Serum coding and non-coding RNAs as biomarkers of NAFLD and fibrosis severity

Stefania Di Mauro1 | Alessandra Scamporrino1 | Salvatore Petta2 | Francesco Urbano1 | Agnese Filippello1 | Marco Ragusa3 | Maria T. Di Martino5 | Francesca Scionti5 | Stefania Grimaudo2 | Rosaria M. Pipitone2 | Graziella Privitera1 | Antonino Di Pino1 | Roberto Scicali1 | Luca Valenti6 | Paola Dongiovanni7 | Anna Fracanzani7 | Agata M. Rabuazzo1 | Antonio Craxì2 | Michele Purrello3 | Francesco Purrello1 | Salvatore Piro1

Abstract

Background & Aims: In patients with non-alcoholic fatty liver disease (NAFLD), liver biopsy is the gold standard to detect non-alcoholic steatohepatitis (NASH) and stage liver fibrosis. We aimed to identify differentially expressed mRNAs and non-coding RNAs in serum samples of biopsy-diagnosed mild and severe NAFLD patients with respect to controls and to each other.

Methods: We first performed a whole transcriptome analysis through microarray (n = 12: four Control: CTRL; four mild NAFLD: NAS ≤ 4 F0; four severe NAFLD NAS ≥ 5 F3), followed by validation of selected transcripts through real-time PCRs in an independent internal cohort of 88 subjects (63 NAFLD, 25 CTRL) and in an external cohort of 50 NAFLD patients. A similar analysis was also performed on liver biopsies and HepG2 cells exposed to olate:palmitate or only palmitate (cellular model of NAFL/NASH) at intracellular/extracellular levels. Transcript correlation with histological/clinical data was also analysed.

Results: We identified several differentially expressed coding/non-coding RNAs in each group of the study cohort. We validated the up-regulation of UBE2V1, BNIP3L mRNAs, RP11-128N14.5 IncRNA, TGFB2/TGFB2-OT1 coding/IncRNA in patients with NAS ≥ 5 (vs NAS ≤ 4) and the up-regulation of HBA2 mRNA, TGFB2/TGFB2-OT1 coding/IncRNA in patients with Fibrosis stages ≥ 3-4 (vs F = 0-2). In vitro models: UBE2V1, RP11-128N14.5 and TGFB2/TGFB2-OT1 had an increasing expression...
1 | BACKGROUND

Non-alcoholic fatty liver disease (NAFLD) accounts for the most increasing cause of chronic liver disease, hepatocellular carcinoma and of end-stage liver disease leading to liver transplantation.1,2

Cohort studies demonstrated that among NAFLD populations, those with non-alcoholic steatohepatitis (NASH) have a higher risk of fibrosis progression, and that the presence of severe liver fibrosis is the main driver of hepatic and extra-hepatic prognosis.2,3 Even if liver biopsy is considered the diagnostic gold standard, the availability of non-invasive markers to be used in NAFLD patients to predict NASH and/or severity of fibrosis represents a relevant medical need.4 To date, the diagnosis of NASH in clinical practice by using non-invasive scores is difficult because of the lack of well-performing and well-validated tools.5 Furthermore, novel-proposed NAFLD blood biomarkers, such as CK18-Asp396 fragments, showed low sensitivity for NASH detection.5,6 Different scores are available for the non-invasive assessment of fibrosis (eg ‘aspartate aminotransferase (AST)-to-platelet ratio index’ (APRI) and the ‘four factors-based fibrosis index’ (FIB-4)); some of them are very easy to use and implement in clinical practice even if there are limitations such as false-positive results and high uncertainty areas.5,7 When looking at imaging devices, liver stiffness measurement (LSM) through transient elastography (FibroScan) is a widely diffused tool, even if its accuracy can be influenced by obesity and severity of steatosis.7,8

Circulating RNAs in plasma or serum have been attracting exponential attention as novel non-invasive diagnostic biomarkers of several kinds of diseases. Body fluid RNAs, including long non-coding RNAs (lncRNAs) and messenger RNAs (mRNAs), have many of the essential characteristics of good biomarkers, for example: (i) they are non-invasive accessible; (ii) their levels can easily be determined by basic molecular biology methods, first of all real-time PCR; (iii) they are remarkably stable in spite of the high amounts of endogenous ribonuclease in body fluids such as serum and plasma.9-11

Several studies have highlighted the stability of lncRNAs and mRNAs, also under various experimental and pre-analytical oppressive conditions, including multiple freeze-thaw cycles, prolonged incubation at room temperature (up to 24h), exogenous RNase treatments, time delay in processing of blood after venipuncture, low/high pH.10,12-15

The stability of lncRNAs and mRNAs in different body fluids against endogenous ribonuclease degradation may be provided by vesicles encapsulation and RNA binding protein association and also their highly stable secondary structures.10,11,16,17

Two very recent studies reported the aberrant expression of long non-coding RNAs (lncRNAs) in liver tissue of NAFL/NASH patients and in mice models of NAFLD.18,19 Furthermore, increasing
experimental evidence is beginning to characterize the role of specific IncRNAs in the pathogenesis of liver fibrosis, and several metabolic functions, such as free fatty acid β-oxidation, lipogenesis and insulin secretion. Numerous studies performed high-throughput analysis of mRNAs in liver tissues from NAFLD patients or in vivo animal models, which allowed the identification of several pathways associated with disease progression. To date, no study has been performed on circulating mRNA and IncRNA signatures in sera from NAFLD patients. Therefore, the main aim of this work was to analyze the whole transcriptome profile in serum samples of NAFLD biopsy-diagnosed patients to identify novel non-invasive biomarkers which are able to identify NAFLD patients with NASH and/or advanced fibrosis.

2 | MATERIALS AND METHODS

2.1 | Study subjects

This study involved two subject cohorts, namely: a first study cohort recruited in University Hospital of Palermo which included 71 patients with biopsy-proven NAFLD and 29 controls and a second external cohort recruited in University Hospital of Milan which included 50 NAFLD patients. A liver biopsy was performed to confirm the diagnosis of NAFLD in patients with ultrasound findings of fatty liver and/or persistent (>6 months) elevation of alanine aminotransferase (ALT) or AST. We excluded patients with alcohol-induced or drug-induced liver disease autoimmune or viral hepatitis and cholestatic or genetic liver disease from our study. We also excluded patients with current or past consumption of ethanol of more than 20 g per day. This study was conducted according to the Declaration of Helsinki and was approved by the ethics committee of the University Hospital of Palermo and Milan. Before the procedures, a written informed consent was obtained from all patients participating in this study. Further information regarding histopathological evaluation and sample processing can be found as Supplementary Data.

3 | RNA EXTRACTION

We performed RNA extraction from serum samples of study cohorts, from liver biopsies of 12 subjects (four CTRL, four mild NAFLD NAS score ≤ 4 Fibrosis stage = 0, four severe NAFLD NAS ≥ 5 Fibrosis Stage = 3), and finally from HepG2 cells exposed to lipid treatment and their culture media. Total RNA was extracted by using miRNeasy mini kit (Qiagen). Further information regarding RNA extraction protocol can be found as Supplementary Data.

4 | MICROARRAY ANALYSIS

High-throughput profiling of serum coding/non-coding RNA, through microarray technology by Clariom D Pico Assay (Thermo Fisher Scientific), was performed on 12 study subjects: four mild NAFLD (NAS ≤ 4, Fibrosis stage = 0), four severe NAFLD (NAS ≥ 5, Fibrosis Stage = 3) and four CTRL. Further information regarding microarray experiments can be found as Supplementary Data.

5 | COMPUTATIONAL ANALYSIS

To understand the function of statistically significant deregulated transcripts for each comparison (mild NAFLD vs CTRL, severe NAFLD vs CTRL, severe vs mild NAFLD), we performed pathway enrichment analysis through Transcriprome Analysis Console v. 4 which retrieves canonical biological pathways from WikiPathways database and establishes P-values using two-sided Fisher’s Exact Test (P < 0.05).

6 | IN VITRO CELL CULTURE EXPERIMENTS

Human hepatoblastoma cell line HepG2 was cultured as previously reported. To obtain in vitro models of NAFL or NASH, HepG2 at 75% confluence were exposed to a mixture of oleate/palmitate or only palmitate at a final fatty acid concentration of 0.5 mM for 48 h respectively.

7 | SINGLE REAL-TIME PCR ASSAYS

Validation of candidate coding/non-coding RNAs both in an independent internal cohort of 88 study subjects (63 NAFLD and 25 CTRL) and an in external cohort of 50 NAFLD patients was performed through real-time PCR assays (Power SYBR Green RNA-to-CT1-Step Kit Thermo Fisher Scientific) according to the manufacturer’s protocol. Candidate transcripts were selected according to the following three criteria: (i) high levels of fluorescence microarray signals; (ii) low P-value in severe NAFLD vs mild NAFLD comparison (P ≤ 0.02); (iii) involvement in oxidative stress, inflammatory, apoptotic, autophagy or fibrogenic pathways according to the literature data. Validated transcript expression was also analyzed, through real-time PCR assays, in liver biopsies and in NAFL/NASH in vitro models at intracellular and extracellular level.

7.1 | STATISTICAL ANALYSIS

Differentially expressed (DE) coding and non-coding RNA identification was performed by Transcriprome Analysis Console (TAC) Software v.4, according to the following parameters: Analysis Type: Expression Gene; Summarization Method: Gene Level – RNA; Gene-Level P < 0.05; ANOVA Method: ebayes. GraphPad Prism (version 6.02) was used to: (i) perform unpaired t test among ΔCt values obtained after single real-time PCR assays; (ii) perform linear regression analysis between clinical/histological data and candidate coding/non-coding RNA -ΔCt values. Receiver operating characteristics
(ROC) curve analysis and multivariate logistic analyses were performed through SPSS PASW Statistics.

8 | RESULTS

8.1 | WHOLE TRANSCRIPTOME ANALYSIS

In order to identify novel potential biomarker signatures associated with NAFLD spectrum, we performed a whole transcriptome analysis in sera from four mild NAFLD patients (NAS ≤ 4; F = 0), four severe NAFLD patients (NAS ≥ 5; F = 3) and four healthy matched controls. Statistically significant deregulated transcripts in each comparison are represented as scatter plots in Figure 1A. We identified: 990 deregulated transcripts in mild NAFLD vs CTRL comparison; 1842 deregulated transcripts in severe NAFLD vs CTRL comparison and 1098 deregulated transcripts in severe NAFLD vs mild NAFLD comparison. Fold change (FC) deregulation counts of statistically significant DE coding and non-coding RNAs (ANOVA Method: ebayes P < 0.05) for each comparison are reported in Figure S1. RNA class percentage variations for each comparison are reported in Figure S2. Hierarchical clustering analysis of statistically significant deregulated transcripts showed distinguishable signatures in each comparison: mild NAFLD vs CTRL, severe NAFLD vs CTRL, severe NAFLD vs mild NAFLD (Figures S3, S4 and S5).

9 | PATHWAY ANALYSIS OF DIFFERENTIALLY EXPRESSED TRANSCRIPTS

To explore the potential functions of deregulated transcripts, we identified significantly enriched biological pathways through Transcriptome Analysis Console. This bioinformatics analysis demonstrated that DE transcripts were statistically significant (Fisher’s Exact Test P < 0.05) associated with several signalling pathways known to be involved in NASH/Fibrosis pathogenesis including inflammatory pathways (eg interleukins signalling pathways), metabolism deregulation (eg electron transport chain, TCA cycle nutrient utilization, Sterol Responsive Element Binding Protein, insulin signalling), cell death (eg apoptosis, ferroptosis), UPR stress (eg proteosome degradation, cytoplasmic ribosomal proteins), and extracellular matrix biosynthesis/fibrosis (eg collagen biosynthesis and modifying enzymes, VEGFA-VEGFR signalling pathway, EGF/EGFR signalling pathway) (Figure 1B).

10 | VALIDATION BY qPCRs

In order to confirm microarray data, we analysed the expression of candidate transcripts through qPCRs in a larger internal independent cohort of 88 subjects (CTRL = 25 and NAFLD = 63) and in an external cohort of 50 NAFLD patients. Clinical and demographic

![FIGURE 1](image-url) Scatter plots were used to assess the variation between mild NAFLD vs control (left panel); severe NAFLD vs CTRL (middle panel); severe NAFLD vs mild NAFLD (right panel) subjects. The values plotted on the X and Y axes are the averaged normalized signal values in each group (log2 scaled). Red points in the plot indicate > 2.0-fold up-regulation of expression, green points indicate > 2.0-fold down-regulation of expression and grey points indicate < 2.0-fold change in expression (panel A). Enriched Pathway analysis. Statistically significant pathways regulated by DE transcripts in mild NAFLD vs CTRL comparison (left panel); severe NAFLD vs CTRL comparison (middle panel); in severe NAFLD vs mild NAFLD comparison (right panel). Values plotted in X axes represent -log10 (P-value), Fisher’s Exact Test P < 0.05 (panel B)
characteristics of each cohort subjects are reported in Table 1. We selected the following transcripts: HBA2, UBE2V1, BNIIP3L coding RNAs, RP11-128N14.5 IncRNA and TGFβ2/TGFβ2-OT1 coding/non-coding RNA. We chose these transcripts because: (i) they presented high levels of fluorescence microarray signals; (ii) they presented a low P-value in severe NAFLD vs mild NAFLD comparison (P ≤ 0.02); (iii) they were potentially linked to oxidative stress, inflammatory, apoptotic, autophagy or fibrogenic pathways according to literature data.29,30,35 We also selected one ncRNA GenBank ID: AC020558.4 (‘AceView’ database gene name: ‘sherveebu’) as a down-regulated control molecule.

First, we grouped the validation NAFLD cohort according to the presence or absence of NASH (defined as simultaneous presence of steatosis, ballooning and lobular inflammation) and according to Kleiner NAS ≥ 5, associated to defined NASH diagnosis. According to microarray results, we confirmed the up-regulation of UBE2V1, RP11-128N14.5, BNIIP3L and TGFβ2/TGFβ2-OT1 in serum samples of patients with a NAS ≥ 5 with respect to those with NAS ≤ 4 (Figure 2A). Moreover, taking into account NASH diagnosis only RP11-128N14.5 reached statistical significance (NOT NASH n = 18 vs NASH n = 45 FC = 7.13, P = 2.00E-02) (Figure 2B).

We also divided the NAFLD patient cohort according to fibrosis severity. We compared patients with absent or mild to moderate fibrosis (F = 0-2) with respect to those with advanced fibrosis and cirrhosis (F = 3-4). According to this classification, we validated the up-regulation of HBA2, UBE2V1, BNIIP3L and TGFβ2/TGFβ2-OT1. Therefore HBA2 reached statistical significance only in F = 3-4 vs F = 0-2 comparison and RP11-128N14.5 reached statistical significance only in NAS ≥ 5 vs NAS ≤ 4 and NASH vs not NASH comparisons (Figure 3). It is important to note that, analysed transcripts were up-regulated also in F = 3-4 or NAS ≥ 5 with respect to control comparisons as previously showed in microarray data. Furthermore, we also validated AC020558.4 down-regulation in NAFLD patients with respect to controls. FC values and P-values of selected coding/non-coding RNAs, obtained through microarray analysis and qPCRs, are reported in Table S1.

The up-regulation of UBE2V1, RP11-128N14.5, BNIIP3L and TGFβ2/TGFβ2-OT1 in NAS ≥ 5 vs NAS ≤ 4 patients, RP11-128N14.5 in NAS vs simple steatosis and the up-regulation of HBA2, and TGFβ2/TGFβ2-OT1 in F = 3-4 vs F = 0-2 patients were also confirmed in the external validation cohort. Transcript FC deregulation and relative P-values obtained in the external validation cohort are reported in Figures S6 and S7.

11 | VALIDATED TRANSCRIPTS ARE ASSOCIATED WITH CLINICAL MARKERS AND HISTOLOGICAL SCORES

We analysed whether ΔCt values of validated transcripts were associated through a linear regression relationship both with the metabolic syndrome and liver damage-associated routine biochemical markers as with histological markers used in clinical practice to diagnose and stage NAFLD and fibrosis severity (Table S2).

UBE2V1 expression levels correlated with AST, ALT, triglycerides, Kleiner ballooning and NAS scores (Figure S8). BNIIP3L expression was directly associated with Kleiner lobular inflammation, ballooning and fibrosis scores (Figure S9). RP11-128N14.5 expression was associated with AST, Kleiner ballooning, NAS and fibrosis scores (Figure S10). TGFβ2/TGFβ2-OT1 was associated through a positive linear regression relationship with FIB-4, Liver Stiffness Measurements, Kleiner lobular inflammation and fibrosis scores (Figure S11). Finally, we found an inverse linear relationship among sherveebu AC020558.4 expression levels and triglycerides (Figure S12) and between HBA2 and total cholesterol levels. P-values and R² values are reported in Table S2.

12 | EXPRESSION ANALYSIS OF LIVER TISSUES AND IN VITRO MODELS

In order to further investigate the possible link between validated transcripts and cellular/molecular mechanisms associated with disease evolution, we evaluated whether RNA deregulation observed in serum of NAFLD patients was also present in liver tissue of NAFLD patients and in in vitro models of NAFL and NASH, both at intracellular and extracellular levels. Total RNA was extracted from liver biopsies of 12 subjects (four CTRL, four mild NAFLD NAS ≤ 4 Fibrosis stage = 0 and four severe NAFLD NAS ≥ 5 Fibrosis Stage = 3). We used an in vitro model that we already used in a previous work.27 HepG2 cells were exposed for 48 h to a mixture of oleate:palmitate (OA:PA) or only to palmitate (PA), to simulate simple steatosis or NASH respectively.

As shown in Figure 4A in liver tissue, we observed a statistically significant up-regulation of UBE2V1 mRNA, RP11-128N14.5 IncRNA and TGFβ2/TGFβ2-OT1 coding/non-coding RNA in severe NAFLD patients with respect to mild NAFLD patients and controls. HBA2 resulted to be statistically significant up-regulated only in severe NAFLD vs CTRL comparison. BNIIP3L and sherveebu AC020558.4 ncRNA did not reach statistical significance in analysed liver biopsies. In accordance with liver tissue data, UBE2V1 mRNA, RP11-128N14.5 IncRNA and TGFβ2/TGFβ2-OT1 coding/non-coding RNA were up-regulated in cells treated with PA with respect to those treated with OA:PA and to controls (Figure 4B). Differently from tissue data, intracellular HBA2 expression levels were unaffected by free fatty acid treatments and sherveebu AC020558.4 ncRNA presented a statistical significant increasing trend of expression ranging from NAFL to NASH in vitro models.

At the extracellular level, UBE2V1 mRNA and sherveebu AC020558.4 ncRNA were statistically significantly deregulated only in PA vs OA:PA condition comparison; medium-secreted RP11-128N14.5 IncRNA, BNIIP3L mRNA and TGFβ2/TGFβ2-OT1 coding/non-coding RNA were up-regulated in a statistically significant manner both in PA vs CTRL comparison as in PA vs OA:PA comparison (Figure 4B). Furthermore, it is important to note that HBA2 was undetectable in HepG2 culture medium.
13 | RECEIVER OPERATING CHARACTERISTICS CURVE ANALYSIS

By performing ROC curve analysis, we investigated the diagnostic performance of validated transcripts for \( F = 3-4 \) or \( \text{NAS} \geq 5 \) patient identification in NAFLD validation cohorts. The transcripts that reached statistical significance in ROC curve analysis were: (I) TGFB2/TGFB2-OT1 (internal cohort: \( AUC = 0.797, 95\% \text{ CI} = 0.675-0.918 \), sensitivity = 65\%, specificity = 81.3\%, \( P = 3.5E-04 \); external cohort: \( AUC = 0.786, 95\% \text{ CI} = 0.623-0.950 \), sensitivity = 62.5\%, specificity = 94.4\%, \( P = 4.4E-03 \) in \( F = 3-4 \) vs \( F = 0-2 \) comparison; (II) RP11-128N14.5

---

**TABLE 1** Clinical and demographic data of internal and external cohort

<table>
<thead>
<tr>
<th>Clinical variables</th>
<th>Clinical and demographic data of internal and external cohort</th>
<th>Controls vs NAFLD patients</th>
<th>NAFLD patients vs external cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>Unaffected controls (25) NAFLD patients (63) NAFLD (50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (% of male)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AST (IU/l)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALT (IU/l)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Platelets (×10⁹/l)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total cholesterol (mg/dL)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDL cholesterol (mg/dL)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Triglycerides (mg/dL)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIB-4 score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APRI score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LSM – KPa</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Histology**

- Kleiner steatosis grade (0-3)
- Kleiner lobular inflammation score (0-3)
- Kleiner Ballooning score (0-2)
- Kleiner NAS (0-8)
- Kleiner fibrosis stage (0-4)

Clinical and demographic indicators were checked for significant differences by t test or Mann-Whitney U-test, respectively for normally distributed or not normally distributed variables. The Chi-square test was used for categorical variables. The continuous normally distributed variables were represented as mean ± SD categorical and non-normal variables were summarized as median and percentage.

The bold values are significance.
FIGURE 2  Fold Change of UBE2V1 mRNA, RP11-128N14.5 IncRNA, BNIP3L mRNA, TGFB2/TGFB2-OT1 coding/ IncRNA validated through real-time PCRs in serum samples of patients with NAS score ≤ 4 or NAS score ≥ 5 respect to healthy controls (panel A). Fold Change of RP11-128N14.5 IncRNA validated through real-time PCRs in serum samples of patients with simple steatosis or defined NASH respect to healthy controls (panel B). n = 88:25 CTRL, 38 NAS ≤ 4, 25 NAS ≥ 5.18 NOT NASH, 45 NASH. t test *P ≤ 0.05,**P ≤ 0.02

FIGURE 3  Fold Change of HBA2 mRNA, UBE2V1 mRNA, BNIP3L mRNA, TGFB2/TGFB2-OT1 coding/ IncRNA validated through real-time PCRs in serum samples patients with Fibrosis Stages F = 0-2 or F = 3-4 respect to healthy controls. n = 88:25 CTRL, 37 F = 0-2, 26 F = 3-4. t test *P ≤ 0.05,**P ≤ 0.01,***P ≤ 0.001
FIGURE 4 Histograms of coding/noncoding RNA expression levels observed in liver biopsies (n = 12, four CTRL, four mild NAFLD: NAS ≤ 4 F = 0, four severe NAFLD NAS SCORE ≥ 5 F = 3) (panel A) and in OA:PA and PA treated HepG2 with respect to controls at intracellular and extracellular levels (panel B). OA:PA = oleate:palmitate; PA = palmitate; CTRL = control. FC = fold change
(internal cohort: AUC = 0.706, 95% CI = 0.554-0.857, sensitivity = 73.7%, specificity = 70.4%, \( P = 1.9E-02 \); external cohort: AUC = 0.694, 95% CI = 0.545-0.843, sensitivity = 78.3%, specificity = 63%, \( P = 2.0E02 \)) and BNI3P4L (internal cohort: AUC = 0.676, 95% CI = 0.512-0.841, sensitivity = 57.9%, specificity = 81.5%, \( P = 4.3E-02 \); external cohort: AUC = 0.686, 95% CI = 0.537-0.835, sensitivity = 51.9%, specificity = 91.3%, \( P = 3.0E02 \)) in NAS \( \geq 5 \) vs NAS \( \leq 4 \) comparison.

It is important to note that TGFβ2/TGFβ2-OT1 expression levels were associated with the blood-based index FIB-4 and LSM used in clinical practice for non-invasive diagnosis of F \( \geq 3 \) stages and histological Kleiner fibrosis stage. Comparison of the AUCs revealed the combinations TGFβ2/TGFβ2-OT1 plus FIB-4 (internal cohort: AUC = 0.891, 95% CI = 0.799-0.982, sensitivity = 80%, specificity = 87.5%, \( P = 3.0E-06 \); external cohort: AUC = 0.889, 95% CI = 0.771-0.998, sensitivity = 87.5%, specificity = 93.8%, \( P = 1.1E-04 \)) or TGFβ2/TGFβ2-OT1 plus LSM (AUC = 0.892, 95% CI = 0.802-0.983, sensitivity = 80%, specificity = 90.6%, \( P = 2.0E-06 \)) were superior to APRI (internal cohort: AUC = 0.759, 95% CI = 0.626-0.893, sensitivity = 55%, specificity = 93.8%, \( P = 1.8E-03 \); external cohort: AUC = 0.667, 95% CI = 0.479-0.854, sensitivity = 68.8%, specificity = 66.7%, \( P = 9.8E-02 \)); FIB-4 (internal cohort: AUC = 0.819, 95% CI = 0.700-0.937, sensitivity = 60%, specificity = 93.8%, \( P = 1.3E-04 \); external cohort: AUC = 0.839, 95% CI = 0.698-0.979, sensitivity = 93.8%, specificity = 66.7%, \( P = 7.7E-04 \)); TGFβ2/TGFβ2-OT1 (internal cohort: AUC = 0.797, 95% CI = 0.675-0.918, sensitivity = 65%, specificity = 91.3%, \( P = 3.5E-04 \); external cohort: AUC = 0.786, 95% CI = 0.623-0.950, sensitivity = 62.5%, specificity = 94.4%, \( P = 4.4E-03 \)) and LSM (AUC = 0.841, 95% CI = 0.723 –0.960, sensitivity = 85%, specificity = 81.3%, \( P = 4.0E-05 \)) (Figure 5, Figure S13).

In NAS \( \geq 5 \) patient discrimination, RP11-128N14.5 diagnostic performance was comparable or superior to AST performance (internal cohort: AUC = 0.707, 95% CI = 0.554-0.860 sensitivity = 73.7%, specificity = 74.1%, \( P = 1.8E-02 \); external cohort: AUC = 0.659, 95% CI = 0.470-0.849, sensitivity = 38.9, specificity = 99%, \( P = 8.3E02 \)) and superior to ALT (internal cohort: AUC = 0.367, 95% CI = 0.203-0.532, sensitivity = 94.7%, specificity = 3.7%, \( P = 1.3E-01 \); external cohort: AUC = 0.633, 95% CI = 0.449-0.817, sensitivity = 55.6%, specificity = 82.6%, \( P = 1.5E01 \)). Combination of RP11-128N14.5 with AST or ALT did not improve its diagnostic performance.

In NASH vs not NASH comparison, RP11-128N14.5 did not reach statistical significance; however, its AUC (internal cohort: AUC = 0.632, 95% CI = 0.477-0.787, sensitivity = 86.7%, specificity = 38.9%, \( P = 1.0E-01 \); external cohort: AUC = 0.650, 95% CI = 0.480-0.819, sensitivity = 53.3%, specificity = 4.3%, \( P = 9.7E-02 \)) was superior with respect to ALT (internal cohort: AUC = 0.412, 95% CI = 0.258-0.567, sensitivity = 25%, specificity = 83.3%, \( P = 2.9E-01 \); external cohort: AUC = 0.567, 95% CI = 0.390-0.744, sensitivity = 57.7%, specificity = 73.3%, \( P = 4.8E-01 \)) and similar to AST (AUC = 0.558, 95% CI = 0.397-0.720, sensitivity = 69.2%, specificity = 55.6%, \( P = 4.8E-01 \); external cohort AUC = 0.672, 95% CI = 0.503-0.842, sensitivity = 62.5%, specificity = 80%; \( P = 7.3E-02 \)). Combination of RP11-128N14.5 with AST or ALT in both cohorts decreased its AUC: 0.5 \( P = 1 \).

14 | DISCUSSION

In this multiphase case-control study, we performed an initial screening of coding mRNAs and ncRNAs in biopsy-proven NAFLD patients at different stages of histopathological severity and controls \((n = 12: 4 \text{CTRL, 4 NA} S \leq 4 \text{F} = 0, 4 \text{NA} S \geq 5 F = 3)\). Interestingly, performing computational analysis, we observed that deregulated transcripts, resulting from this initial screening, resulted to be statistically associated to key pathways known to be involved in fibrosis and NASH physiopathogenesis such as inflammation, UPR stress and metabolism. To better characterize our results, we chose, for subsequent qPCR validation, five DE transcripts in severe NAFLD vs mild NAFLD with a very compelling \textit{P}-value (\( \leq 0.02 \)) and potentially linked to NASH progression and fibrogenic pathways. The potential of validated transcripts for predicting disease severity (NAS \( \geq 5 \)) or advanced fibrosis (F = 3-4) in NAFLD patients was also evaluated. Interestingly, combinations of TGFβ2/TGFβ2-OT1 plus FIB-4 (AUC = 0.891, \( P = 3.00E-06 \)) or TGFβ2/TGFβ2-OT1 plus LSM (AUC = 0.892, \( P = 2.00E-06 \)) were able to powerfully discriminate patients with F 3-4 with respect to patients F 0-2. Therefore, the combination of coding/non-coding RNA expression values and routine clinical data could be conveniently applied as a diagnostic tool for non-invasive screening of NAFLD patients according to fibrosis stage. Moreover, the best-performing transcript for NAS \( \geq 5 \) patient identification, RP11-128N14.5, achieved an AUC similar to AST, and their combined use did not improve the diagnostic power. Therefore, the identification of novel NASH diagnosis biomarkers is still open and needs to be addressed.

In order to determine a possible link between selected coding/ non-coding RNAs and synergic and complex mechanisms underlying disease and fibrosis progression, we tested the hypothesis of a statistical association between validated transcript expression and histological features of NAFLD spectrum. In our study, we observed that UBE2V1, BNI3P3, RP11-128N14.5, TGFβ2/TGFβ2-OT1 were correlated with Kleiner histological scores: Kleiner NAS correlated with UBE2V1 (\( P = 4.0E-03 \)) and RP11-128N14.5 (\( P = 3.10E-02 \)); Kleiner lobular inflammation score correlated with BNI3P3 (\( P = 8.00E-03 \)) and TGFβ2/TGFβ2-OT1 (\( P = 3.10E-02 \)); Kleiner ballooning score correlated with UBE2V1 (\( P = 4.50E-02 \)), BNI3P3 (\( P = 2.60E-02 \)), RP11-128N14.5 (\( P = 1.20E-02 \)); Kleiner fibrosis score correlated with BNI3P3 (\( P = 3.60E-02 \)), RP11-128N14.5 (\( P = 3.70E-02 \)) and TGFβ2/TGFβ2-OT1 (\( P = 1.60E-02 \)). The statistical correlation with histological scores of disease severity and fibrosis stages suggests that analysed transcripts could be potentially associated with NASH and hepatic fibrogenesis molecular features.

Explaining the relationship between RNA extracellular expression and affected tissues is a difficult challenge. We tried to respond to this question by verifying whether RNA deregulation in patients could be somehow mirrored in liver tissue of NAFLD subjects and into in vitro models of NAFLD disease, both at intracellular and extracellular levels.
FIGURE 5 Univariate and Multivariate ROC curve analysis for predicting APRI, FIB-4, LSM, TGFB2/TGFB2-OT1, TGFB2/TGFB2-OT1 + Fibroscan as \( F = 3-4 \) patient diagnosis biomarkers with respect to \( F = 0-2 \) patients. APRI AUC = 0.759 (95% CI = 0.626-0.893, sensitivity = 55%, specificity = 93.8%, \( P = 1.80 \times 10^{-3} \)); TGFB2/TGFB2-OT1 + FIB4 AUC = 0.891 (95% CI = 0.799-0.982, sensitivity = 80%, specificity = 87.5%, \( P = 3.00 \times 10^{-6} \)). TGFB2/TGFB2-OT1 AUC = 0.797 (95% CI = 0.675-0.918, sensitivity = 65%, specificity = 81.3%, \( P = 1.80 \times 10^{-3} \)); FIB-4 AUC = 0.819 (95% CI = 0.700-0.937, sensitivity = 60%, specificity = 93.8%, \( P = 0-2 \) patients.

APRI AUC = 0.759 (95% CI = 0.626-0.893, sensitivity = 55%, specificity = 93.8%, \( F = 3-4 \) patient diagnosis biomarkers with respect to \( F = 0-2 \) patients. In support of the first hypothesis, it has been reported that oxidative stress induces an increased erythrocyte susceptibility to haemolysis in animal models of NASH and in obese subjects. Akyuz et al reported an increased expression of HBA1 and HBB in liver...
biopsies of NASH patients with respect to controls. They also demonstrated that H$_2$O$_2$ treatment of HepG2 cells increases gene and protein expression of HBA1 and HBB. Furthermore, haemoglobin over-expression induced oxidative stress decrease. In accordance with these results we observed a statistically significant up-regulation of HBA2 transcript in severe NAFLD patients respect to controls. However, despite the fact that palmitate is a known oxidative stress inducer, we did not observe any HBA2 gene expression alteration in our in vitro models and we were not able to detect HBA2 expression in culture medium. Probably, in NAFLD, the main contribution of the increase in HBA2 gene expression is given by other cell types, different from hepatocytes, inside the affected liver tissue.

Although any direct link among UBE2V1 gene expression and NAFLD has not been reported, UBE2V1 is involved in non-canonical IKK poly-ubiquitination, which in turn activates NF-kappa-B cascade, which is one of the key pro-inflammatory signalling pathways active in NASH. There are not any literature data concerning the involvement of RP11-128N14.5 in NAFLD. RP11-128N14.5 is a long intronic RNA transcribed in the same direction of its protein-coding host gene RAP2A. RP11-128N14.5, as already reported for several lncRNAs, could have a role in the epigenetic regulation of its host gene locus. There are no published data on the association between the RAP2A coding gene and NAFLD spectrum; however, it has been reported that in hepatocellular carcinoma, RAP2A regulates MAP4K4, which leads to JNK and NF-xB signalling activation which are the members of common molecular pathways of NASH pathogenesis.

TGF-β is a homodimer that exists in three different isoforms (-β1, -β2 and -β3) in mammals and all TGF-β ligands act through the same receptor signalling systems. It has been widely demonstrated that, in the liver, TGF-β signalling is involved in the fibrogenic response through hepatic stellate cell activation. Furthermore, it has been reported that TGF-β signalling in hepatocytes contributes to hepatocyte death and lipid accumulation that promote the development of NASH. In real-time validation experiments, we used primers able to amplify both mRNA encoding for TGFβ2 and IncRNA TGFβ2-Overlapping Transcript 1, a newly discovered IncRNA deriving from the 3’-UTR of TGFβ2. We found that levels of TGFβ2/TGFβ2-OT1 were highly increased in serum samples of patients with a NAS ≥ 5 and fibrosis F = 3-4. Several studies have highlighted the role of TGFβ2-OT1 in the regulation of autophagy, apoptotic, inflammatory and fibrogenic pathways, all known to be involved in NAFLD progression. It has been reported that TGFβ2-OT1 is involved in the negative regulation of apoptosis and positive regulation of autophagy and inflammation in vascular endothelial cells. Long Xiang Tu and colleagues reported the up-regulation of TGFβ2-OT1 in fibroblasts derived from hypertrophic scars, suggesting its involvement in excessive proliferation of fibroblasts and accumulation of extracellular matrix. As concerns TGFβ2, its pro-fibrogenic role and/or its increased gene or protein expression has been reported in lens fibrosis, kidney fibrosis, cardiac fibrosis and liver fibrosis. Several studies reported the involvement of BNIP3L in the regulation of pathways associated with NASH pathogenesis, specifically, apoptosis, necrosis, autophagy and fibrosis. Several studies supported BNIP3L involvement in apoptosis and necrosis of cardiomyocyte in heart diseases. HIF1b-dependent BNIP3L increased gene and protein expression has been reported in primary cultured hepatocytes exposed to ethanol and in mice models of alcohol-induced liver injury and steatosis. Although these data support the potential activation of this pathway in liver steatosis, we did not observe its activation in liver biopsies and in vitro model at intracellular level. Concerning the pro-fibrogenic role of BNIP3L, Wei Li and colleagues reported that the expression of BNIP3L is increased in cardiac fibrosis.

Although the major limitation of our study is the small sample size, our results may contribute to increase the understanding of NAFLD physiopathology. Our qPCR results, or other results deriving from microarray analysis or from similar studies, need to be validated in larger patient cohorts with a clearer stratification according to NAS (NAS ≤ 3 vs NAS ≥ 5). Serum RNAs should be analysed in vivo animal models in which genetic and environmental confounding factors should be minimized. Finally, intracellular RNAs should be assessed in liver biopsies of larger patient cohort. Notwithstanding these limitations, our data contribute to expand the knowledge on molecular mechanisms associated with NAFLD pathophysiology. Moreover, they could provide the evidence for identification of novel biomarkers as future possible alternatives to liver biopsy.

ACKNOWLEDGEMENTS

We wish to thank the Scientific Bureau of the University of Catania for language support.

CONFLICT OF INTEREST

The authors do not have any disclosures to report.

ORCID

Alessandra Scamporrino https://orcid.org/0000-0003-4483-3365
Salvatore Petta https://orcid.org/0000-0002-0822-9673
Luca Valenti https://orcid.org/0000-0001-8909-0345
Anna Fracanzani https://orcid.org/0000-0001-5918-0171

REFERENCES


SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Di Mauro S, Scamporrino A, Petta S, et al. Serum coding and non-coding RNAs as biomarkers of NAFLD and fibrosis severity. Liver Int. 2019;00:1–13. https://doi.org/10.1111/liv.14167