type 2 diabetic patients complicated with non-alcoholic fatty liver disease in China. J Gastroenterol Hepatol 2009;24:228–232.
[42] Zain SM, Mohamed Z, Mahadeva S, Cheah PL, Rampal S, Chin KF, et al. Impact of leptin receptor gene variants on risk of non-alcoholic fatty liver disease and its interaction with adiponutrin gene. J Gastroenterol Hepatol 2013;28:873–879.
[43] Yamamoto Y, Yamasuge W, Imai S, Kunisawa K, Hoshi M, Fujigaki H, et al. Lipopolysaccharide shock reveals the immune function of indoleamine 2, 3-dioxygenase 2 through the regulation of IL-6/stat3 signalling. Sci Rep 2018;8:15917.
[44] Vogel H, Kamitz A, Hallahan N, Lebek S, Schallschmidt T, Jonas W, et al. A collective diabetes cross in combination with a computational framework to dissect the genetics of human obesity and Type 2 diabetes. Hum Mol Genet 2018;27:3099–3112.

[45] Anstee QM, Seth D, Day CP. Genetic factors that affect risk of alcoholic and nonalcoholic fatty liver disease. Gastroenterology 2016;150: 1728–1744.e7.
[46] Brunt EM, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. Am J Gastroenterol 1999;94:2467–2474.
[47] Sookoian S, Flichman D, Garaycoechea ME, Gazzi C, Martino JS, Castano GO, et al. Lack of evidence supporting a role of TMC4-rs641738 missense variant-MBOAT7- intergenic downstream variant-in the susceptibility to nonalcoholic fatty liver disease. Sci Rep 2018;8:5097.
[48] Strnad P, Buch S, Hamesch K, Fischer J, Rosendahl J, Schmelz R, et al. Heterozygous carriage of the alpha1-antitrypsin Pi*Z variant increases the risk to develop liver cirrhosis. Gut 2019;68:1099–1107.
[49] Petta S, Valenti L, Tuttolomondo A, Dongiovanni P, Pipitone RM, Camma C, et al. Interferon lambda 4 rs368234815 TT>deltaG variant is associated with liver damage in patients with nonalcoholic fatty liver disease. Hepatology 2017;66:1885–1893.
[50] Ufnal M, Zadlo A, Ostaszewski R. TMAO: a small molecule of great expectations. Nutrition 2015;31:1317–1323.
[51] Estes C, Anstee QM, Arias-Loste MT, Bantel H, Bellentani S, Caballeria J, et al. Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016-2030. J Hepatol 2018;69:896–904.
[52] Kleiner DE, Brunt EM, Wilson LA, Behling C, Guy C, Contos M, et al. Association of histologic disease activity with progression of nonalcoholic fatty liver disease. JAMA Netw Open 2019;2:e1912565.
[53] Chen F, Esmaili S, Rogers G, Bugianesi E, Petta S, Marchesini G, et al. Lean NAFLD: a distinct entity shaped by differential metabolic adaptation. Hepatology 2020;71:1213–1227.