Increased serum miR-193a-5p during non-alcoholic fatty liver disease progression: diagnostic and mechanistic relevance

Katherine Johnson[1], Peter J Leary[2], Olivier Govaere[1], Matthew J Barter[3], Sarah H Charlton[3], Simon J Cockell[2,3], Dina Tiniakos[1], Michalina Zatorska[1], Pierre Bedossa[1], M. Julia Brosnan[4§], Jeremy Cobbold[5], Mattias Ekstedt[6], Guruprasad P. Aithal[7], Karine Clément[8,9], Jörn M Schattenberg[10], Jerome Boursier[11], Vlad Ratziu[8,9], Elisabetta Bugianesi[12], Quentin M Anstee[1,13*], Ann K Daly[1*] *on behalf of the LITMUS Consortium Investigators*[†]

[1]Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

[2]Bioinformatics Support Unit, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

[3]Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

[4]Internal Medicine Research Unit, Pfizer Inc, Cambridge, MA, USA

[5]Oxford Liver Unit, NIHR Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK

[6]Division of Diagnostics and Specialist Medicine, Department of Health, Medicine, and Caring Sciences, Linköping University, Linköping, Sweden

[7]NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and University of Nottingham, Nottingham, UK

[8]Institute of Cardiometabolism and Nutrition, Pitié Salpêtrière Hospital, Paris 75013, France

[9]Assistance Publique - Hopitaux de Paris, Paris 75004, France

[10]University Medical Center of Johannes Gutenberg-University Mainz, Mainz 55131, Germany

[11]Hepatology Department, Angers University Hospital, Angers, France

[12]University of Turin, Turin 10124, Italy

[13]Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Trust, Newcastle upon Tyne, UK

[*]These authors contributed equally

[§]Deceased

[†]LITMUS Consortium Investigators

Newcastle University, UK: Quentin M. Anstee, Pierre Bedossa, James Clark, Simon Cockell, Heather J. Cordell, Ann K. Daly, Rebecca Darlay, Christopher P. Day, Olivier Govaere, Tim Hardy, Katherine Johnson, Yang-Lin Liu, Fiona Oakley, Jeremy Palmer, Rachel Queen, Dina Tiniakos, Kristy Wonders, Michalina Zatorska

Academisch Medisch Centrum bij de Universiteit van Amsterdam, Netherlands: Patrick M Bossuyt, Adriaan G. Holleboom, Hadi Zafarmand, Yasaman Vali, Jenny Lee

Hôpital Beaujon, Assistance Publique: Hôpitaux de Paris, France and Institute of Cardiometabolism and Nutrition, Paris, France: Karine Clement, Raluca Pais, Vlad Ratziu

University Medical Centre Mainz, Germany: Jörn M Schattenberg, Detlef Schuppan

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

Örebro University, Sweden: Tuulia Hyötyläinen, Aidan McGlinchey, Matej Orešič, Partho Sen

Centre for Cooperative Research in Biosciences, Spain: Jose Mato, Óscar Millet

University of Bern, Switzerland: Jean-Francois Dufour

University of Oxford, UK: Jeremy Cobbold, Stephen Harrison, Stefan Neubauer, Michael Pavlides, Ferenc Mozes, Salma Akhtar

Perspectum Diagnostics Ltd, UK: Rajarshi Banerjee, Matt Kelly, Elizabeth Shumbayawonda, Andrea Dennis, Charlotte Erpicum

Servicio Andaluz de Salud, Spain: Manuel Romero-Gomez, Rocío Gallego-Durán, Isabel Fernández

Nordic Bioscience A/S, Denmark: Morten Karsdal, Diana Leeming, Mette Juul Fisker, Elisabeth Erhardtsen, Daniel Rasmussen, Per Qvist, Antonia Sinisi

Integrated BioBank of Luxembourg Foundation, Luxembourg: Estelle Sandt, Maria Manuela Tonini

University of Torino, Italy: Elisabetta Bugianesi, Maurizio Parola, Chiara Rosso

University of Firenze, Italy: Fabio Marra

Consiglio Nazionale Delle Ricerche, Pisa, Italy: Amalia Gastaldelli

University of Angers, France: Jerome Boursier

Antwerp University Hospital, Belgium: Sven Francque

Linköping University, Sweden: Mattias Ekstedt, Stergios Kechagias

University of Helsinki, Finland: Hannele Yki-Järvinen, Kimmo Porthan

University Medical Centre Utrecht, Netherlands: Saskia van Mil

National & Kapodistrian University of Athens, Greece: George Papatheodoridis

Faculdade de Medicina da Universidade de Lisboa, Portugal: Helena Cortez-Pinto

Università degli Studi di Milano, Italy: Luca Valenti

University of Palermo, Italy: Salvatore Petta

Università Cattolica del Sacro Cuore, Italy: Luca Miele

University Hospital Würzburg, Germany: Andreas Geier

Universitätsklinikum Aachen University, Germany: Christian Trautwein

University of Nottingham, UK: Guruprasad P. Aithal

Antaros Medical AB, Sweden: Paul Hockings

University of Birmingham, UK: Phil Newsome

iXscient Ltd, UK: David Wenn

Faculdade de Farmácia da Universidade de Lisboa, Lisbon, Portugal: Cecília Maria Pereira Rodrigues

Genfit SA, France: Rémy Hanf, Pierre Chaumat, Christian Rosenquist

Intercept Pharma Europe Ltd, UK: Aldo Trylesinski

One Way Liver S.L., Spain: Pablo Ortiz

Eli Lilly and Company Ltd, USA: Kevin Duffin

Pfizer Ltd, USA: Carla Yunis, Melissa Miller, M. Julia Brosnan, Theresa Tuthill

Boehringer-Ingelheim International GmbH, Germany: Judith Ertle, Ramy Younes

SomaLogic Inc., USA: Leigh Alexander, Rachel Ostroff

Novo Nordisk A/S, Denmark: Mette Skalshøi Kjær

Ellegaard Göttingen Minipigs AS, Denmark: Lars Friis Mikkelsen

Novartis Pharma AG, Switzerland: Clifford Brass, Lori Jennings, Maria-Magdalena Balp, Miljen Martic

Takeda Pharmaceuticals International GmbH, Switzerland: Guido Hanauer

AstraZeneca AB, Sweden: Sudha Shankar, Richard Torstenson

Echosens, France: Céline Fournier

Resoundant Inc., USA: Richard Ehman, Michael Kalutkiewicz, Kay Pepin

Bristol-Myers Squibb Company Corp., USA: Joel Myers, Diane Shevell

HistoIndex Pte Ltd, Singapore: Gideon Ho

Allergan: Henrik Landgren

Gilead Science Inc., USA: Rob Myers

Research Triangle Institute: Lynda Doward, Diane Whalley, James Twiss

Contact information

Prof. Ann K Daly PhD

Translational & Clinical Research Institute

Faculty of Medical Sciences

Newcastle University

Framlington Place

Newcastle upon Tyne

NE2 4HH

UK

Telephone: +44 191 208 7031

Email: a.k.daly@ncl.ac.uk

Prof. Quentin M Anstee PhD, FRCP

Translational & Clinical Research Institute

Faculty of Medical Sciences

Newcastle University

Fourth Floor, William Leech Building

Framlington Place

Newcastle upon Tyne

NE2 4HH

UK

Telephone: + 44 (0) 191 208 7012

Email: quentin.anstee@newcastle.ac.uk

Keywords: Micro RNA, non-alcoholic fatty liver disease, biomarker, sequencing

Electronic word count: 5650

Number of figures and tables 7

Conflicts of interest

The authors have no potential conflicts (financial, professional, or personal) directly relevant to the manuscript.

Financial support

This study has been performed as part of the LITMUS (Liver Investigation: Testing Marker Utility in Steatohepatitis) project which has received funding from the Innovative Medicines Initiative (IMI2) Program of the European Union under Grant Agreement 777377; this Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. It was also supported by the Newcastle NIHR Biomedical Research Centre and the European NAFLD Registry.

Author Contributions

Conceptualisation: KJ, OG, QMA and AKD; methodology: KJ, PJL and OG; software: PJL and SJC; validation: KJ; formal analysis: KJ, MZ and PJL; investigation: KJ, PJL, OG, MJB, SHC, DT and PB; resources: JC, ME, GA, KC, JMS, JB, VR, EB, QMA and AKD; data curation: KJ and PJL; writing – original draft: KJ, QMA and AKD; writing – review and editing: KJ, PJL, OG, MJB, SHC, SJC, DT, PB, MJB, JC, ME, GA, KC, JMS, JB, VR, EB, QMA and AKD; project supervision: QMA and AKD; funding acquisition: QMA and AKD.

Abstract

Background & Aims: Serum microRNAs (miRNAs) levels are known to change in non-alcoholic fatty liver disease (NAFLD) and may serve as useful biomarkers. This study aimed to profile miRNAs comprehensively at all NAFLD stages.

Methods: We profiled 2,083 serum miRNAs in a discovery cohort (183 NAFLD cases representing the complete NAFLD spectrum and 10 population controls). MiRNA libraries generated by HTG EdgeSeq were sequenced by Illumina NextSeq. Selected serum miRNAs were profiled in 372 additional NAFLD cases and 15 population controls by quantitative reverse transcriptase-polymerase chain reaction.

Results: Levels of 275 miRNAs differed between cases and population controls. Fewer differences were seen within individual NAFLD stages but miR-193a-5p consistently the showed increased levels in all comparisons. Relative to NAFL/NASH with mild fibrosis (stage 0/1), three miRNAs (miR-193a-5p, miR-378d and miR378d) were increased in cases with NASH and clinically significant fibrosis (stage 2-4), seven (miR193a-5p, miR-378d, miR-378e, miR-320b, c, d & e) increased in cases with NAFLD Activity Score (NAS) 5-8 compared with lower NAS, and three (miR-193a-5p, miR-378d, miR-378e) increased but one (miR-19b-3p) decreased in steatosis, activity, and fibrosis "activity" (SAF-A) score 2-4 compared with lower SAF-A. The significant findings for miR-193a-5p were replicated in the additional NAFLD cohort. Studies in Hep G2 cells showed that following palmitic acid treatment, miR-193a-5p expression decreased significantly. Gene targets for miR-193a-5p were investigated in liver RNAseq data for a case subgroup (n=80); liver *GPX8* levels correlated positively with serum miR-193a-5p.

Conclusions: Serum miR-193a-5p levels correlate strongly with NAFLD activity grade and fibrosis stage. MiR-193a-5p may have a role in the hepatic response to oxidative stress and is a potential clinically tractable circulating biomarker for progressive NAFLD.

1 # Introduction

2 NAFLD affects approximately one quarter of the adult general population and is the most
3 common cause of liver disease in the Western world.(1) The complex, multisystem disease is
4 characterised by intrahepatic triglyceride accumulation in the absence of excess alcohol
5 intake.(1, 2) In a significant minority of patients, NAFLD can progress from steatosis (non-
6 alcoholic fatty liver, NAFL) to its inflammatory form (NASH) with hepatocyte injury and
7 progressive fibrosis, and may lead to cirrhosis, hepatocellular carcinoma and end-stage liver
8 disease.(3)

9 NAFLD pathogenesis is multifactorial and so, in addition to metabolic comorbidities and
10 environmental influences, genetic and epigenetic factors confer an increased risk of NAFLD-
11 associated end-stage liver disease.(4, 5) Inflammatory disease severity may be assessed
12 histologically using one of two widely adopted semi-quantitative scoring systems: the NASH
13 CRN NAS (NAFLD Activity Score), which conflates the degree of steatosis with
14 steatohepatitic activity grade, and the FLIP SAF (Steatosis – Activity – Fibrosis) score, which
15 preserves the distinction between steatosis and activity.(6, 7) Histological scoring of a liver
16 biopsy remains the current reference standard for grading and staging NAFLD for clinical
17 trials, even though sampling error as well as inter- and intra-observer discrepancies are well
18 recognised to increase variability and result in misclassification.(8, 9) Consequently,
19 circulating biomarkers have been sought to circumvent the need for invasive biopsies.

20 Micro RNAs (miRNA) are small (~22 nucleotides in length) non-coding RNA molecules that
21 can post-transcriptionally regulate gene expression.(10) Mature miRNAs, complexed with
22 facilitative proteins (i.e. the Argonaute protein family), exert their effects by binding mRNA
23 molecules and inducing mRNA degradation or inhibiting translation.(11) Currently over
24 2,000 known human miRNAs are recorded on the miRBase database,(12) some of which
25 have been characterised as diagnostic biomarkers for diseases such as cardiovascular
26 disease(13) and cancer.(14) Despite several studies postulating the association of various
27 miRNAs with NAFLD, very few have been unequivocally replicated and validated.(15) The
28 lack of consensus may in part be due to differences in study design, relatively small numbers
29 of cases being studied, technical approaches and variances within the chosen techniques. The
30 most abundantly expressed miRNA known in adult liver, miR-122-5p, is possibly the
31 strongest candidate NAFLD biomarker, having been consistently corroborated in independent

32 studies.(16-18) More recently, levels of another miRNA, miR-34a-5p, have been shown to
33 increase in serum as NAFLD progresses and miR-34a-5p has been incorporated in an *in vitro*
34 diagnostic test to assist in the identification of patients with fibrosing-steatohepatitis.(19) Our
35 aim was to identify additional serum miRNAs biologically relevant to NAFLD
36 pathophysiology and progression with the capacity to serve as biomarkers. We report a
37 comprehensive and unbiased profile of over 2,000 miRNAs in a large international cohort of
38 histologically-characterised NAFLD patients, with separate discovery and replication
39 performed using diverse techniques in independent patient cohorts.

40 Materials and Methods

41 Patient recruitment and sample collection

42 The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki.
43 Cases were derived from the European NAFLD Registry (NCT04442334) with the collection
44 and use of NAFLD patient data and samples approved by the relevant Ethical Committees in
45 the participating centres, and all patients having provided informed consent.(20) The
46 collection of blood samples from 'healthy' controls (no reported comorbidities nor evidence
47 of NAFLD) was approved by the County Durham and Tees Valley Research Ethics
48 Committee (12/NE/012), and informed consent was given by all donors. Disease stage and
49 grade of NAFLD samples were histologically proven by liver biopsy, with the serum samples
50 taken within six months of said biopsy. Liver biopsy specimens were scored centrally by two
51 expert liver pathologists (DT, PB) to ensure consistency in the analysis and interpretation of
52 histology data. NAFLD was defined by the presence of steatosis in >5% of hepatocytes.
53 Cases were divided into those with Non-Alcoholic Fatty Liver (NAFL) if steatosis only or
54 steatosis with lobular inflammation were present and those with NASH which is
55 characterised by the presence of both lobular inflammation and hepatocellular ballooning in
56 addition to steatosis.(21, 22) NAS was calculated as the unweighted sum of steatosis,
57 ballooning, and lobular inflammation while SAF activity was calculated as the unweighted
58 sum of hepatocyte ballooning and lobular inflammation.(6, 7) NAFLD activity was graded as
59 low (NAS 0-4 or SAF activity 0-2) and high (NAS 5-8 or SAF activity 3-4), and fibrosis
60 staged as minimal fibrosis stage (F0-F1) and clinically significant (F2-F4).

61 **Serum sample RNA extraction, library preparation, sequencing and quality control**

62 Serum samples (15 µl) were processed through the HTG EdgeSeq platform (HTG Molecular
63 Diagnostics, Inc., Tucson, AZ, USA) using the EdgeSeq miRNA Whole Transcriptome
64 Assay according to the manufacturer's instructions. Briefly, the samples were lysed with a
65 1:1 ratio of HTG biofluids lysis buffer to extract RNA and the lysates input into nuclease
66 protection assays run on the HTG EdgeSeq processor. Sequence adapters were added to the
67 samples in a 16-cycle PCR reaction. PCR products were purified using AMPure XP beads
68 (Beckman Coulter, High Wycombe, UK) and quantified using a KAPA Library
69 Quantification Kit (KAPA Biosystems, Wilmington, MA, USA). The barcoded libraries were
70 pooled and 2,083 miRNAs (plus five negative controls, one positive control and 13
71 housekeeping genes) sequenced at the Genomics Core Facility, Newcastle University, UK,
72 using an Illumina NextSeq 500 System (Illumina Inc., Cambridge, UK) according to the
73 manufacturer's instructions (single end run, 50 bp). FASTQ files were processed by the pre-
74 installed HTG EdgeSeq parser software to align the probe sequences to the results, giving an
75 output of raw count data. Quality control was conducted using the raw counts according to
76 HTG guidelines. Delta mean values within two standard deviations (calculated by the
77 manufacturer as 4.8 for serum samples) were considered acceptable.

78 **MiRNA sequencing data analysis**

79 Data for the 2,083 sequenced miRNAs were analysed using RStudio version 1.0.143
80 (RStudio: Integrated Development for R. RStudio, Inc., Boston, MA,
81 http://www.rstudio.com/). Raw counts were first transformed into CPM (counts per million)
82 and the miRNAs with a CPM ≤100 were filtered out of the raw count dataset. DESeq2
83 version 1.18.1(23) was used to generate a dataset – corrected for batch, age and centre – from
84 which principal component analysis (PCA) plots were constructed. Outliers were defined by
85 visualisation of the PCA plots and removed from the dataset. Comparisons between each of
86 the histological groupings (control, NAFL, NASH-F0/F1, NASH-F2, NASH-F3 and NASH-
87 F4) and various subgroups were performed using limma version 3.34.9.(24) Statistical
88 significance was determined using a Benjamini-Hochberg adjusted p value ≤0.05; no logFC
89 ($\log_2$ fold-change) threshold was set. AUROC (area under the receiver operating
90 characteristic) curves were generated using IBM SPSS Statistics for Windows, version 24
91 (IBM Corp., Armonk, N.Y., USA) and significance between two AUROC curves was

92 assigned based on the DeLong, DeLong and Clarke-Pearson method(25) in RStudio version
93 1.0.143 (RStudio: Integrated Development for R. RStudio, Inc., Boston, MA,
94 http://www.rstudio.com/).

95 **MiRNA analysis in the replication cohort**

96 *Total RNA extraction for analysis by quantitative reverse transcriptase PCR*

97 Total RNA was extracted from serum using the MACHEREY-NAGEL NucleoSpin® miRNA
98 Plasma Kit (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany; distributed by
99 Fisher Scientific UK, Loughborough, UK) according to the manufacturer's instructions with
100 only slight modifications to the protocol. Serum (130 ul) was vortexed with 90 µl Buffer
101 MLP and incubated for 3 min at room temperature. As processing controls, 100 fmol of each
102 of a 5′ phosphorylated and a non-phosphorylated exogenous synthetic cel-miR-39-3p were
103 spiked-in (5′-UCACCGGGUGUAAAUCAGCUUG-3′; Integrated DNA Technologies, Inc.,
104 Coralville, IA, USA). Total RNA was eluted in 30 µl nuclease-free water and stored at -80ºC.

105 *Complementary DNA synthesis and quantitative reverse transcriptase PCR*

106 Using the TaqMan™ Advanced miRNA cDNA Synthesis Kit (Thermo Fisher Scientific,
107 Paisley, UK), complementary DNA (cDNA) was synthesised from 3 µl total RNA by reverse
108 transcriptase according to the manufacturer's instructions. The miR-Amp reaction product
109 was diluted 1:10 in nuclease-free water and stored at -20ºC. The expressions of three
110 miRNAs, including the phosphorylated spike-in, were quantified using pre-formulated
111 TaqMan™ Advanced miRNA Assays (Thermo Fisher Scientific, Paisley, UK): 478293_mir
112 (for the spiked-in cel-miR-39-3p), 477954_mir (hsa-miR-193a-5p) and 477855_mir (hsa-
113 miR-122-5p). Each quantitative polymerase chain reaction (qPCR) mixture contained 5 µl
114 diluted cDNA, 1 × TaqMan™ Fast Advanced Master Mix (Thermo Fisher Scientific, Paisley,
115 UK) and 1 × TaqMan™ Advanced miRNA Assay (Thermo Fisher Scientific, Paisley, UK),
116 and was made up to 15 µl with nuclease-free water.

117 TaqMan™ MicroRNA Reverse Transcription Kits (Thermo Fisher Scientific, Paisley, UK)
118 were used to synthesise cDNA from 5 µl total RNA according to the manufacturer's
119 instructions for miRNAs where analysis with material generated via the Advanced miRNA
120 kit was unsuccessful due to low levels. The reverse transcription reaction products were

121 diluted 1:10 in nuclease-free water and stored at -20°C. The expressions of three miRNAs,
122 including the non-phosphorylated spike-in, were quantified using pre-formulated standard
123 TaqMan™ MicroRNA Assays (Thermo Fisher Scientific, Paisley, UK): 000200 (cel-miR-39-
124 3p), 464645_mat (hsa-miR-3687) and 000426 (hsa-miR-34a-5p). Each qPCR reaction
125 mixture contained 2.5 µl diluted cDNA, 1 × TaqMan™ Fast Advanced Master Mix (Thermo
126 Fisher Scientific, Paisley, UK) and 1 × TaqMan™ Small RNA Assay (Thermo Fisher
127 Scientific, Paisley, UK), and was made up to 20 µl with nuclease-free water.

128 *Quantitative PCR procedure and analysis of replication cohort data*

129 All qPCR reactions were run in triplicate in MicroAmp™ Fast Optical 96-Well Reaction
130 Plates (Thermo Fisher Scientific, Paisley, UK) on a QuantStudio™ 5 Real-time PCR machine
131 (Thermo Fisher Scientific, Paisley, UK). PCR cycling conditions were 50°C for 2 min and
132 95°C for 20 sec, followed by 40 cycles of 95°C for 1 sec and 60°C for 20 sec. A reference
133 control, made of pooled cDNA, for each assay was run in triplicate on every plate, meaning
134 the raw cycle threshold (C_t) values could be corrected for batch effects. The batch corrected
135 data were analysed per miRNA by the $2^{-\Delta\Delta Ct}$ method (26). Data were plotted using GraphPad
136 Prism version 8.3.1 (GraphPad Software, San Diego, California USA; www.graphpad.com).
137 Statistical analyses were performed using Mann-Whitney *U* and Kruskal-Wallis one-way
138 analysis of variance tests.

139 **Target gene predictions and bioinformatic searches**

140 Two online databases were searched for predicted gene targets of miR-193a-5p –
141 TargetScan(27) and miRDB(28) and the overlapping genes between the two were identified.
142 Common genes were identified between the resulting predicted genes and the differentially
143 expressed genes in NAFLD liver as described by Govaere *et al*. (29) Normalised and
144 transformed counts of miR-193a-5p, *GPX8* and *COL1A1* were correlated for the 80
145 overlapping samples for which miRNA-seq and RNA-seq data were available; a linear model
146 was performed to obtain the slopes and *p* values of the correlations. Single cell RNA-seq data
147 for *GPX8* and *COL1A1* were extracted from the Liver Cell Atlas gene browser on 13 August
148 2020.(30) Data on gene expression were obtained from the Gene Page of the GTEx
149 (Genotype-Tissue Expression) Portal (GTEx Analysis Release V8; dbGaP Accession
150 phs000424.v8.p2).

151 **Hep G2 cell culture and treatment**

152 The human hepatoblastoma Hep G2 cell line (ATCC® HB-8065™, Middlesex, UK), was
153 cultured as described previously.(29) To assess effects on miRNA expression, cells were
154 treated for 24 hours with: (i) 1% bovine serum albumin-conjugated palmitic acid (250 µM),
155 (ii) 1% bovine serum albumin-conjugated oleic acid (500 µM), (iii) 1% bovine serum
156 albumin-conjugated palmitic and oleic acid (250/500 µM) combined solution, and (iv) 1%
157 bovine serum albumin as a lipid-loading control. All treatments were performed in triplicate.
158 Total RNA was isolated from the cells as described previously,(29) and cDNA was
159 synthesised and qPCR performed as described above using the TaqMan™ Advanced miRNA
160 cDNA Synthesis Kit and TaqMan™ Advanced miRNA assay for miR-193a-5p.

161 Results

162 **Study cohort description**

163 The current study comprised histologically characterised cases representing the full spectrum
164 of NAFLD severity from NAFL to NASH F4 (cirrhosis). The samples were divided between
165 discovery (*n*=183) and replication (*n*=372) cohorts. Discovery cases to cover the range of
166 NAFLD phenotypes were selected from those already recruited within the European NAFLD
167 Registry(20) by July 2018 with replication cases selected from the larger cohort available by
168 June 2019. Demographic and clinical data are reported in Table 1. There were some
169 significant differences between the cohorts. In particular, the discovery cohort included a
170 relatively large number of NAFL cases without detectable fibrosis or inflammation to allow
171 detailed comparisons between early and late disease. The replication cohort was significantly
172 older with more severe disease and included a wider spectrum of phenotypes. To obtain
173 baseline measurements for comparison with NAFLD cases, 10 individuals from a healthy
174 control group (3 male, 7 female, mean age 31 years) were included in the analysis with the
175 discovery cohort. A further 15 control individuals (9 male, 6 female, mean age 42 years)
176 were included in the replication cohort analyses.

177 **Comparison of NAFLD serum miRNA profiles with controls**

178 The serum miRNA profiles of 183 patients across the NAFLD spectrum and 10 population
179 controls were generated using the HTG EdgeSeq and Illumina NextSeq technologies. Initial
180 PCA confirmed that there was no clustering by centre, plate/batch or age, indicating that

181 batch correction was effective (Supplementary Figure 1). PCA showed that the control
182 samples appeared to cluster together, suggesting they had a different miRNA profile to the
183 NAFLD samples, although some overlap was observed (Figure 1). Relative to controls, levels
184 of 275 serum miRNAs were different in NAFLD patients, 165 increasing and 110 decreasing
185 (Supplementary Table 1). The most significantly different miRNA was miR-3687 for which
186 decreased levels in NAFLD were seen (fold change (FC) = 0.10, $p = 4.57\times10^{-28}$);
187 (Supplementary Table 2) When cases with NASH and different stages of fibrosis were
188 compared with the population controls, in excess of 200 significant differences in miRNA
189 levels were detected for each fibrosis stage, with the top 200 for each listed in Supplementary
190 Table 2. The most significant alteration was again the decrease seen for miR-3687 but for
191 NASH F0/F1, F2 and F3 only.

192 **Comparison of NAFLD miRNA profiles with NAFL as baseline**

193 We next conducted an analysis amongst NAFLD cases only, excluding population controls.
194 Using NAFL as the baseline, we initially looked for serum miRNA changes in NASH cases
195 generally; four miRNAs showed increased levels with no miRNAs decreased (Table 2), The
196 miRNA showing the largest increase was miR-193a-5p (FC 1.34, $p = 3.70\times10^{-03}$). We
197 subsequently characterised the miRNA profiles of NASH cases with particular stages of
198 fibrosis (F0-F4) against NAFL (Table 3). The miRNA profile for NASH-F0/F1 was similar to
199 that for NAFL; thereafter, progressive changes in miRNA levels were seen for cases with
200 NASH-F2, NASH-F3 and NASH-F4 (cirrhosis) suggesting that it was the stage of fibrosis
201 rather than grade of steatohepatitic activity that was driving the association. All changes seen
202 in F3 and F4 were increases in miRNA levels but for F2, both increases and decreases in
203 levels was detected. MiR-193a-5p showed the most significant change with a progressive
204 increase in levels from F2 (FC 1.32) via F3 (FC 1.65) to F4 (FC 1.84) (Table 3). These
205 increases in miR-193a-5p levels with increasing fibrosis score were also in line with those
206 seen for the NAFLD cases compared with population controls (Supplementary Table 2).

207 To assess the potential for distinguishing between mild and more clinically significant
208 disease, we undertook subgroup comparisons of miRNA profile in relation to disease activity
209 and fibrosis stage. Three miRNAs showed increased levels in clinically significant fibrosis
210 compared with minimal disease (Table 2); the most statistically significant of these was miR-
211 193a-5p (FC 1.35, $p = 1.79\times10^{-03}$). When the grade of steatohepatitis was considered using

212 NAS and SAF activity, seven and four miRNAs respectively showed altered levels, with all
213 but one of these changes an increase with increasing disease severity (Table 2). MiR-193a-5p
214 again showed the most significant increase (NAS FC = 1.36, $p = 2.61\times10^{-04}$; SAF activity FC
215 = 1.35, $p = 2.70\times10^{-02}$).

216 In light of its association with disease severity, we evaluated miR-193a-5p as a potential
217 circulating biomarker of disease stage and activity by assessing its diagnostic performance in
218 the sequencing dataset using AUROC statistics. AUROCs were also calculated and plotted
219 for miR-122-5p as well as simple composite scores (the aspartate aminotransferase/alanine
220 aminotransferase (AST/ALT) ratio (25, 26) and the FIB-4 score (27)). Relevant curves and
221 values are detailed in Supplementary Figure 2. The AUROC for miR-193a-5p was 0.92 for
222 discriminating NAFLD from population controls, while the values for differentiating mild
223 disease from significant fibrosis (F0-1 vs. F2-4), high NAS or high SAF-Activity were 0.68,
224 0.73 and 0.71 respectively. MiR-193a-5p significantly outperformed AST/ALT ratio when
225 discriminating between mild vs. significant fibrosis ($p = 0.04$), high NAS ($p = 6.00 \times 10^{-05}$)
226 and high SAF-Activity ($p = 0.03$). It was also more discriminating than FIB-4 as an indicator
227 of high NAS ($p = 3.40 \times 10^{-03}$) and exhibited comparable performance to FIB4 for fibrosis.
228 The discriminatory ability of miR-193a-5p was better than miR-122-5p for fibrosis (p =
229 0.019), NAS ($p = 0.012$) and SAF activity ($p = 1.53 \times10^{-03}$).

230 **Replication of miR193a-5, together with previously reported miRNAs, as predictors of**
231 **NAFLD progression using quantitative PCR in an independent cohort of samples**

232 We sought to replicate selected findings from our sequencing studies relating to miR-193-5p
233 and miR-3687 and additionally quantify serum levels of two miRNAs (miR-122-5p and miR-
234 34a-5p) that have previously been reported as relevant to NAFLD,(16-19) using the
235 independent replication cohort of 372 NAFLD cases, Healthy control samples (n=15) were
236 also included. Limited volumes of serum and inability to analyse using the Advanced miRNA
237 kit prevented replication of other interesting miRNA signals. Consistent with the results of
238 our sequencing analysis, miR-193a-5p showed significantly higher levels in NAFLD cases
239 with greater disease severity (i.e. more significant fibrosis and high NAS/SAF activity)
240 (Figure 2). However, the differences observed in miR-3687 levels did not vary significantly
241 with grade or stage of disease in the replication cohort (Figure 2). Levels of miR-34a-5p were
242 also significantly increased with a higher grade of NAFLD activity (by both NAS and SAF

243 activity) and with more significant fibrosis (Supplementary Figure 3). In contrast, miR-122-
244 5p levels did not show significance by fibrosis stage but levels increased significantly in
245 higher grades of NAFLD activity by both NAS and SAF activity (Supplementary Figure 3).

246 **Correlation of miR-193a-5p target gene predictions with differential hepatic gene**
247 **expression measured by RNA-seq in NAFLD**

248 In view of the significant increase in serum miR-193a-5p levels seen as disease activity
249 increased and its strong discriminatory ability, we sought to investigate the functional
250 significance of miR-193a-5p in disease progression. Bioinformatic analysis using
251 TargetScan(27) and miRDB(28) predicted that 204 and 327 genes, respectively, were targets
252 of miR-193a-5p, with an overlap of 78 genes between both prediction sets. To better examine
253 the transcriptional correlates of miR-193a-5p, we reanalysed our recently reported hepatic
254 transcriptomics dataset from 206 histologically-characterised NAFLD cases to identify genes
255 differentially expressed between mild and severe steatohepatitis and/or fibrosis.(29) We then
256 projected the list of 78 *a priori* predicted miR-193a-5p targets onto these data. The resulting
257 differentially expressed genes for each comparison are detailed in Figure 3. Nine predicted
258 gene targets of miR-193a-5p were significantly differentially expressed in the liver of patients
259 with greater steatohepatitic activity by NAS (*COL1A1*, *CRYBG3*, *GPX8*, *OLA1*, *RAP2A*,
260 *SLC7A1*, *XK*, *ANKS1A*, *IGF2*), three by SAF activity (*COL1A1*, *GPX8*, *IFFO2*), and 11 gene
261 targets were differentially expressed in the liver of patients with significant fibrosis
262 (*COL1A1*, *GPX8*, *IFFO2*, *NETO2*, *PCDHA12*, *SLC7A1*, *ZNF827*, *GOLGA6A*, *ITSN1*,
263 *KCNH1*, *KMT2A*). Additionally, the correlation between serum miR-193a-5p and levels in
264 the liver of the two genes that were differentially expressed in all three comparisons,
265 *COL1A1* and *GPX8*, was investigated in a subgroup of 80 samples for which both serum
266 miRNA-seq and liver RNA-seq data were available (Figure 3). A significant positive
267 correlation between miR-193a-5p and *GPX8* was observed (slope = 0.35, $p = 0.021$) and
268 there was a similar trend with *COL1A1* (slope = 0.29, $p = 0.074$). Moreover, single cell RNA-
269 seq data extracted from the Liver Cell Atlas gene browser,(30) showed that both *GPX8* and
270 *COL1A1* followed a similar pattern of increased expression in cirrhotic compared to
271 uninjured tissue in the mesenchyme (Supplementary Figure 4).

272 **Quantification of miR-193a-5p expression in a hepatoma cell line**

273 We utilised data from the GTEx Portal to visualise the tissue-specific expression of miR-
274 193a-5p, confirming that – in addition to the renal medulla, omentum and subcutaneous
275 adipose tissue, mammary tissue and cultured fibroblasts – miR-193a-5p was expressed in the
276 liver (Supplementary Figure 5). Considering this evidence, we characterised the expression of
277 miR-193a-5p *in vitro* in a model system of human hepatocytes in the context of lipid
278 accumulation as seen in NAFLD (Figure 4). Hep G2 cells were loaded with fatty acids and
279 miR-193a-5p expression quantified by qPCR. After a 24-hour treatment with lipotoxic
280 palmitic acid, miR-193a-5p expression was decreased significantly relative to the control
281 condition ($p = 0.009$). A less lipotoxic but more steatogenic lipid, oleic acid,(31) had a
282 negligible effect relative to the control condition, and, also in line with a previous report,(31)
283 there was no effect when the two lipids were combined.

284 **Discussion**

285 An unbiased profile of circulating miRNAs in NAFLD patients and controls has been
286 generated using high-throughput sequencing techniques. While other studies have identified
287 some changes in serum miRNA levels in NAFLD,(17, 32, 33) our study is the largest of its
288 kind – in terms of sample number and number of miRNAs analysed – spanning the full
289 spectrum of the disease including cirrhosis. Although targeting only known miRNAs and
290 unable to detect novel miRNAs, within these limitations we conducted a comprehensive and
291 unbiased profile of miRNAs in NAFLD. Our sequencing approach revealed hundreds of
292 miRNAs showing different levels at every NAFLD fibrosis stage relative to the controls,
293 suggesting that these levels might correlate with, and therefore potentially predict, disease
294 progression. We also saw numerous significant differences in miRNA levels within the entire
295 NAFLD group compared with controls. Altered serum miRNA levels could be a useful
296 diagnostic tool for NAFLD generally but identifying such changes was not an aim of this
297 study.

298 By combining the NAFLD sequencing data with an independent qPCR replication, we have
299 shown consistently that serum levels of miR-193a-5p can distinguish between mild and
300 advanced disease activity and fibrosis. Our finding for miR-193a-5p is in line with a recent
301 report that performed miRNA sequencing of plasma samples from NAFLD cases but
302 focussed on predominantly mild steatosis and fibrosis.(33) We also saw increased levels of

303 miR-122-5p, which was generally in line with previous reports, (16, 17, 34) apart from a
304 failure to detect significant differences with increasing fibrosis severity. This suggests that
305 miR-193a-5p is a more sensitive biomarker of NAFLD progression than miR-122-5p. In
306 addition to miR-193a-5p and miR-122-5p, we detected several other miRNAs that show
307 differences in levels as NAFLD progresses, though the effect was not as significant as for
308 miR-193a-5p and overall levels of these miRNAs were lower. These additional significant
309 miRNAs included miR-378d and miR-378e which both increased in NASH and advanced
310 fibrosis. This finding is in line with the recent sequencing study by others (33) who, while
311 reporting that these two miRNAs were present at higher levels in plasma from NAFLD cases
312 compared with controls and that they showed a significant relationship with extent of fibrosis,
313 also found lower levels in plasma compared with miR-193a-5p, both in NAFLD cases and
314 controls. We also saw significantly raised levels of miR-4484 in NASH compared with
315 NAFL. The previous sequencing study reported levels of this miRNA increased in severe
316 steatosis but did not include data on inflammation.(33) Our sequencing analysis did not detect
317 changes in levels of miR-34a, another previously reported miRNA increased in NAFLD,(19)
318 but we were able to confirm previous data for this relatively low abundance miRNA as a
319 significant marker for NAFLD progression by use of a highly sensitive qPCR assay.
320 Indicating its potential usefulness as a biomarker, serum miR-193a-5p achieved high
321 AUROC values in our study. There is increasing evidence that at least a proportion of
322 circulating miRNAs is encapsulated within extracellular vesicles in NAFLD.(35) The
323 sequencing approach we used detects all miRNAs within serum without distinguishing
324 between those present in vesicles and those circulating outside vesicles which are possibly
325 protein-bound.

326 There is evidence from other studies that levels of miR-193a-5p are decreased in liver tissue
327 in fibrosis (36) and in hepatocellular carcinoma,(37) and that hepatic miR-193a-5p levels
328 exhibit an inverse correlation with serum levels (36) similar to that reported for other
329 miRNAs including miR-122-5p.(17, 38) This would also be broadly consistent with our
330 observations in Hep G2 cells where proapoptotic palmitic acid (31) decreases miR-193a
331 levels but steatogenic oleic acid, either alone or in combination with palmitic acid, has no
332 significant effect. The miR-193a target genes found to show altered expression during
333 NAFLD in our transcriptomic analysis all increased in expression which is typical of genes
334 regulated by miRNA with a decreased miRNA level facilitating increased mRNA levels. In a
335 previous study, miR-193a-5p has been shown to act within a network of miRNAs to modify

336 the TGF-β-dependent regulation of extracellular matrix-related genes in hepatic fibrosis.(36)
337 It is likely, therefore, that miR-193a-5p in NAFLD will exert its effects on target genes as
338 part of a large interaction network over a longer time course rather than, for example, as an
339 immediate response to an acute stress. One of the other miRNAs that showed increased
340 serum levels in severe fibrosis, miR-378d, has also been demonstrated to show decreased
341 expression during fibrosis in an animal model.(39)

342 Significantly differentially expressed predicted gene targets of miR-193a-5p in our
343 transcriptomics data include *RAP2A*, *IGF2* and *SLC7A1*, which have all been associated with
344 hepatic function, liver disease and/or steatosis.(40-42) Most notably, we found *COL1A1* and
345 *GPX8* showed increased expression by all three comparisons. A separate comparison using
346 single cell RNA-seq data also showing increased expression in cirrhotic cells compared to
347 uninjured mesenchyme cells.(30) *COL1A1* encodes an extracellular matrix component that is
348 extensively deposited in the development of scarring and fibrosis.(43) Accordingly, it is a
349 highly biologically plausible and relevant target of miR193a-5p and has previously been
350 identified as part of a transcriptomic gene signature associated with advanced NAFLD.(29)
351 MiR-193a-5p has also been shown to interact with *COL1A1* in colorectal cancer cells and is
352 known to be involved in extracellular matrix deposition.(36, 44) It is therefore plausible that
353 secretion of miR-193a-5p by hepatocytes permits an increase in *COL1A1* expression in
354 stellate cells and thus contributes to hepatic fibrosis as proposed by others.(45)

355 GPX8 is an endoplasmic reticulum(ER)-resident member of the glutathione peroxidase
356 family that functions to protect cells from oxidative damage;(46, 47) oxidative damage may
357 in turn lead to ER stress and is associated with metabolically-driven NAFLD pathologies.(48)
358 Roy *et al*. have previously described the expression of miR-193a-5p in the liver and its role
359 as a member of a network of miRNAs modifying the TGF-beta-dependent regulation of
360 extracellular matrix-related genes expression in hepatic stellate cells to modify the balance
361 between deposition and resorption of liver fibrosis. (36) Recently, it has been demonstrated
362 that *GPX8* expression increases during epithelial to mesenchymal transition in several cell
363 lines with evidence that increased GPX8 affects IL-6/JAK-STAT3 signalling via the soluble
364 IL6 receptor.(49) Given that hepatic miR-193a-5p expression is known to exhibit an inverse
365 correlation with serum levels (36), we therefore postulate that the increase in serum miR-
366 193a-5p we observed in patients with advanced NAFLD is indicative of increased hepatocyte
367 export. This could induce increased *GPX8* expression in other hepatic cells as they respond to

368 oxidative stress and undergo further transitions during NAFLD progression. Indeed, we
369 observed a significant positive correlation between serum miR-193a-5p and hepatic *GPX8*.
370 Nonetheless, further investigations are required to identify the specific cells in which *GPX8*
371 acts, as low hepatocyte expression in the single cell RNA-seq dataset suggests that it may
372 exert its effects elsewhere in the liver.(30)

373 Published reports on miRNAs in NAFLD are seldom in complete agreement,(50) and as
374 such, one should be mindful of pertinent limitations. Discordance between studies could be
375 due to sampling and interpretation errors, and so the histological classification of patients'
376 disease might be inconsistent from the outset. Despite being a component of an *in vitro*
377 diagnostic test for NASH,(19) our sequencing analysis filtered out miR-34a-5p due to low
378 levels, yet subsequent qPCR was in fact able to detect it. Conversely, the strongest signal in
379 NAFLD relative to controls, miR-3687, was not confirmed by qPCR. These examples
380 highlight that sequencing should always be replicated independently using a different
381 methodology to test the strength of the original dataset. This approach combined with use of a
382 replication cohort enabled us to confirm the novel miR-193a-5p signal. However, the qPCR
383 approach also has limitations and we were unable to replicate all the associations found by
384 sequencing due to limited serum availability. A higher throughput approach for replication,
385 possibly using Nanostring or a custom microarray would be helpful if a further replication
386 cohort was available. Further studies on the functional significance of miR-193a-5p,
387 especially direct confirmation of effects on specific target genes, and the identification of
388 cells and possible extrahepatic tissues in which these interactions and downstream effects
389 occur are also still needed.

390 Overall, we have obtained an unbiased global profile of circulating miRNAs in NAFLD and
391 our study has identified a number of miRNAs that display changes in levels during disease
392 progression. In particular, the increase in circulating miR-193a-5p was replicated
393 independently, and we showed that several functionally relevant gene targets of this miRNA
394 were differentially expressed in liver tissue. We have also further investigated the reported
395 association of miR-122-5p and miR-34a-5p with NAFLD finding that miR-122-5p was not a
396 significant predictor of fibrosis progression but confirming that miR-34a-5p is a sensitive
397 though difficult to detect marker. MiR-193a-5p had a superior ability to discriminate between
398 mild and advanced NASH compared with AST/ALT and FIB-4 and could have value as an
399 additional biomarker for progression, possibly in combination with other biologically

400 relevant indices. The current study demonstrates how global miRNA profiling may provide
401 both new insights into disease pathophysiology and identify promising new biomarkers of
402 liver disease.

403 **List of abbreviations**

404 NAFL: non-alcoholic fatty liver

405 NASH: non-alcoholic steatohepatitis

406 NAS: NAFLD activity score

407 SAF activity: activity component of the “steatosis, activity and fibrosis” score

408 miRNA: micro RNA

409 LogFC: $\log_2$ fold-change

410 AUROC: area under the receiver operating characteristic

411 cDNA: complementary DNA

412 qPCR: quantitative polymerase chain reaction

413 AST/ALT: aspartate aminotransferase/alanine aminotransferase

414

References

1. Younossi ZM. Non-alcoholic fatty liver disease - A global public health perspective. J Hepatol 2019;70:531-544.
2. Anstee QM, Day CP. The genetics of NAFLD. Nat Rev Gastroenterol Hepatol 2013;10:645-655.
3. Anstee QM, Reeves HL, Kotsiliti E, Govaere O, Heikenwalder M. From NASH to HCC: current concepts and future challenges. Nat Rev Gastroenterol Hepatol 2019;16:411-428.
4. Anstee QM, Darlay R, Cockell S, Meroni M, Govaere O, Tiniakos D, Burt AD, et al. Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort. J Hepatol 2020;73:505-515.
5. Anstee QM, Seth D, Day CP. Genetic Factors That Affect Risk of Alcoholic and Nonalcoholic Fatty Liver Disease. Gastroenterology 2016;150:1728-1744 e1727.
6. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, Ferrell LD, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology 2005;41:1313-1321.
7. Bedossa P, Consortium FP. 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.
8. Ratziu V, Charlotte F, Heurtier A, Gombert S, Giral P, Bruckert E, Grimaldi A, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. Gastroenterology 2005;128:1898-1906.
9. Rinella ME, Tacke F, Sanyal AJ, Anstee QM, participants of the A-EW. Report on the AASLD/EASL joint workshop on clinical trial endpoints in NAFLD. J Hepatol 2019;71:823-833.
10. Pillai RS. MicroRNA function: multiple mechanisms for a tiny RNA? RNA 2005;11:1753-1761.
11. O'Brien J, Hayder H, Zayed Y, Peng C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Front Endocrinol (Lausanne) 2018;9:402.
12. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic Acids Res 2019;47:D155-D162.
13. Zhou SS, Jin JP, Wang JQ, Zhang ZG, Freedman JH, Zheng Y, Cai L. miRNAS in cardiovascular diseases: potential biomarkers, therapeutic targets and challenges. Acta Pharmacol Sin 2018;39:1073-1084.
14. Wang H, Peng R, Wang J, Qin Z, Xue L. Circulating microRNAs as potential cancer biomarkers: the advantage and disadvantage. Clin Epigenetics 2018;10:59.
15. Lopez-Riera M, Conde I, Quintas G, Pedrola L, Zaragoza A, Perez-Rojas J, Salcedo M, et al. Non-invasive prediction of NAFLD severity: a comprehensive, independent validation of previously postulated serum microRNA biomarkers. Sci Rep 2018;8:10606.
16. Dongiovanni P, Meroni M, Longo M, Fargion S, Fracanzani AL. miRNA Signature in NAFLD: A Turning Point for a Non-Invasive Diagnosis. Int J Mol Sci 2018;19.
17. Pirola CJ, Fernandez Gianotti T, Castano GO, Mallardi P, San Martino J, Mora Gonzalez Lopez Ledesma M, Flichman D, et al. Circulating microRNA signature in non-alcoholic fatty liver disease: from serum non-coding RNAs to liver histology and disease pathogenesis. Gut 2015;64:800-812.
18. Cheung O, Puri P, Eicken C, Contos MJ, Mirshahi F, Maher JW, Kellum JM, et al. Nonalcoholic steatohepatitis is associated with altered hepatic MicroRNA expression. Hepatology 2008;48:1810-1820.

19. Harrison SA, Ratziu V, Boursier J, Francque S, Bedossa P, Majd Z, Cordonnier G, et al. A blood-based biomarker panel (NIS4) for non-invasive diagnosis of non-alcoholic steatohepatitis and liver fibrosis: a prospective derivation and global validation study. Lancet Gastroenterol Hepatol 2020;5:970-985.
20. Hardy T, Wonders K, Younes R, Aithal GP, Aller R, Allison M, Bedossa P, et al. The European NAFLD Registry: A real-world longitudinal cohort study of nonalcoholic fatty liver disease. Contemp Clin Trials 2020;98:106175.
21. Bedossa P, Poitou C, Veyrie N, Bouillot JL, Basdevant A, Paradis V, Tordjman J, et al. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. Hepatology 2012;56:1751-1759.
22. Sanyal AJ, Brunt EM, Kleiner DE, Kowdley KV, Chalasani N, Lavine JE, Ratziu V, et al. Endpoints and clinical trial design for nonalcoholic steatohepatitis. Hepatology 2011;54:344-353.
23. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
24. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
25. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-845.
26. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001;25:402-408.
27. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife 2015;4.
28. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. Nucleic Acids Res 2020;48:D127-D131.
29. Govaere O, Cockell S, Tiniakos D, Queen R, Younes R, Vacca M, Alexander L, et al. Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. Sci Transl Med 2020;12:eaba4448.
30. Ramachandran P, Dobie R, Wilson-Kanamori JR, Dora EF, Henderson BEP, Luu NT, Portman JR, et al. Resolving the fibrotic niche of human liver cirrhosis at single-cell level. Nature 2019;575:512-518.
31. Ricchi M, Odoardi MR, Carulli L, Anzivino C, Ballestri S, Pinetti A, Fantoni LI, et al. Differential effect of oleic and palmitic acid on lipid accumulation and apoptosis in cultured hepatocytes. J Gastroenterol Hepatol 2009;24:830-840.
32. Tan Y, Ge G, Pan T, Wen D, Gan J. A pilot study of serum microRNAs panel as potential biomarkers for diagnosis of nonalcoholic fatty liver disease. PLoS One 2014;9:e105192.
33. Zhang X, Mens MMJ, Abozaid YJ, Bos D, Darwish Murad S, de Knegt RJ, Ikram MA, et al. Circulatory microRNAs as potential biomarkers for fatty liver disease: the Rotterdam study. Aliment Pharmacol Ther 2021;53:432-442.
34. Jampoka K, Muangpaisarn P, Khongnomnan K, Treeprasertsuk S, Tangkijvanich P, Payungporn S. Serum miR-29a and miR-122 as Potential Biomarkers for Non-Alcoholic Fatty Liver Disease (NAFLD). Microrna 2018;7:215-222.
35. Arrese M, Eguchi A, Feldstein AE. Circulating microRNAs: emerging biomarkers of liver disease. Semin Liver Dis 2015;35:43-54.
36. Roy S, Benz F, Vargas Cardenas D, Vucur M, Gautheron J, Schneider A, Hellerbrand C, et al. miR-30c and miR-193 are a part of the TGF-beta-dependent regulatory network controlling extracellular matrix genes in liver fibrosis. J Dig Dis 2015;16:513-524.

37. Roy S, Hooiveld GJ, Seehawer M, Caruso S, Heinzmann F, Schneider AT, Frank AK, et al. microRNA 193a-5p Regulates Levels of Nucleolar- and Spindle-Associated Protein 1 to Suppress Hepatocarcinogenesis. Gastroenterology 2018;155:1951-1966.e1926.
38. Butt AM, Raja AJ, Siddique S, Khan JS, Shahid M, Tayyab GU, Minhas Z, et al. Parallel expression profiling of hepatic and serum microRNA-122 associated with clinical features and treatment responses in chronic hepatitis C patients. Sci Rep 2016;6:21510.
39. Hyun J, Wang S, Kim J, Rao KM, Park SY, Chung I, Ha CS, et al. MicroRNA-378 limits activation of hepatic stellate cells and liver fibrosis by suppressing Gli3 expression. Nat Commun 2016;7:10993.
40. Zheng X, Zhao W, Ji P, Zhang K, Jin J, Feng M, Wang F, et al. High expression of Rap2A is associated with poor prognosis of patients with hepatocellular carcinoma. Int J Clin Exp Pathol 2017;10:9607-9613.
41. Lopez MF, Zheng L, Miao J, Gali R, Gorski G, Hirschhorn JN. Disruption of the Igf2 gene alters hepatic lipid homeostasis and gene expression in the newborn mouse. Am J Physiol Endocrinol Metab 2018;315:E735-E744.
42. Hatzoglou M, Fernandez J, Yaman I, Closs E. Regulation of cationic amino acid transport: the story of the CAT-1 transporter. Annu Rev Nutr 2004;24:377-399.
43. Liu X, Xu J, Brenner DA, Kisseleva T. Reversibility of Liver Fibrosis and Inactivation of Fibrogenic Myofibroblasts. Curr Pathobiol Rep 2013;1:209-214.
44. Sun J, Liu J, Zhu Q, Xu F, Kang L, Shi X. Hsa_circ_0001806 Acts as a ceRNA to Facilitate the Stemness of Colorectal Cancer Cells by Increasing COL1A1. Onco Targets Ther 2020;13:6315-6327.
45. Povero D, Eguchi A, Li H, Johnson CD, Papouchado BG, Wree A, Messer K, et al. Circulating extracellular vesicles with specific proteome and liver microRNAs are potential biomarkers for liver injury in experimental fatty liver disease. PLoS One 2014;9:e113651.
46. Brigelius-Flohe R, Kipp A. Glutathione peroxidases in different stages of carcinogenesis. Biochim Biophys Acta 2009;1790:1555-1568.
47. Ramming T, Hansen HG, Nagata K, Ellgaard L, Appenzeller-Herzog C. GPx8 peroxidase prevents leakage of H2O2 from the endoplasmic reticulum. Free Radic Biol Med 2014;70:106-116.
48. Pagliassotti MJ. Endoplasmic reticulum stress in nonalcoholic fatty liver disease. Annu Rev Nutr 2012;32:17-33.
49. Khatib A, Solaimuthu B, Ben Yosef M, Abu Rmaileh A, Tanna M, Oren G, Schlesinger Frisch M, et al. The glutathione peroxidase 8 (GPX8)/IL-6/STAT3 axis is essential in maintaining an aggressive breast cancer phenotype. Proc Natl Acad Sci U S A 2020;117:21420-21431.
50. Liu CH, Ampuero J, Gil-Gomez A, Montero-Vallejo R, Rojas A, Munoz-Hernandez R, Gallego-Duran R, et al. miRNAs in patients with non-alcoholic fatty liver disease: A systematic review and meta-analysis. J Hepatol 2018;69:1335-1348.

Table 1. Clinical data for the NAFLD serum samples used in sequencing and quantitative PCR replication

	Sequencing cohort (n=183)	Replication cohort (n=372)	p value
Patient demographic and clinical characteristics			
Sex (% female)	40%	41%	0.79
Age in years (Median (IQR))	54 (44-60)	56 (47-64)	0.007
BMI (Median (IQR))	31.6 (27.7-36.1)	31.5 (28.4-36.5)	0.46
AST in IU/L (Median (IQR))	37 (27-52)	42 (30-60)	0.014
ALT in IU/L (Median (IQR))	53 (36-79)	57 (38-84)	ns
Histological characteristics			
Fibrosis grade			<0.0001
F0	59 (32%)	43 (12%)	
F1	21 (11%)	127 (34%)	
F2	47 (26%)	66 (18%)	
F3	38 (21%)	92 (25%)	
F4	18 (10%)	44 (12%)	
NASH			
Yes	128 (70%)	315 (85%)	<0.0001
No	55 (30%)	57 (15%)	
NAS score >4			0.93
Yes	83 (45%)	171 (46%)	
No	100 (55%)	201 (54%)	
SAF activity score >2			0.52
Yes	67 (37%)	147 (40%)	
No	116 (63%)	225 (60%)	

p-values for continuous or categorical datasets were estimated by Mann–Whitney and Chi Square/Fisher's tests respectively

Table 2. Differences in miRNA levels detected in the NAFLD-only sequencing analysis

NASH grouped relative to NAFL

miRNA	LogFC	LogFC SE	Adj.p.val.†
miR-193a-5p	0.49	0.11	3.70E-03
miR-378d	0.37	0.09	4.97E-03
miR-378e	0.35	0.09	1.01E-02
miR-4484	0.35	0.09	2.05E-02

NASH with significant fibrosis (F2-F4) relative to NAFL/NASH with minimal fibrosis (F0/F1)

miRNA	LogFC	LogFC SE	Adj.p.val.†
miR-193a-5p	0.43	0.10	1.79E-03
miR-378e	0.34	0.07	1.79E-03
miR-378d	0.33	0.07	1.85E-03

High NAS (NAS 5-8) relative to low NAS (NAS 1-4)

miRNA	LogFC	LogFC SE	Adj.p.val.†
miR-193a-5p	0.44	0.09	2.61E-04
miR-378d	0.29	0.07	7.14E-03
miR-320d	0.32	0.09	2.39E-02
miR-378e	0.26	0.07	2.39E-02
miR-320b	0.32	0.09	2.89E-02
miR-320c	0.32	0.09	2.89E-02
miR-320e	0.29	0.08	2.89E-02

High SAF-Activity (SAF Activity 3-4) relative to low SAF-Activity (SAF Activity 0-2)

miRNA	LogFC	LogFC SE	Adj.p.val.†
miR-193a-5p	0.43	0.11	2.70E-02
miR-378d	0.34	0.09	2.70E-02

miR-378e	0.34	0.09	2.70E-02
miR-19b-3p	-0.74	0.20	2.73E-02

LogFC, log_2 fold-change; logFC SE, log_2 fold-change standard error, adj.p.val., adjusted *p* value.

† Adjusted *p* value significance threshold ≤0.05.

Table 3. miRNAs showing significantly different levels in NASH grouped by fibrosis stage

F2			F3			F4		
miRNA	**logFC**	**adj p val**	**miRNA**	**logFC**	**adj p val**	**miRNA**	**logFC**	**adj p val**
miR-193a-5p	0.41	4.77E-02	miR-193a-5p	0.72	2.00E-04	miR-193a-5p	0.88	2.32E-03
miR-19b-3p	-0.79	4.77E-02	miR-378d	0.48	4.97E-03	miR-4484	0.63	1.44E-02
miR-3135a	-0.46	4.77E-02	miR-378e	0.43	1.32E-02	miR-378g	0.51	1.79E-02
miR-378d	0.35	4.77E-02	miR-4484	0.45	1.32E-02	miR-6873-3p	-1.24	3.08E-02
miR-378e	0.35	4.77E-02	miR-378a-3p	0.38	1.50E-02			
miR-4534	-0.46	4.77E-02	miR-6780a-5p	0.82	2.59E-02			
miR-649	-0.59	4.77E-02						
miR-670-3p	0.37	4.77E-02						
miR-6769a-5p	0.22	4.77E-02						
miR-6852-5p	0.27	4.77E-02						
miR-25-3p	-0.61	4.83E-02						
miR-330-3p	0.25	4.83E-02						
miR-548ay-5p	-0.51	4.83E-02						
miR-548d-5p	-0.63	4.83E-02						

MiRNA levels were assessed by sequencing and are shown relative to NAFL. Adjusted p value threshold ≤0.05. F0/F1 showed no significant differences to NAFL

Figure legends

Figure 1. Principal component analysis plot for the sequencing data. Quality controlled, filtered and batch corrected miRNA sequencing data within the plot were coloured according to histological group to aid the identification of clustering by either the first or second principal component.

Figure 2. Replication by quantitative PCR (qPCR) of miR-193a-5p and miR-3687 associations. Levels of miR-193a-5p are shown for **(A)** significant fibrosis (NASH F2-F4) relative to minimal fibrosis (NAFL-NASH F0/F1) (n = 359), **(B)** advanced NAS (NAS 5-8) relative to mild NAS (NAS 0-4) (n = 359), and **(C)** advanced SAF activity (SAF activity 3-4) relative to mild SAF activity (SAF activity 0-2) (n = 359). Levels of miR-3687 are shown for **(D)** significant fibrosis (NASH F2-F4) relative to minimal fibrosis (NAFL-NASH F0/F1) (n = 371), **(E)** advanced NAS (NAS 5-8) relative to mild NAS (NAS 0-4) (n = 371), and **(F)** advanced SAF activity (SAF activity 3-4) relative to mild SAF activity (SAF activity 0-2) (n = 371). $2^{-\Delta\Delta Ct}$ was calculated relative to controls for all samples. Median values are shown with 95% confidence intervals. Mann-Whitney *U* tests were performed for all comparisons.

Figure 3. Differentially expressed predicted target genes of miR-193a-5p in liver RNA-seq. (A) RNA-seq data from liver tissue were analysed for three comparisons: advanced NAS (NAS 5-8) relative to mild NAS (NAS 1-4), advanced SAF activity (SAF activity 3-4) relative to mild SAF activity (SAF activity 0-2), and significant fibrosis (NASH F2-F4) relative to minimal fibrosis (NAFL-NASH F0/F1). The genes that overlapped with those predicted by *in silico* tools to be targets of miR-193a-5p are shown in the heatmap. Hierarchical clustering is based on levels of fold change in gene expression in the liver tissue. Statistical significance in the RNA-seq data was determined using a Benjamini-Hochberg adjusted p value ≤0.05: * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$. (B) and (C) Linear models of 80 overlapping NAFLD samples between the miRNA-seq and RNA-seq datasets for miR-193a-5p and **(B)** *GPX8*, and **(C)** *COL1A1*.

Figure 4. *In vitro* functional assessment of miR-193a-5p expression in Hep G2 cells. Hep G2 cells were treated with fatty acids (oleic acid (500 uM), palmitic acid (250 uM) or a combination of oleic (500 uM) and palmitic acid (250 uM)) for 24 hours. All treatments

were performed in triplicate and qPCR reactions of each were performed in triplicate. Data were normalised to the control condition (untreated) using the $2^{-\Delta Ct}$ method. An unpaired Student's *t*-test was performed for all conditions relative to the control conditions (** $p \leq 0.01$). Data are presented as the mean with error bars representing the standard error of the mean.