

# SCIENTIFIC REPORTS

OPEN

## MBOAT7 rs641738 variant and hepatocellular carcinoma in non-cirrhotic individuals

Benedetta Donati<sup>1</sup>, Paola Dongiovanni<sup>2</sup>, Stefano Romeo<sup>3,4</sup>, Marica Meroni<sup>1</sup>, Misti McCain<sup>5,6</sup>, Luca Miele<sup>7</sup>, Salvatore Petta<sup>8</sup>, Silvia Maier<sup>9</sup>, Chiara Rosso<sup>10</sup>, Laura De Luca<sup>9</sup>, Ester Vanni<sup>10</sup>, Stefania Grimaudo<sup>8</sup>, Renato Romagnoli<sup>11</sup>, Fabio Colli<sup>11</sup>, Flaminia Ferri<sup>12</sup>, Rosellina Margherita Mancina<sup>3</sup>, Paula Iruzubieta<sup>13</sup>, Antonio Craxi<sup>8</sup>, Anna Ludovica Fracanzani<sup>1,2</sup>, Antonio Grieco<sup>7</sup>, Stefano Ginanni Corradini<sup>12</sup>, Alessio Aghemo<sup>14</sup>, Massimo Colombo<sup>1,15</sup>, Giorgio Soardo<sup>9</sup>, Elisabetta Bugianesi<sup>10</sup>, Helen Reeves<sup>5,6</sup>, Quentin M. Anstee<sup>6,13</sup>, Silvia Fargion<sup>1,2</sup> & Luca Valenti<sup>1,2</sup>

Nonalcoholic fatty liver disease (NAFLD) represents an emerging cause of hepatocellular carcinoma (HCC), especially in non-cirrhotic individuals. The rs641738 C > T *MBOAT7/TMC4* variant predisposes to progressive NAFLD, but the impact on hepatic carcinogenesis is unknown. In Italian NAFLD patients, the rs641738 T allele was associated with NAFLD-HCC (OR 1.65, 1.08–2.55; n = 765), particularly in those without advanced fibrosis (p < 0.001). The risk T allele was linked to 3'-UTR variation in *MBOAT7* and to reduced *MBOAT7* expression in patients without severe fibrosis. The number of *PNPLA3*, *TM6SF2*, and *MBOAT7* risk variants was associated with NAFLD-HCC independently of clinical factors (p < 0.001), but did not significantly improve their predictive accuracy. When combining data from an independent UK NAFLD cohort, in the overall cohort of non-cirrhotic patients (n = 913, 41 with HCC) the T allele remained associated with HCC (OR 2.10, 1.33–3.31). Finally, in a combined cohort of non-cirrhotic patients with chronic hepatitis C or alcoholic liver disease (n = 1121), the T allele was independently associated with HCC risk (OR 1.93, 1.07–3.58). In conclusion, the *MBOAT7* rs641738 T allele is associated with reduced *MBOAT7* expression and may predispose to HCC in patients without cirrhosis, suggesting it should be evaluated in future prospective studies aimed at stratifying NAFLD-HCC risk.

Following the epidemics of obesity and insulin resistance and the recent improvements in the prevention and treatment of viral hepatitis, nonalcoholic fatty liver disease (NAFLD) is becoming a major cause of hepatocellular carcinoma (HCC) in Western countries<sup>1–4</sup>. NAFLD affects 16–38% of the general population worldwide<sup>5</sup> and is leading cause of cirrhosis, the main risk factor for HCC. However, HCC frequently develops in NAFLD

<sup>1</sup>Department of Pathophysiology and Transplantation, Università degli Studi di Milano, 20122, Milano, Italy.

<sup>2</sup>Internal Medicine, Fondazione IRCCS Ca' Granda Ospedale Policlinico Milano, Milano, Italy. <sup>3</sup>Sahlgrenska Center for Cardiovascular and Metabolic Research, Wallenberg Laboratory, Department of Molecular and Clinical Medicine, Department of Cardiology, University of Gothenburg, Gothenburg, Sweden. <sup>4</sup>Clinical Nutrition Unit, Department of Medical and Surgical Sciences, University Magna Graecia, Catanzaro, Italy. <sup>5</sup>Northern Institute of Cancer Research, The Medical School, Newcastle University, Newcastle-upon-Tyne, UK. <sup>6</sup>Hepatology, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. <sup>7</sup>Internal Medicine, Policlinico Gemelli, Roma, Italy. <sup>8</sup>Department of Gastroenterology, Università di Palermo, Palermo, Italy. <sup>9</sup>Internal Medicine and Liver Unit, Department of Experimental and Clinical Medical Sciences, University of Udine, Udine, Italy. <sup>10</sup>Gastroenterology, Department of Medical Sciences, University of Turin, Turin, Italy. <sup>11</sup>General Surgery and Liver Transplantation Center, Department of Surgical Sciences, University of Turin, Turin, Italy. <sup>12</sup>Department of Clinical Medicine, Gastroenterology division, Sapienza University, Rome, Italy. <sup>13</sup>Liver Research Group, Institute of Cellular Medicine, The Medical School, Newcastle University, Newcastle-upon-Tyne, UK. <sup>14</sup>Gastroenterology and Hepatology, Fondazione IRCCS Ca' Granda Ospedale Policlinico Milano, Milano, Italy. <sup>15</sup>Hepatology, Humanitas Clinical and Research Center, Rozzano, Italy. Benedetta Donati and Paola Dongiovanni contributed equally to this work. Correspondence and requests for materials should be addressed to L.V. (email: [luca.valenti@unimi.it](mailto:luca.valenti@unimi.it))

	Hepatocellular carcinoma		p
	Yes (n = 132)	No (n = 633)	
Age, years	67.0 ± 8.8	47.5 ± 12.3	<0.001
Sex, Female	25 (19)	163 (26)	0.25
Obesity, yes	39 (32)	202 (32)	0.99
T2DM, yes	84 (64)	124 (20)	<0.001
Severe fibrosis, F3-4	111 (84)	99 (16)	<0.001
<i>PNPLA3</i> , I148M			<0.001
I/I	31 (23)	231 (36)	
I/M	55 (42)	283 (45)	
M/M	46 (35)	120 (19)	
<i>TM6SF2</i> , E167K			0.30
E/E	109 (83)	538 (85)	
E/K	19 (14)	88 (14)	
K/K	4 (3)	7 (1)	
<i>MBOAT7</i> , rs641738 C > T			0.003
C/C	26 (20)	213 (34)	
C/T	69 (52)	285 (45)	
T/T	37 (28)	135 (21)	

**Table 1.** Clinical features of 765 Italian NAFLD patients stratified by HCC diagnosis. ( ): % values; T2DM: type 2 diabetes mellitus. Comparisons were performed by logistic regression setting HCC as dependent variable, and the association of genetic variants was analyzed assuming additive models.

patients even in the absence of severe liver fibrosis<sup>6</sup>. The high prevalence of NAFLD and development of HCC in non-cirrhotic patients unaware of being at risk renders therefore classic HCC screening strategies aimed at early diagnosis and curative treatment<sup>7</sup> unfeasible. Thus, there is an urgent need of non-invasive biomarkers able to stratify NAFLD-HCC risk, especially in patients without severe fibrosis<sup>8,9</sup>.

Inherited factors contribute to HCC susceptibility, and strong familial aggregation is observed<sup>10</sup>. NAFLD too has a strong heritable component<sup>11</sup>, and the *PNPLA3* I148M variant is the main common genetic determinant of hepatic fat content and of progressive NAFLD<sup>12–15</sup>. The mechanism is related to accumulation of the mutated protein<sup>16</sup>, which interferes with lipid droplets remodeling in hepatocytes<sup>15,17,18</sup>, and with retinol release by hepatic stellate cells<sup>19,20</sup>. The *PNPLA3* variant predicts HCC development in European patients with NAFLD<sup>21</sup>, suggesting that genetic risk factors may prove helpful to select high-risk individuals for screening<sup>21–23</sup>, but has a low specificity to be used as single prognostic biomarker<sup>24</sup>. Furthermore, the *PNPLA3* variant also predisposes to HCC in other liver diseases associated with steatosis, namely alcoholic liver disease (ALD) and chronic hepatitis C (CHC)<sup>23</sup>. The *TM6SF2* E167K variant also predisposes to progressive NAFLD by altering the secretion of very low-density lipoproteins<sup>25–27</sup>, but its direct role in HCC predisposition is disputed<sup>26,28</sup>.

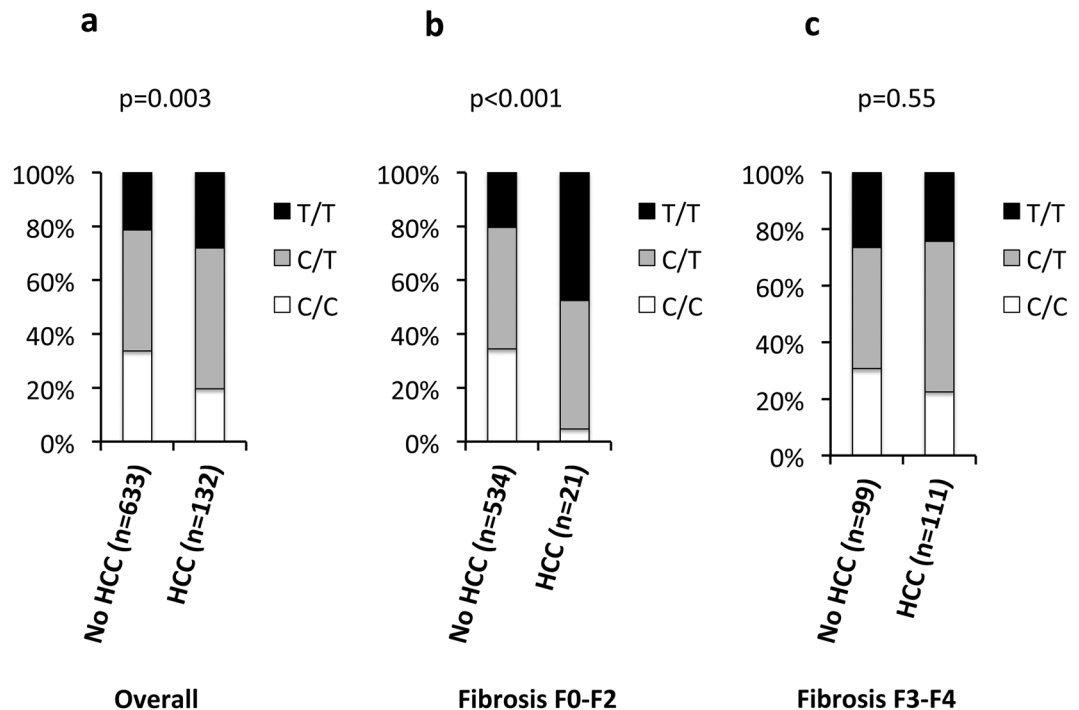
The *MBOAT7/TMC4* locus rs641738 C > T sequence variant predisposes to liver fibrosis development in individuals with excessive alcohol intake<sup>29</sup> and chronic hepatitis C<sup>30</sup>, and to the development and the progression of NAFLD in individuals of European descent<sup>31,32</sup>. However, whether the rs641738 variant is also associated with HCC risk is still unknown. Aim of this study was therefore to evaluate whether the rs641738 variant predisposes to HCC in NAFLD patients stratified by the presence of severe fibrosis, and in other liver diseases characterized by hepatic steatosis.

## Results

**The Italian NAFLD cohort.** The clinical features of Italian NAFLD patients stratified by HCC diagnosis are presented in Table 1. Patients who developed HCC were older ( $p < 0.001$ ), had higher prevalence of type 2 diabetes (T2DM;  $p < 0.001$ ), and of severe fibrosis (stage F3-F4) than those who did not (Table 1), whereas sex distribution and prevalence of obesity were not different ( $p = \text{NS}$ ). Concerning genetic risk factors, at univariate analysis HCC development was associated with *PNPLA3* and *MBOAT7* variants ( $p < 0.001$  and  $p = 0.003$ , respectively; Table 1).

The clinical features of Italian NAFLD-HCC patients according to presence of severe fibrosis are reported in Table S1. Patients who developed HCC in the absence of severe fibrosis ( $n = 21$ , 17%) were more frequently males ( $p = 0.040$ ), and carriers of the *MBOAT7* rs641738 risk T allele ( $p = 0.006$ ).

***MBOAT7* variation is associated with NAFLD-HCC in Italian patients.** The frequency distribution of the *MBOAT7* rs641738 C > T polymorphism in Italian NAFLD patients stratified by the presence of HCC is shown in Fig. 1A. There was a significant over-representation of the rs641738 T allele in HCC vs. non-HCC NAFLD patients ( $p = 0.003$ , Fig. 1a and Table 1). In the NAFLD population, which was selected due to referral for suspected steatohepatitis/HCC, there was a borderline deviation from Hardy-Weinberg equilibrium for the frequency distribution of the rs641738 *MBOAT7* variant ( $p = 0.03$  in both groups). However, the frequency



**Figure 1.** Frequency distribution of the *MBOAT7* locus rs641738 T allele in 765 Italian NAFLD patients stratified by the presence of hepatocellular carcinoma (HCC). (a) Overall cohort; (b) patients with stage F0-F2 fibrosis; (c) patients with stage F3-F4 fibrosis. Comparisons were performed by logistic regression setting HCC as dependent variable, and the association with the *MBOAT7* variant was analyzed assuming an additive model.

distribution of the rs641738 variant did not violate Hardy-Weinberg equilibrium in 243 unselected healthy control subjects ( $p = \text{NS}$ ; Table S3).

The clinical features of Italian NAFLD patients stratified by the rs641738 genotype are shown in Table S2. The rs641738 T allele was borderline associated with T2DM ( $p = 0.05$ ) in the overall cohort, but this was not confirmed in patients stratified by HCC diagnosis, and was possibly explained by the confounding effect of the association between both rs641738 T allele and T2DM with HCC. In HCC patients, the T allele was associated with obesity ( $p = 0.035$ ), and lack of severe fibrosis F3-F4 ( $p = 0.006$ ).

As expected, the rs641738 T allele was nearly associated with severe fibrosis (stage F3-F4; OR 1.24, 95% c.i. 1.00–1.54;  $p = 0.052$ ). The frequency distribution of *PNPLA3*, *TM6SF2*, and *MBOAT7* variants according to hepatocellular carcinoma (HCC) diagnosis in Italian NAFLD patients stratified by the severity of fibrosis (stage F0-F2 vs. F3-F4) is presented in Table 2. While the *PNPLA3* variant was associated with HCC development in patients with ( $p = 0.011$ ), but not in those without severe fibrosis ( $p = \text{NS}$ ), the *MBOAT7* T allele was associated with HCC in patients without ( $p < 0.001$ ; Fig. 1b and Table 2), but not in those with ( $p = 0.55$ ; Fig. 1c and Table 2) severe fibrosis.

**Independent predictors of NAFLD-HCC.** The independent predictors of NAFLD-HCC are presented in Table 3. At univariate analysis (left panel), development of HCC was associated with older age, T2DM, and severe fibrosis ( $p < 0.001$  for all), whereas among the genetic factors with *PNPLA3* I148M ( $p < 0.001$ ) and *MBOAT7* T rs641738 T alleles (OR 2.18, 95% c.i. 1.30–3.63;  $p = 0.003$ ).

At multivariate logistic regression analysis including as independent variables noninvasive predictors of HCC (Model 1, middle panel), which can be applied even in NAFLD patients without histological evaluation of liver damage, HCC development was associated with older age ( $p < 0.001$ ), male sex ( $p = 0.045$ ), T2DM ( $p < 0.001$ ), *PNPLA3* I148M alleles ( $p = 0.010$ ), *TM6SF2* E167K alleles ( $p = 0.027$ ), and remained strongly associated with *MBOAT7* rs641738 alleles (OR 1.81, 95% c.i. 1.24–2.69;  $p = 0.002$ ).

After further adjustment for the presence of severe fibrosis stage F3-F4 (Model 2), the *TM6SF2* E167K ( $p = 0.008$ ) and *MBOAT7* rs641738 T (OR per allele 1.65, 95% c.i. 1.08–2.55;  $p = 0.021$ ; OR for T/T vs. C/C 2.73, 95% c.i. 1.17–6.51,  $p = 0.008$ ) alleles remained significantly associated with HCC risk, whereas the effect of the *PNPLA3* I148M variant was lost.

In Model 1, the effect of *MBOAT7* variant was larger in patients without severe fibrosis (OR per allele 2.78, 95% c.i. 1.04–8.71;  $p = 0.050$ ), whereas it was not significant considering only patients with severe fibrosis (OR per allele 1.19, 95% c.i. 0.78–2.03;  $p = 0.3$ ).

**Combined effect of genetic risk factors for NAFLD-HCC.** The relationship between the total number of risk alleles including *PNPLA3* I148M, *TM6SF2* E167K, and *MBOAT7* rs641738 T and HCC risk is presented in Fig. 2. There was a significant association between the number of risk alleles and HCC (OR per allele 1.56, 95% c.i.

Fibrosis stage						
HCC	F0-F2		p	F3-F4		p
	Yes	No		Yes	No	
	Yes (n = 21)	No (n = 534)		Yes (n = 111)	No (n = 99)	
PNPLA3, I148M			0.92			0.011
I/I	10 (48)	200 (37)		21 (19)	31 (31)	
I/M	6 (28)	236 (44)		49 (44)	47 (48)	
M/M	5 (24)	98 (19)		41 (37)	22 (22)	
TM6SF2, E167K			0.11			0.50
E/E	16 (76)	458 (86)		93 (84)	80 (80)	
E/K	4 (19)	70 (13)		15 (13)	18 (19)	
K/K	1 (5)	6 (1)		3 (3)	1 (1)	
MBOAT7, rs641738			<0.001			0.55
C/C	1 (4)	183 (35)		25 (22)	30 (30)	
C/T	10 (48)	243 (45)		59 (53)	42 (43)	
T/T	10 (48)	108 (20)		27 (25)	27 (27)	

**Table 2.** Frequency distribution of *PNPLA3*, *TM6SF2*, and *MBOAT7* variants according to hepatocellular carcinoma (HCC) in 765 Italian NAFLD patients stratified by the severity of fibrosis (stage F0-F2 vs. F3-F4). ( ): % values. HCC. Hepatocellular carcinoma. Comparisons were performed by logistic regression setting HCC as dependent variable, and the association of genetic variants was analyzed assuming additive models.

	Unadjusted			Model 1			Model 2		
	OR	95% c.i.	p	OR	95% c.i.	p	OR	95% c.i.	p
Age, years	1.20	1.16–1.24	<0.001	1.19	1.15–1.23	<0.001	1.16	1.11–1.21	<0.001
Sex, Female	0.57	0.41–1.06	0.09	0.52	0.27–0.99	0.045	0.45	0.21–0.91	0.026
Obesity, yes	1.11	0.75–1.69	0.61	1.69	1.05–2.56	0.069	2.50	1.32–4.76	0.008
T2DM, yes	7.18	4.81–10.84	<0.001	4.73	2.75–8.30	<0.001	3.33	1.75–6.44	<0.001
Severe fibrosis, F3-F4	28.9	17.6–49.5	<0.001	NA	NA	NA	12.5	6.36–6.1	<0.001
<i>PNPLA3</i> , n I148M alleles	1.70	1.32–2.21	<0.001	1.61	1.12–2.32	0.010	1.31	0.86–2.03	0.24
<i>TM6SF2</i> , n E167K alleles	1.27	0.82–1.92	0.27	1.99	1.08–3.65	0.027	2.80	1.33–6.10	0.008
<i>MBOAT7</i> , n T alleles	2.18	1.30–3.63	0.003	1.81	1.24–2.69	0.002	1.65	1.08–2.55	0.021

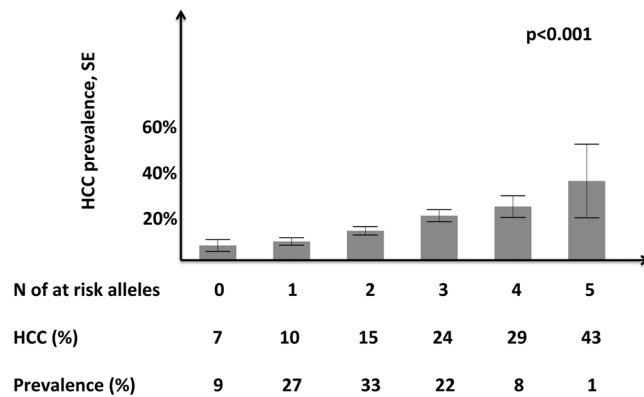
**Table 3.** Independent predictors of HCC in 765 Italian patients with NAFLD. OR: odds ratio of HCC, 95% c.i.: 95% confidence interval; T2DM: type 2 diabetes mellitus; n: number of at risk alleles. Comparisons were performed by logistic regression setting HCC as dependent variable, and the association of genetic variants was analyzed assuming additive models. Model 1: adjusted for age, sex, obesity, and T2DM; Model 2: further adjusted for presence of advanced fibrosis; NA: not addressed.

1.31–1.86; OR 9.25, 95% c.i. 3.83–22.8 between the extremes of the distribution, i.e. 5 vs. 0 risk alleles;  $p < 0.001$  for both). HCC risk was 9% in the 36% of the population with 0–1 risk alleles, 19% in the 55% of the population with 2–3 risk alleles, and 31% in the 9% of the population with 4–5 risk alleles. The association held constant after correction for other risk factors as in Model 2 (OR per allele 1.68, 95% c.i. 1.30–2.20; OR 13.4, 95% c.i. 3.71–51.5 between the extremes of the distribution;  $p < 0.001$  for both).

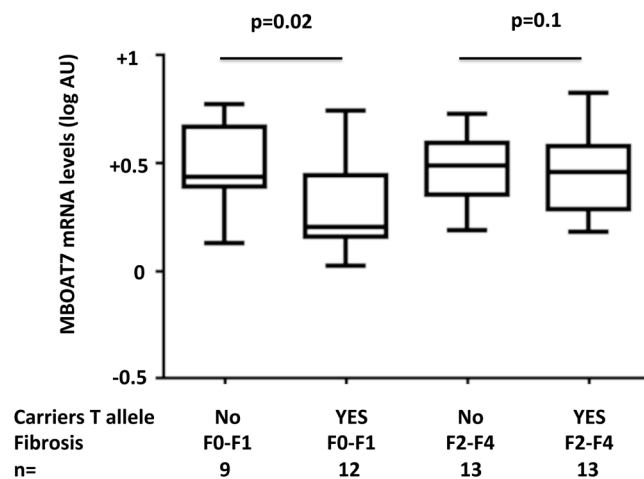
A combined risk score considering acquired and genetic risk factors was developed to predict HCC:  $1/(1 + e^{-((-12.588 + (0.162 * \text{age}) + (0.404 * \text{Sex}: 1 \text{ if male, } -1 \text{ if female}) + (0.259 * \text{Obesity}: 1 \text{ present, } -1 \text{ absent}) + (0.587 * \text{T2DM}: 1 \text{ present, } -1 \text{ absent}) + (1.299 * \text{Severe Fibrosis}: 1 \text{ yes, } -1 \text{ no}) + (0.442 * \text{number of risk alleles}))})$ ). The model had a  $0.96 \pm 0.4$  area under the receiving operating characteristic curve (AUROC) for detecting HCC cases. The optimal cutoff (identifying the best combination of sensitivity and specificity) had 96% sensitivity and 89% specificity for HCC (Fig. S1). The corresponding AUROC of a model taking into consideration only clinical factors was slightly lower ( $0.93 \pm 0.5$ ). In the subgroup of patients without severe fibrosis, the AUROC for clinical factors alone was  $0.91 \pm 0.5$ , whereas the full model incorporating genetic risk factors maintained an AUROC of  $0.96 \pm 0.4$  ( $p = \text{NS}$  vs. clinical factors alone).

**Relationship between *MBOAT7* locus variants and gene expression.** To investigate the biological basis for the stronger association of *MBOAT7* locus variation with HCC in patients without severe fibrosis, we next examined the association of the rs641738 variant with possibly variants that may influence *MBOAT7* mRNA stability, and with hepatic *MBOAT7* mRNA expression levels in patients stratified by the severity of liver fibrosis.

In 98 severely obese patients, the rs641738 variant was in high linkage with the *MBOAT7* 3'-UTR variant rs8736 C > T polymorphism ( $R^2 = 0.98$ ; only 1/98 discordant case). Despite this, the rs8736 polymorphism was non-significantly more closely associated with NAFLD ( $p = 0.048$  vs.  $p = 0.057$ ) and *MBOAT7* expression



**Figure 2.** Risk of hepatocellular carcinoma according to the number of *PNPLA3* I148M, *TM6SF2* E167K, and *MBOAT7* rs641738 C > T risk variants in 765 Italian patients with NAFLD. HCC: hepatocellular carcinoma; SE: standard error. Comparisons were performed by a multivariate logistic regression setting HCC as dependent variable, and the association with genetic risk factors (numbers of at risk alleles carried) was analyzed assuming an additive model.  $p < 0.001$  for the association of the number of risk alleles with HCC, both at unadjusted analysis and after adjustment for age, sex, obesity, T2DM, and presence of advanced fibrosis stage F2-F4.



**Figure 3.** Impact of the presence of rs643718 risk T allele on *MBOAT7* expression (log mRNA levels) in 47 patients with NAFLD from the Milan Hepatology service stratified by the presence of clinically significant hepatic fibrosis (stage F2-F4). Data were compared by Student's t-test.

( $p = 0.042$  vs.  $p = 0.046$ ) than rs641738. These data are in line with the hypothesis that rs641738 is not the causal variant, but may be in linkage with variants influencing *MBOAT7* expression.

Gene expression of *MBOAT7* in 47 patients from the Hepatology service characterized by more severe liver damage (Table S4) is shown in Fig. 3. The rs641738 T allele was associated with reduced hepatic *MBOAT7* expression in patients absent or mild fibrosis (stage F0-F1;  $p = 0.02$ ), but not in those with moderate-severe fibrosis (stage F2-F4;  $p = 0.1$ ).

***MBOAT7* variation and NAFLD-HCC risk in UK non-cirrhotic NAFLD patients.** To increase the study power, we next evaluated the association of the rs641738 T allele with HCC risk in non-cirrhotic NAFLD patients from UK, whose clinical features are presented in Table S5. The frequency distribution of the rs641738 C > T genotype in patients stratified by HCC diagnosis is shown in Table 4. Although in the UK non-cirrhotic NAFLD cohort ( $N = 358$ , of whom 20 with HCC) *MBOAT7* variation was not significantly associated with HCC ( $p = 0.32$ ), in the overall combined UK/Italian cohort of NAFLD patients without advanced fibrosis/cirrhosis the T allele remained associated with an increased risk of HCC (allelic OR 2.10, 95% c.i. 1.33–3.31).

**Impact of *MBOAT7* variation on HCC risk in non-cirrhotic patients with other liver diseases.** We finally evaluated the impact of the rs641738 T allele on HCC risk in 1121 non-cirrhotic patients with CHC and ALD (25, 2% with HCC, Table S6). Results are presented in Table 5. The rs641738 T allele was associated with increased risk of HCC, independently of age, sex, and the etiology of liver disease (OR 1.93 for each T allele, 95% c.i. 1.07–3.58;  $p = 0.035$ ), with an effect size comparable to that observed in NAFLD. We observed a similar

Hepatocellular carcinoma				
	YES	NO	p*	OR (95% c.i.)
Italy, n=	21	534		
<i>MBOAT7</i> , rs641738 C > T			<0.001	3.09 (1.58–6.02)
C/C	1 (4)	183 (35)		
C/T	10 (48)	243 (45)		
T/T	10 (48)	108 (20)		
UK, n=	20	338		
<i>MBOAT7</i> , rs641738 C > T			0.32	1.39 (0.71–2.71)
C/C	4 (20)	97 (29)		
C/T	11 (55)	179 (53)		
T/T	5 (25)	62 (18)		
OVERALL, n=	41	872		
<i>MBOAT7</i> , rs641738 C > T			0.001	2.10 (1.33–3.31)
C/C	5 (12)	280 (32)		
C/T	21 (51)	422 (49)		
T/T	15 (37)	170 (19)		

**Table 4.** Frequency distribution of the rs641738 C > T genotype in 913 patients without advanced liver fibrosis/cirrhosis from the Italian and UK cohorts stratified by HCC diagnosis. \*Comparisons were performed by logistic regression setting HCC as dependent variable, and the association of genetic variants was analyzed assuming additive models. ( ): % values; OR, 95% c.i.: allelic odds ratio and 95% confidence intervals.

HCC					
	No (n = 1096)	Yes (n = 25)	p value	Allelic OR, 95% c.i.*	p value*
<i>PNPLA3</i> I148M	558/437/101 (51/40/9)	8/10/7 (32/40/28)	0.016	1.92, 1.07–3.45	0.021
<i>TM6SF2</i> E167K	973/122/1 (89/11/0)	20/5 (80/20)	0.16	1.96, 0.61–5.27	0.16
<i>MBOAT7/TMC4</i> rs641738 C > T	327/510/259 (30/46/24)	2/15/8 (8/60/32)	0.028	1.93, 1.07–3.58	0.035

**Table 5.** Independent predictors of hepatocellular carcinoma (HCC) in 1121 non-cirrhotic patients with chronic liver diseases associated with hepatic fat accumulation (597 with chronic hepatitis C and 524 with alcoholic liver disease). ( ): % values; OR: odds ratio; c.i.: confidence interval. Comparisons were performed by logistic regression setting HCC as dependent variable, and the association of genetic variants was analyzed assuming additive models. \* Adjusted for age, sex, liver disease etiology, and *PNPLA3*, *TM6SF2* and *MBOAT7/TMC4* genetic variants.

trend for association of the T allele with non-cirrhotic HCC in patients with CHC and ALD analyzed separately (Table S7). The *PNPLA3* I148M variant was also associated with HCC development outside cirrhosis, with a similar effect size of that of *MBOAT7* variation (Table 5; p = 0.021).

## Discussion

In this study, we evaluated whether the rs641738 C > T *MBOAT7* locus sequence variant, associated with the development and progression of NAFLD<sup>31,32</sup>, influences susceptibility to NAFLD-HCC. The main result is that in the Italian NAFLD cohort each *MBOAT7* rs641738 T allele conferred an approximately 80% increased risk of HCC. The association was mostly driven by a strong enrichment in the risk T allele in patients without advanced liver fibrosis, suggesting that *MBOAT7* variation predisposes to HCC development particularly in non-cirrhotic patients. To confirm this hypothesis, we also examined the frequency of the *MBOAT7* T allele in an independent UK cohort of non-cirrhotic NAFLD patients. Although the T allele was not significantly associated with NAFLD-HCC in this replication cohort, it remained significant in the combined cohort of 913 non-cirrhotic European NAFLD patients (41 with HCC) where it was associated with a greater than 2-fold increased risk of NAFLD-HCC; a relatively large effect size for a common genetic variant. This genetic polymorphism might thus represent a first useful biomarker to stratify HCC risk among individuals affected by NAFLD without advanced liver fibrosis.

The *MBOAT7* protein catalyzes the transfer of polyunsaturated fatty acids such as arachidonoyl-CoA to lyso-phosphatidylinositol, thereby allowing to achieve an adequate level of desaturation in cell membranes<sup>31</sup>. The rs641738 T allele is associated with reduced *MBOAT7* expression and altered phosphatidyl-inositol plasma and hepatic composition<sup>31,32</sup>, favoring hepatocellular fat accumulation and the production of inflammatory mediators<sup>31</sup>. However, rs641738 is not likely the causal variant underpinning susceptibility to NAFLD and HCC, as we observed that it is in strong linkage with other polymorphisms in 3'-UTR of *MBOAT7*, which may be more closely related to the phenotype and are potentially involved in the regulation of *MBOAT7* mRNA stability.



It could be speculated that the effect size of rs641738 on HCC risk was larger in patients without severe fibrosis because the presence of the risk variant may somewhat compensate for the lack of the cirrhotic pro-carcinogenic environment. However, we observed that the rs641738 T allele is associated with reduced hepatic expression of *MBOAT7*<sup>31</sup> only in NAFLD patients without severe fibrosis. Therefore, the *MBOAT7* variant may exert its deleterious effect specifically at early stages of liver disease. Alteration of hepatic parenchymal structure and relative cell-types representation may then hamper the impact of the *MBOAT7* variant during severe fibrosis, because *MBOAT7* is highly expressed in hepatic stellate cells and inflammatory cells<sup>31,33</sup>. In keeping with this interpretation, the rs641738 T allele was also associated with development of early stages, but not severe fibrosis in patients at risk of NASH<sup>31</sup>, and in a large cohort of CHC patients<sup>30</sup>. Therefore, the *MBOAT7* variation might have a dual impact on liver disease during initial stages: either predisposes to HCC development before severe fibrosis ensues, or it facilitates the evolution to early-intermediate fibrosis. In line with this hypothesis, we also showed that the rs641738 T allele was associated with HCC development in non-cirrhotic patients with ALD or CHC.

In the Italian NAFLD cohort, the overall impact of the *MBOAT7* rs641738 on HCC risk was similar to that of the I148M *PNPLA3* variant. However, the effect of the I148M variant on HCC risk was not independent of severe fibrosis, suggesting that the mechanism is partly mediated by promotion of hepatic fibrogenesis and alteration of hepatic stellate cells biology<sup>15,19,34</sup>. Notably, the size effect of the *PNPLA3* I148M variant was larger and only partially attenuated by the impact on liver fibrosis in a previous study conducted in a UK cohort<sup>21</sup>. This difference may be due to lifestyle factors, and to the higher prevalence of clinical cofactors, as opposed to genetic risk variants (lower frequency of the I148M variant) in the UK cohort. In addition, we also report for the first time an association between the *TM6SF2* E167K variant and NAFLD-HCC. However, this association was not detected by univariate analysis due to an interaction of the *TM6SF2* variant with clinical factors, and it has previously been absent when sought in the UK NAFLD cohort<sup>26</sup>, whereas a predisposing effect on HCC was reported in a Italian cohort of patients with alcoholic cirrhosis<sup>28</sup>. Further studies are therefore required to confirm whether the E167K variant is an independent risk factor for HCC.

All in all, data suggest that genetic variants predisposing to hepatic fat accumulation promote hepatic carcinogenesis. Indeed, hepatocellular fat accumulation represents a key feature of hepatic carcinogenesis<sup>35,36</sup>. Therefore, they might represent useful biomarkers for risk stratification. In fact, in the Italian NAFLD cohort the number of genetic risk variants carried was strongly associated with HCC, with 13.4-fold higher risk in those carrying five risk variants as compared to none. Remarkably, the number of genetic risk variants was able to classify NAFLD patients in three groups with different HCC risk: 9% in the 36% of patients with 0–1 risk alleles, 19% in the 55% with 2–3 risk alleles, and 31% in those carrying more than 3 risk alleles. This could in principle allow a better stratification of HCC risk than those allowed by the *PNPLA3* I148M variant alone, which was proposed by the EASL-EASD-EASO NAFLD guidelines<sup>21,37</sup>. However, in the present cross-sectional Italian cohort genetic risk variants did not significantly improve the predictive accuracy of clinical factors.

Limitations of the study include its cross-sectional retrospective nature, resulting similarly to previous reports<sup>21</sup> in an uneven representation of clinical risk factors (age, sex, T2DM, severe fibrosis) between HCC cases and controls. However, the majority of NAFLD-HCC patients are still diagnosed incidentally outside regular follow-up<sup>38</sup>, so that prospective studies in patients with advanced disease would not be more informative, especially for the risk of non-cirrhotic NAFLD-HCC. This could have led to an underestimation of the impact of inherited genetic risk variants on NAFLD-HCC, whereas the impact of clinical risk factors may have been overestimated. Therefore, the weight of specific factors in determining the HCC risk score should be reassessed in larger prospective cohorts with long follow-up and availability of the genetic risk profile before evaluation of genetic risk variants can be considered for implementation in clinical practice. Due to the relatively low number of patients included, these limitations are particularly relevant for the association of *MBOAT7* variation with HCC development in patients without severe liver fibrosis. Finally, these results may not be applicable to other ethnic groups.

In conclusion, the *MBOAT7* rs641738 T allele is associated with reduced *MBOAT7* expression and may predispose to HCC in European individuals without cirrhosis, suggesting it should be evaluated in future prospective studies aimed at stratifying NAFLD-HCC risk.

## Methods

**Patients.** We enrolled 132 consecutive unrelated patients with NAFLD-HCC of Italian descent, referred between January 2008 and January 2015 to the Milan, Udine, Turin, Rome, and Palermo hospitals, for whom DNA samples were available. Diagnosis of HCC was based on the EASL-EORTC Clinical Practice Guidelines<sup>7</sup>. In the absence of liver biopsy, diagnosis of NAFLD required detection of ultrasonographic steatosis plus at least one criterion of the metabolic syndrome.

As controls, we selected Italian-ancestry patients with histologically confirmed NAFLD followed at the same referral outpatient Hepatology services during the same study period<sup>27,39</sup>, from a recently published database<sup>31</sup>, who did not develop HCC during follow-up. We did not consider for this study bariatric patients from Milan because of the different recruitment criteria (indication to metabolic surgery), leading to different demographic and clinical features<sup>27,39</sup>, and the lack of incident HCC cases in the cohort.

All patients were tested for secondary causes of steatosis including alcohol abuse ( $\geq 30/20$  g/day in M/F) and the use of drugs known to precipitate steatosis. Viral and autoimmune hepatitis, hereditary hemochromatosis, Wilson's disease, alpha-1-antitrypsin deficiency and present or previous active infection with HBV and HCV were ruled out using standard clinical and laboratory evaluation, as well as liver biopsy features.

Advanced fibrosis was defined in the presence of fibrosis stage F3-F4<sup>40</sup>, when liver biopsy was available. In HCC patients with radiological diagnosis, advanced fibrosis was defined in the presence of clinical, endoscopic or ultrasonographic signs of portal hypertension or cirrhosis ( $n = 46$ ), or of liver stiffness  $\geq 8.4$  kPa evaluated by elastometry ( $n = 3$ ), or by a positive NAFLD fibrosis score ( $n = 5$ )<sup>41</sup>. Obesity was defined when BMI  $> 30$  Kg/m<sup>2</sup>. All NAFLD patients without HCC included in the study underwent liver biopsy, while among those who developed

HCC, fibrosis staging was histologically performed in 78 (59%) of the cases. Clinical features of subjects included according to the presence of HCC are shown in Table 1.

The UK NAFLD cohort comprising HCC cases and controls recruited at a single tertiary centre (the Newcastle upon Tyne Hospitals NHS Foundation Trust, UK) has previously been described<sup>21</sup>. The study had all the necessary ethical approvals and all participants gave informed consent. For this study, we considered patients for whom liver disease staging scored according to the NASH CRN histopathological system and DNA samples for *MBOAT7* genotyping were available. Clinical features of the 358 individuals analyzed in this study are presented in Table S5.

We next evaluated the impact of the rs641738 variant on HCC risk in an Italian multicenter cohort of non-cirrhotic patients with chronic hepatitis C (CHC; n = 597) and alcoholic liver disease (ALD; n = 524). CHC patients were from the well-described histological Milan CHC cohort<sup>42,43</sup>, whereas HCC cases were previously described by our group (Milan HCC cohort, where presence of cirrhosis was carefully assessed)<sup>44</sup>.

ALD patients with and without HCC included patients from the Milan center, which were partly described previously (again the Milan HCC cohort), and for whom liver disease was evaluated by histology or as described above for NAFLD<sup>44</sup>. We also considered consecutive individuals, who were admitted to the Outpatient Clinic at the Department of Clinical Medicine, Policlinico Umberto I, Rome for alcohol abuse or dependence between 2005 and 2014, for whom DNA samples were still available and genotyping was successful<sup>45</sup>. At-risk alcohol consumption was defined as  $\geq 3/2$  alcohol units per day for M/F, respectively. Cirrhosis was ruled out based on the presence of at least one of the following features: (i) current or past cirrhosis complications; (ii) the presence of at least two parameters among hyperbilirubinaemia, hypoalbuminaemia, prolonged prothrombin time, low platelet count, irregular liver surface at ultrasound/CT, reduced portal vein flow at ultrasound, gastroesophageal varices at endoscopy, or by histological analysis. Individuals with other coexistent liver diseases were excluded. Clinical features of this cohort are presented in Table S6.

The study protocol was conformed to the ethical guidelines of the 1975 Declaration of Helsinki, was approved by the Ethical Committee of the Fondazione IRCCS Ca' Granda of Milan, as well as by the other involved Institutions, and was performed according to the recommendations of the hospitals involved. Informed consent was obtained from each patient.

**Genotyping.** Patients were genotyped for rs738409 (*PNPLA3* I148M) rs58542926 (*TM6SF2* E167K), as previously described<sup>27</sup>. The rs641738 (located within *TMC4* coding sequence, chr19:54676763 positive strand, p.G17E protein variant) and rs8736 (located within the 3'-untranslated region – UTR – of *MBOAT7*) *MBOAT7/TMC4* locus genotyping has been performed in duplicate by TaqMan 5'-nuclease assays at the Metabolic Liver Disease lab, at the University of Milan (Life Technologies, Carlsbad, CA). Genotyping success rate was >98%. The duplicate genotype concordance rate was 100%. Genotyping was confirmed by direct sequencing of random samples, and concordance rate was 100% (primers are available upon demand).

**Gene expression analysis.** Expression of *MBOAT7* and *TMC4* was determined in two different subsets of patients. The first one was made of 98 severely obese patients, with a very low prevalence of advanced liver fibrosis, and has previously been described in details<sup>31</sup>. This was used to analyze the association of a 3'-untranslated region *MBOAT7* locus variant, rs8736 possibly influencing *MBOAT7* mRNA stability, with *MBOAT7* expression. The second one was made up of 47 patients from the Hepatology service, and was characterized by a higher prevalence of liver fibrosis. Clinical features of these patients are presented in Table S4. This was used to evaluate the impact of liver fibrosis on the association between the rs641738 variant and *MBOAT7* expression.

*MBOAT7* expression was quantified as previously described<sup>31</sup>. Association analysis between rs641738 variant (additive model) and gene expression, and linkage with rs8736 were conducted by the PLINK v1.07 genetic analysis software.

**Statistical analysis.** For descriptive statistics, continuous variables are shown as mean and standard deviation or median. Categorical variables are presented as number and proportion. The OR of HCC or severe fibrosis per *MBOAT7* rs641738 T alleles and other risk factors were estimated by logistic regression models, assuming an additive effect of genetic variants, and adjusted for clinical risk factors (the major known clinical risk factors for NAFLD-HCC in previous studies) as specified, *PNPLA3* I148M and *TM6SF2* E167K genotypes (clinical and genetic factors previously associated with or candidate for liver disease evolution to HCC)<sup>21,26</sup>. A NAFLD-HCC risk score was developed according to a previously described procedure<sup>41</sup>.

Statistical analyses were carried out using the JMP 13.0 Statistical analysis software (SAS Institute, Cary, NC), R statistical analysis software version 3.3.2 (<http://www.R-project.org/>), and PLINK v1.07<sup>46</sup>. P-values < 0.05 were considered statistically significant. The study methods and results have been reported according to the STROBE/STREGA guidelines for genetic association studies.

## References

- Dongiovanni, P., Romeo, S. & Valenti, L. Hepatocellular carcinoma in nonalcoholic fatty liver: role of environmental and genetic factors. *World J Gastroenterol* **20**, 12945–12955 (2014).
- Dyson, J. *et al.* Hepatocellular cancer: The impact of obesity, type 2 diabetes and a multidisciplinary team. *J Hepatol* **60**, 110–117 (2014).
- Baffy, G., Brunt, E. M. & Caldwell, S. H. Hepatocellular carcinoma in non-alcoholic fatty liver disease: An emerging menace. *J Hepatol* **56**, 1384–1391 (2012).
- Bugianesi, E. *et al.* Expanding the natural history of nonalcoholic steatohepatitis: from cryptogenic cirrhosis to hepatocellular carcinoma. *Gastroenterology* **123**, 134–140 (2002).



5. Younossi, Z. & Henry, L. Contribution of Alcoholic and Nonalcoholic Fatty Liver Disease to the Burden of Liver-Related Morbidity and Mortality. *Gastroenterology* **150**, 1778–1785 (2016).
6. Kawada, N. *et al.* Hepatocellular carcinoma arising from non-cirrhotic nonalcoholic steatohepatitis. *J Gastroenterol* **44**, 1190–1194 (2009).
7. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol* **56**, 908–943 (2012).
8. Paradis, V. *et al.* Hepatocellular carcinomas in patients with metabolic syndrome often develop without significant liver fibrosis: a pathological analysis. *Hepatology* **49**, 851–859 (2009).
9. Leung, C. *et al.* Characteristics of hepatocellular carcinoma in cirrhotic and non-cirrhotic non-alcoholic fatty liver disease. *World J Gastroenterol* **21**, 1189–1196 (2015).
10. Turati, F. *et al.* Family history of liver cancer and hepatocellular carcinoma. *Hepatology* **55**, 1416–1425 (2012).
11. Dongiovanni, P. & Valenti, L. Genetics of nonalcoholic fatty liver disease. *Metabolism* **65**, 1026–1037 (2016).
12. Romeo, S. *et al.* Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* **40**, 1461–1465 (2008).
13. Valenti, L. *et al.* Homozygosity for the PNPLA3/adiponutrin I148M polymorphism influences liver fibrosis in patients with nonalcoholic fatty liver disease. *Hepatology* **51**, 1209–1217 (2010).
14. Sookoian, S. & Pirola, C. J. Meta-analysis of the influence of I148M variant of patatin-like phospholipase domain containing 3 gene (PNPLA3) on the susceptibility and histological severity of nonalcoholic fatty liver disease. *Hepatology* **53**, 1883–1894 (2011).
15. Dongiovanni, P. *et al.* PNPLA3 I148M polymorphism and progressive liver disease. *World J Gastroenterol* **19**, 6969–6978 (2013).
16. Donati, B. *et al.* The rs2294918 E434K Variant Modulates Patatin-Like Phospholipase Domain-Containing 3 Expression and Liver Damage. *Hepatology* **63**, 787–798 (2016).
17. Ruhanen, H. *et al.* PNPLA3 mediates hepatocyte triacylglycerol remodeling. *J Lipid Res* **55**, 739–746 (2014).
18. Smagris, E. *et al.* Pnpla3I148M knockin mice accumulate PNPLA3 on lipid droplets and develop hepatic steatosis. *Hepatology* **61**, 108–118 (2015).
19. Pirazzi, C. *et al.* PNPLA3 has retinyl-palmitate lipase activity in human hepatic stellate cells. *Hum Mol Genet* **23**, 4077–4085 (2014).
20. Mondul, A. *et al.* PNPLA3 I148M Variant Influences Circulating Retinol in Adults with Nonalcoholic Fatty Liver Disease or Obesity. *J Nutr* **145**, 1687–1691 (2015).
21. Liu, Y. L. *et al.* Carriage of the PNPLA3 rs738409 C > G polymorphism confers an increased risk of non-alcoholic fatty liver disease associated hepatocellular carcinoma. *J Hepatol* **61**, 75–81 (2013).
22. Valenti, L., Dongiovanni, P., Ginanni Corradini, S., Burza, M. A. & Romeo, S. PNPLA3 I148M variant and hepatocellular carcinoma: a common genetic variant for a rare disease. *Dig Liver Dis* **45**, 619–624 (2013).
23. Trepo, E. *et al.* Association between the PNPLA3 (rs738409 C > G) variant and hepatocellular carcinoma: Evidence from a meta-analysis of individual participant data. *Hepatology* **59**, 2170–2177 (2014).
24. Anstee, Q. M., Liu, Y. L., Day, C. P. & Reeves, H. L. Reply to: HCC and liver disease risk in homozygous PNPLA3 p.I148M carriers approach monogenic inheritance. *J Hepatol* **62**, 982–983 (2015).
25. Kozlitina, J. *et al.* Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* **46**, 352–356 (2014).
26. Liu, Y. L. *et al.* TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. *Nature communications* **5**, 4309 (2014).
27. Dongiovanni, P. *et al.* Transmembrane 6 superfamily member 2 gene variant disentangles nonalcoholic steatohepatitis from cardiovascular disease. *Hepatology* **61**, 506–514 (2015).
28. Falletti, E., Cussigh, A., Cmet, S., Fabris, C. & Toniutto, P. PNPLA3 rs738409 and TM6SF2 rs58542926 variants increase the risk of hepatocellular carcinoma in alcoholic cirrhosis. *Dig Liver Dis* **48**, 69–75 (2016).
29. Buch, S. *et al.* A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. *Nat Genet* **47**, 1443–1448 (2015).
30. Thabet, K. *et al.* MBOAT7 rs641738 increases risk of liver inflammation and transition to fibrosis in chronic hepatitis C. *Nature communications* **7**, 12757 (2016).
31. Mancina, R. M. *et al.* The MBOAT7-TMC4 Variant rs641738 Increases Risk of Nonalcoholic Fatty Liver Disease in Individuals of European Descent. *Gastroenterology* **150**, 1219–1230 e1216 (2016).
32. Luukkonen, P. K. *et al.* The MBOAT7 variant rs641738 alters hepatic phosphatidylinositols and increases severity of non-alcoholic fatty liver disease in humans. *J Hepatol* (2016).
33. Gijon, M. A., Riekhof, W. R., Zarini, S., Murphy, R. C. & Voelker, D. R. Lysophospholipid acyltransferases and arachidonate recycling in human neutrophils. *J Biol Chem* **283**, 30235–30245 (2008).
34. Pingitore, P. *et al.* PNPLA3 overexpression results in reduction of proteins predisposing to fibrosis. *Hum Mol Genet* (2016).
35. Bjornson, E. *et al.* Stratification of Hepatocellular Carcinoma Patients Based on Acetate Utilization. *Cell reports* **13**, 2014–2026 (2015).
36. Li, L. *et al.* Differential requirement for de novo lipogenesis in cholangiocarcinoma and hepatocellular carcinoma of mice and humans. *Hepatology* **63**, 1900–1913 (2016).
37. European Association for the Study of the Liver. Electronic address eee, European Association for the Study of D, European Association for the Study of O. EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol* **64**, 1388–1402 (2016).
38. Piscaglia, F. *et al.* Clinical patterns of hepatocellular carcinoma in nonalcoholic fatty liver disease: A multicenter prospective study. *Hepatology* **63**, 827–838 (2016).
39. Dongiovanni, P. *et al.* Statin use and nonalcoholic steatohepatitis in at risk individuals. *J Hepatol* **63**, 705–712 (2015).
40. Kleiner, D. E. *et al.* Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* **41**, 1313–1321 (2005).
41. Angulo, P. *et al.* The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology* **45**, 846–854 (2007).
42. Valenti, L. *et al.* Patatin-Like phospholipase domain-containing 3 I148M polymorphism, steatosis, and liver damage in chronic hepatitis C. *Hepatology* **53**, 791–799 (2011).
43. Milano, M. *et al.* Transmembrane 6 superfamily member 2 gene E167K variant impacts on steatosis and liver damage in chronic hepatitis C patients. *Hepatology* **62**, 111–117 (2015).
44. Valenti, L. *et al.* PNPLA3 I148M polymorphism, clinical presentation, and survival in patients with hepatocellular carcinoma. *PLoS One* **8**, e75982 (2013).
45. Burza, M. A. *et al.* PNPLA3 I148M (rs738409) genetic variant and age at onset of at-risk alcohol consumption are independent risk factors for alcoholic cirrhosis. *Liv Int* **34**, 514–520 (2014).
46. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575 (2007).

## Acknowledgements

We thank all members of the Metabolic Liver Disease center for helpful comments and discussion, Maria Antonella Burza, Rebecca Darlay and Maria Luisa Attilia for technical and clinical assistance. LV was supported by the Ricerca Corrente, Fondazione IRCCS Ca' Granda Milano (Institutional research funds; Public Research Institute), Associazione Malattie Metaboliche del Fegato ONLUS (non-profit organization for the study and care of metabolic liver diseases), and the Fondazione Policlinico – INGM Molecular Medicine grant 2014–2016, My First AIRC Grant project code 16888; intramural research grants from the Catholic University of the Sacred Heart (Linea D.3.2-2013 and Linea D.1 to A.G.) and PRIN 2010–2011 (Prot. N. 2010C4JJWB). QMA and HR were supported by the Horizon 2020 Framework Program of the European Union (EPoS: Grant Agreement 634413) and the facilities of the Newcastle NIHR Biomedical Research Centre.

## Author Contributions

Benedetta Donati: study design, manuscript drafting, data analysis and interpretation, genotypization. Paola Dongiovanni: genotypization, gene expression analysis, data analysis and interpretation. Stefano Romeo: manuscript revision for important intellectual content, data interpretation. Marica Meroni: genotypization, gene expression analysis. Misti McCain: data and samples collection. Luca Miele: data and samples collection. Salvatore Petta: data and samples collection. Silvia Maier: data and samples collection. Chiara Rosso: data and samples collection. Laura De Luca: data and samples collection. Ester Vanni: data and samples collection. Stefania Grimaudo: genotypization. Renato Romagnoli: data and samples collection. Fabio Colli: data and samples collection. Flaminia Ferri: data and samples collection. Rosellina Margherita Mancina: genotypization. Paula Iruzubietta: genotypization. Antonio Craxi: data and samples collection. Anna Ludovica Fracanzani: data and samples collection. Antonio Grieco: data and samples collection. Stefano Ginanni Corradini: data and samples collection. Alessio Aghemo: data and samples collection. Massimo Colombo: data and samples collection. Giorgio Soardo: data and samples collection, revision of manuscript for important intellectual content. Elisabetta Bugianesi: data and samples collection, revision of manuscript for important intellectual content. Quentin M. Anstee: data and samples collection, data analysis and interpretation, revision of manuscript for important intellectual content. Helen Reeves: data and samples collection, revision of manuscript for important intellectual content. Silvia Fargion: data and samples collection, revision of manuscript for important intellectual content. Luca Valenti: study design, manuscript drafting, data analysis and interpretation, study funding and supervision.

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-04991-0](https://doi.org/10.1038/s41598-017-04991-0)

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017