Università degli studi di Milano

Scuola di dottorato in Medicina Clinica e Sperimentale

XXIX ciclo

Identification of rare and common genetic variants associated with hepatocellular carcinoma in patients with nonalcoholic fatty liver disease: mechanisms and clinical implications

Tutor: Prof. Luca Valenti

PhD student: Dr. Benedetta Donati

Index

Abstract

Nonalcoholic fatty liver disease (NAFLD) affects roughly 30% of the general population and its prevalence is increasing worldwide, in particular in Western countries where it is projected to become the main cause of hepatocellular carcinoma (HCC) within 2025. Although the majority of NAFLD progressing towards HCC are individuals with advanced fibrosis, NAFLD-HCC frequently develops even in patients without cirrhosis. Family history and genetic factors play an important role in the pathogenesis of progressive NAFLD and of HCC. Our hypothesis is that both common and rare variants may have a role in influencing HCC predisposition.

Aim of this study was (1) to evaluate whether the rs641738 C>T common genetic variant in the *MBOAT7/TMC4* locus, the last genetic variant recently associated with NAFLD development and progression towards early phases of damage, also predisposes to HCC in patients stratified by the presence of severe fibrosis, and (2) to assess whether telomere attrition and inherited rare *hTERT* mutations associates with NAFLD-HCC development, since mutations in Telomerase gene have been linked to familial liver diseases.

In the first part of the study, we found that the rs641738 T allele is associated with NAFLD-HCC independently of fibrosis severity and of several confounders (allelic OR 1.65, 95% confidence interval 1.08-2.55; n=765), particularly in patients without advanced fibrosis ($p<0.001$). The rs641738 risk T allele is associated with reduced MBOAT7 expression, evaluated by qRT-PCR on liver biopsy, specifically in patients without severe fibrosis, and was more strongly associated with HCC development in this specific subgroup ($p=0.02$). Moreover, the number of *PNPLA3*, *TM6SF2*, and *MBOAT7* risk alleles is associated with NAFLD-HCC development independently of clinical factors ($p<0.001$), but it did not significantly improve their predictive accuracy. The independent association between the

MBOAT7 risk variant and increased HCC risk has also been confirmed in a combined cohort of chronic hepatitis C or alcoholic liver disease patients without cirrhosis (n=1121).

In the second part of the study, we found that telomere length, evaluated by qRT-PCR, is progressively reduced in 40 NAFLD-HCC vs. 45 NAFLD-cirrhosis (p=0.048) and 64 healthy controls (p=0.0006), independently of age and sex. We detected an enrichment of rare germline coding mutations in *hTERT* in NAFLD-HCC patients vs. controls, even after considering 78 European primary liver cancers. No mutations were found in NAFLD-cirrhosis and local controls, and only one in 503 Europeans from the 1000Genomes Project. Thus, overall mutations frequency was 0.025 in cases vs. <0.001 in controls (p=0.0005 at Burden test). *hTERT* mutations occurred predominantly in females (p=0.03) and were predominantly located in the N-terminal template-binding domain of the gene (p=0.037 for specific enrichment). The frameshift Glu113Argfs*79 and the missense Glu668Asp damaging mutations co-segregated with liver disease in families. Besides Glu668Asp, the Ala67Val variant resulted in reduced intracellular protein levels in HEK-293 cells. These data suggest that telomere attrition and certain *hTERT* mutations are involved in the pathogenesis of NAFLD-HCC.

In conclusion, the identification of new common prognostic marker, such as the *MBOAT7* rs641738 variant, in NAFLD patients and the evaluation of *hTERT* rare mutations in NAFLD-HCC may improve HCC screening strategies, by assisting in the identification of subjects at risk even in the absence of severe fibrosis and of family members at high risk of progressive liver disease, respectively. In future, functional studies are needed in order to understand the exact role and the interplay of the several common and rare genetic variants which contribute to the molecular mechanisms that underlie the pathogenesis of NAFLD-HCC.

1.Introduction

1.1 Epidemiology of hepatocellular carcinoma associated with nonalcoholic fatty liver disease

Hepatocellular carcinoma (HCC) is the sixth most common cancer and the third cause of cancer-related mortality worldwide [1]. Nowadays chronic viral hepatitis due to HBV or HCV infection represents the main cause of HCC development. Nevertheless, because of the diffusion of the vaccination against HBV and the availability of new antiviral treatment, and of the pandemic of obesity and insulin resistance on the other side, nonalcoholic fatty liver disease (NAFLD) is projected to become the leading cause of HCC in Western countries within 2025 [1-4]. NAFLD is defined as the hepatic accumulation of neutral lipids greater than 5% of liver weight and affects 16-38% of the general population worldwide [5]. It is a leading cause of cirrhosis, the main risk factor for HCC, which is due to progressive fibrosis and represents *per se* a pre-cancerous condition. Although the growing rates of HCC may be due to the increased number of individuals with advanced fibrosis, NAFLD-HCC frequently develops without overt cirrhosis suggesting that steatosis directly promotes hepatic carcinogenesis [6-9]. However, progression of liver disease to cirrhosis and HCC is more frequent in the subgroup of patients who develop non-alcoholic steatohepatitis (NASH), a condition characterized by active inflammation and fibrosis [10].

Established risk factors for disease progression in NAFLD include older age and presence of features of the metabolic syndrome, such as obesity, severe insulin resistance, and hypertension. Consistently, NAFLD-HCC patients are most commonly older males, with type 2 diabetes (T2D) and meeting criteria for at least one feature of the metabolic syndrome. HCCs arising in patients with features of the metabolic syndrome are larger, more

differentiated and mainly occurred in the absence of significant fibrosis compared than those arising in chronic viral hepatitis [11].

In addition, it has been estimated up to 30% of HCCs in industrialized countries develop in patients with cryptogenic cirrhosis, a condition which is retained to stem in the majority of cases from burnt-out NASH [3, 12, 13].

Several epidemiological studies established an association between overweight and obesity, that are considered the major determinants of insulin resistance and NAFLD, and higher risk to develop HCC (17% and 89% respectively compared to normal weight). Furthermore the relationship linking obesity to HCC risk is stronger in male than in female [14, 15]. Even diabetes has been independently correlated to HCC onset in large epidemiological studies, where it was found that among men affected by T2D the risk for HCC was doubled [16, 17].

On one hand, these data suggest that obesity and diabetes are the major epidemiological determinant of HCC incidence in Western countries. On the other hand, patients with progressive NAFLD are mostly unaware of being affected by a progressive form of liver disease. Therefore, due to the very high prevalence of NAFLD, occurrence in patients without advanced fibrosis, and lack of diseases awareness, classic screening strategies for the detection of early HCC are ineffective [1]. This renders the development of new noninvasive biomarkers able to stratify the risk of HCC development in NAFLD patients a public health priority.

1.2 Pathogenesis of hepatocellular carcinoma (HCC) associated with nonalcoholic fatty liver disease (NAFLD)

The mechanisms linking NAFLD to liver disease progression towards HCC have not yet been identified. Anyhow several pathways may be activated in obesity and diabetes favoring a tumor-promoting environment distinguishing the pathogenesis of NAFLD-HCC from that of

other etiologies [1, 3, 18, 19]. First of all, increased cancer risk is associated with a low-grade of chronic inflammation, a manifestation typical of obesity and metabolic syndrome. Indeed, adipose tissue expansion promotes the release of pro-inflammatory cytokines, namely tumor necrosis factor alpha (TNFα) and interleukin 6 (IL6) both potent activators of key oncogenic signaling pathways [20]. Furthermore, obesity alters the release of adipokines reducing the level of those with anti-inflammatory effects such as adiponectin and arising the level of those with pro-inflammatory and fibrogenic effects, such as leptin [21, 22].

Even though this mechanism remain to be elucidated, cancer cells proliferation rate and HCC poor prognosis correlate with increased lipogenesis and accumulation of lipid droplets, suggesting that increased availability of fatty acids in hepatocytes may support HCC cells growth [23, 24].

Lipotoxicity may also contribute to HCC development by interfering with intracellular signaling pathways. Lipid peroxides and free radicals generated by increased rates of fatty acids oxidation, may induce mitochondrial dysfunction and endoplasmic reticulum stress thus altering gene transcription [25, 26].

Overall the factors listed above collectively induce hyperinsulinemia, resulting in increased bioavailability of insulin-like growth factor-1 (IGF1) which in turn promotes cellular proliferation and inhibits apoptosis [27].

Finally, some studies have been conducted in order to identified the biomarkers able to distinguish the pathogenesis of HCC according or not to the presence of cirrhosis. The deregulation of the Wnt/β-catenin signaling pathways seems to be specific of HCC onset in the absence of advanced liver fibrosis [11]. The activation of hepatic stellate cells (HSCs) is a major step in the development of cirrhotic HCC, however these cells not only secrete collagen that results in liver fibrosis, but may even produce several growth factors (senescence-associated secretory phenotype – SASP) which stimulate oncogenic pathways contributing to the expansion of neoplastic clones [28]. Interestingly, signs of cellular senescence and SASP

have been observed in HSCs not related to fibrotic areas but located in proximity of NAFLD-HCC lesions, indicating that these cells may contribute to the development of HCC also in the absence of cirrhosis [29].

1.3 Role of genetics in NAFLD progression towards HCC

Genetic factors have been shown to influence disease progression in NAFLD, and family history remains the main risk factor for HCC development [10, 30]. The common genetic polymorphism rs738409 C>G encoding for the I148M variant in Patatin-like phospholipase domain-containing protein 3 (PNPLA3 or adiponutrin) has been established as the common genetic determinant of hepatic fat content and of progressive NAFLD [31-35]. The mechanism is related to accumulation of the mutated protein [36], which interferes with lipid droplets remodeling in hepatocytes [34, 37, 38], and with retinol release by hepatic stellate cells [39, 40]. The PNPLA3 variant predicts HCC development in European patients with NAFLD [41] and also in individuals affected by other liver diseases associated with steatosis, namely alcoholic liver disease (ALD) and chronic hepatitis C (CHC) [42]. This evidence suggests that this genetic risk factor may be helpful to select high-risk individuals for screening [42-44], but it has a low sensitivity to be used as single prognostic biomarker [45].

The rs58542926 E167K variant in Transmembrane 6 superfamily member 2 (TM6SF2) also predisposes to progressive NAFLD by altering the secretion of very low-density lipoproteins [46-48], but its direct role in HCC predisposition is disputed [47, 49].

More recently, it has been found that the rs641738 C>T sequence variant in the Membrane bound O-acyltranferase domain containing 7/ Transmembrane channel like 4 (MBOAT7/TMC4) locus, involved in phospholipids remodeling, predisposes to cirrhosis development in individuals with excessive alcohol intake [50], and to the development and

the progression of NAFLD in individuals of European descent [51]. However, whether the rs641738 variant is also associated with HCC risk is still unknown.

Besides the validated common genetic risk factors influencing lipid metabolism, the influence of variants involved in fibrogenesis has recently been described. Fibrosis stage and liver disease progression are strictly linked to cell senescence. Consistently, hepatocyte expression of p21, playing a pivotal role in the induction and maintenance of cellular senescence, was associated with fibrosis stage in NAFLD and increase liver related morbidity and mortality [52]. p21 is encoded by the Cyclin-dependent Kinase 1A (CDKN1A) gene. The rs762623 G>A variant in the promoter region of CDKN1A, inducing a reduction in p21 expression level, has been associated with the development but not the progression of fibrosis in NAFLD independently from well recognized PNPLA3 I148M status [53]. Interestingly, CDKN1A variants have previously been described in association with rapid progression of idiopathic pulmonary fibrosis, a degenerative condition characterized by cellular senescence and impairment of telomeres [54, 55].

Consistently, genetic data indicate that NAFLD is commonly observed in patients with telomeropathies, suggesting that steatosis may either be a consequence of hepatocellular senescence, as also observed in animal models, or a trigger for liver disease progression [56, 57]. Indeed, loss of function germline mutations in the telomerase reverse transcriptase (*hTERT)* can predispose to a spectrum of familial liver diseases characterized by steatosis [56] and possible evolution to cirrhosis and HCC [58, 59]. In keeping, we previously reported the occurrence of NAFLD-HCC in a patient with a rare germline *hTERT* loss-of-function mutation [60]. Furthermore, it has also been reported that rare mutations inducing Mendelian diseases due to severe derangements in the function of encoded proteins may predispose to NAFLD-HCC. Indeed, mutations in Apolipoprotein B (*APOB)* may explain some familial cases through predisposition towards development of severe steatosis caused by hepatocellular retention of lipids [61].

1.4 The *MBOAT7* rs641738 common variant is a risk factor for NAFLD onset and progression

The rs641738 C>T polymorphism located in the region of chromosome 19 proximal to the *MBOAT7* gene, has been identified as a risk factor for cirrhosis development in a genome wide association study conducted in a cohort of alcohol abusers [50]. Afterwards, the impact of this variant on hepatic fat accumulation and metabolism has been demonstrated in two different independent cohorts: the Dallas heart study (DHS), a multiethnic population study conducted in Texas, and the Liver biopsy cross-sectional cohort, including European patients at risk of NASH. The T allele was associated with the entire spectrum of histological liver damage related to NAFLD: from steatosis, to necro-inflammation and early stages of fibrosis [51].

The MBOAT7 protein is involved in Lands' cycle regulating phospholipid acyl-chain remodeling. Particularly, it catalyzes the transfer of polyunsaturated fatty acids such as arachidonoyl-CoA to lyso-phosphatidylinositol and other lyso-phospholipids, thereby allowing to achieve an adequate level of desaturation [51]. The rs641738 T allele is associated with reduced MBOAT7 expression and consequently alters phosphatidyl-inositol plasma and hepatic composition [51, 62].

1.5 *hTERT* mutations and telomere shortening are risk factors for liver disease progression

Telomeres are simple repeat non coding DNA sequence (TTAGGG) located at the end of each chromosome. Their main function is chromosome capping in order to avoid chromosomes erosion by nuclease and chromosomes end-to-end fusion. During each round of cell division, telomere length reduces due to the inability of polymerase to fully replicate the terminal chromosomal segment, functioning as a "mitotic clock" to sense cell aging. When

telomeres become critically short, a DNA-damage program is activated leading to apoptosis or cell senescence (due to the Hayflick limit) [55, 63]. Telomere attrition is exacerbated in degenerative conditions characterized by chronic injury and regeneration with accelerated cell turnover [64]. Thus, limiting the replicative capacity of cells and the number of cells participating in tissue regeneration, shortened telomeres play a causal role in the pathogenesis of liver fibrosis [65-69]. Reduction of telomere length is considered a hallmark of cirrhotic tissue independently of the etiology of liver disease (e.g. viral hepatitis, autoimmune hepatitis, alcohol abuse…) [68]. A causal role of telomere shortening in fibrosis progression has been experimentally demonstrated: *TERT* deficient mice, after three generations, developed shortened telomeres and displayed diminished capacity for liver regeneration, and accelerated development of cirrhosis after liver injury. On the contrary, overexpression of *TERT* activity improved liver function and protected from development of hepatic steatosis and fibrosis [57]. Telomere length is a strongly hereditable tract. Consistently, genetic studies have proven that mutations in Telomerase represent the underlying cause of accelerated telomere shortening and organ failure. A significant enrichment of germline missense mutations in the *hTERT* and *hTERC* genes was observed in both a US cohort and a larger series of German patients with cirrhosis of different etiologies [59]. These mutations impaired *hTERT* enzymatic activity, as they were associated with reduced telomere length in the peripheral blood of patients and reduced telomerase activity in vitro [58]. These observations indicate that, in at least a proportion of patients who developed cirrhosis, fibrosis progression may be favored by genetic risk variants facilitating telomere shortening and cell senescence in the presence of triggering factors.

In this context, telomere shortening may favor carcinogenesis by directly facilitating genomic instability. Indeed, shortened telomeres have been associated with the typical karyotipic alterations in HCC (chromosome 8 alterations), especially in the presence of *TP53* mutations [70, 71]. Moreover, loss of *hTERT* has been shown to affect the overall configuration of

chromatin and to diminish the capacity for DNA repair of double strand breaks (DSB) [72]. Therefore, current data suggest a model whereby telomere shortening drives chromosomal instability during early stages of hepatic carcinogenesis, while telomerase re-activation is involved in progression to malignancy in another subset of tumors, as it confers cellular immortalization. Indeed, in HCC tissues, long telomeres and increased Telomerase activity were shown to be significant poor prognostic factors, associated with clinic-pathological features of aggressive behavior [73].

Thus, HCC tumor progression is associated with the reactivation of Telomerase, which is necessary for the immortalization of the neoplastic clone [74, 75]. *hTERT* promoter mutations, the most frequent somatic genetic alteration in HCC [76, 77], represent an important mechanism of reactivation of Telomerase during hepatocarcinogenesis by creating new binding site for specific transcription factors which consequently induce *hTERT* overexpression [78, 79], whereas in other cases overexpression may be related to HBV insertional mutagenesis or duplication.

2. Aim

NAFLD has a high and increasing prevalence in western countries. Although the growing rates of NAFLD progressing towards HCC may be due to the increased number of individuals with advanced fibrosis, NAFLD-HCC frequently develops without overt cirrhosis suggesting steatosis directly promotes hepatic carcinogenesis [6, 7]. The complexity of the clinical manifestation renders patients mostly unaware of being at risk and classic HCC screening strategies aimed at early diagnosis and curative treatment unfeasible. Family history and genetic factors play an important roles in the pathogenesis of progressive NAFLD and of HCC [10, 30], however there is a lack of a common prognostic marker able to identify NAFLD patients at risk to progress towards cancer.

Our hypothesis is that both common and rare variants may have a role in influencing rare diseases pathogenesis. Common genetic risk factors in combination with acquired risk factors, thanks to their high frequency in the general population and easy detection, maybe useful biomarkers to select subjects at risk of disease development. Rare genetic variants most likely deleterious, due to their very low frequency, may have a direct pathogenic role.

Thus aim of the first part of this study was (1) to evaluate whether the *MBOAT7* rs641738 common genetic variant predisposes to HCC in patients with histologically confirmed NAFLD stratified by the presence of severe fibrosis. Moreover, progression of liver disease towards cirrhosis and possible HCC has been associated with inheritance of *hTERT* and *hTERC* mutations [56, 80] and older age and duration of liver disease remain among the major and more validated risk factors for liver disease progression [81, 82]. Thus, the aim of the second part of this study was (2) to assess whether telomere attrition and inherited rare *hTERT* mutations predispose to NAFLD-HCC development. We therefore compared peripheral telomere length and the frequency of *hTERT* rare germinal mutations between

European individuals with NAFLD-HCC and controls with NAFLD cirrhosis and without liver disease.

3.Material and methods

3.1 Study design

Study design is shown in figure 1. We investigated the role of common and rare genetic factors in the pathogenesis of NAFLD-HCC. (1) Briefly, in the first part of the project we analyzed the prevalence of the common genetic variant rs641738 in the *MBOAT7* locus in a large NAFLD cohort according to the presence of HCC and advanced fibrosis. We subsequently evaluated the impact of the risk variant on hepatic MBOAT7 expression in patients stratified by fibrosis severity. Furthermore, we validated the association of the variant with HCC development in patients affected by other liver diseases of whom steatosis influences progression, namely alcoholic liver disease (ALD) and Chronic hepatitis C (CHC). (2) In the second part, we investigated the presence of rare germline *hTERT* coding mutations in NAFLD-HCC. We defined rare coding mutations as variations with allelic frequency <0.001, according to two of the largest frequency database available namely ExAC (Exome Aggregation Consortium) Non-Finnish European (NFE) population and ESP (Exome Sequencing Project) European-American (EA) population. Furthermore, we evaluated telomere length on a discovery cohort of Italian NAFLD-HCC, NAFLD-cirrhosis and healthy individuals of comparable age and sex distribution. We attempted replication in a validation cohort including both Italian and Northern European subjects with NAFLD, who developed primary liver cancer (PLC). We subsequently conducted family studies to test for segregation of the *hTERT* mutations with liver disease. Finally, we examined the functional effects of the coding mutations identified by bioinformatics using a combination of *in silico* prediction tools and *in vitro* by cell studies with overexpression of recombinant wild-type and mutated proteins.

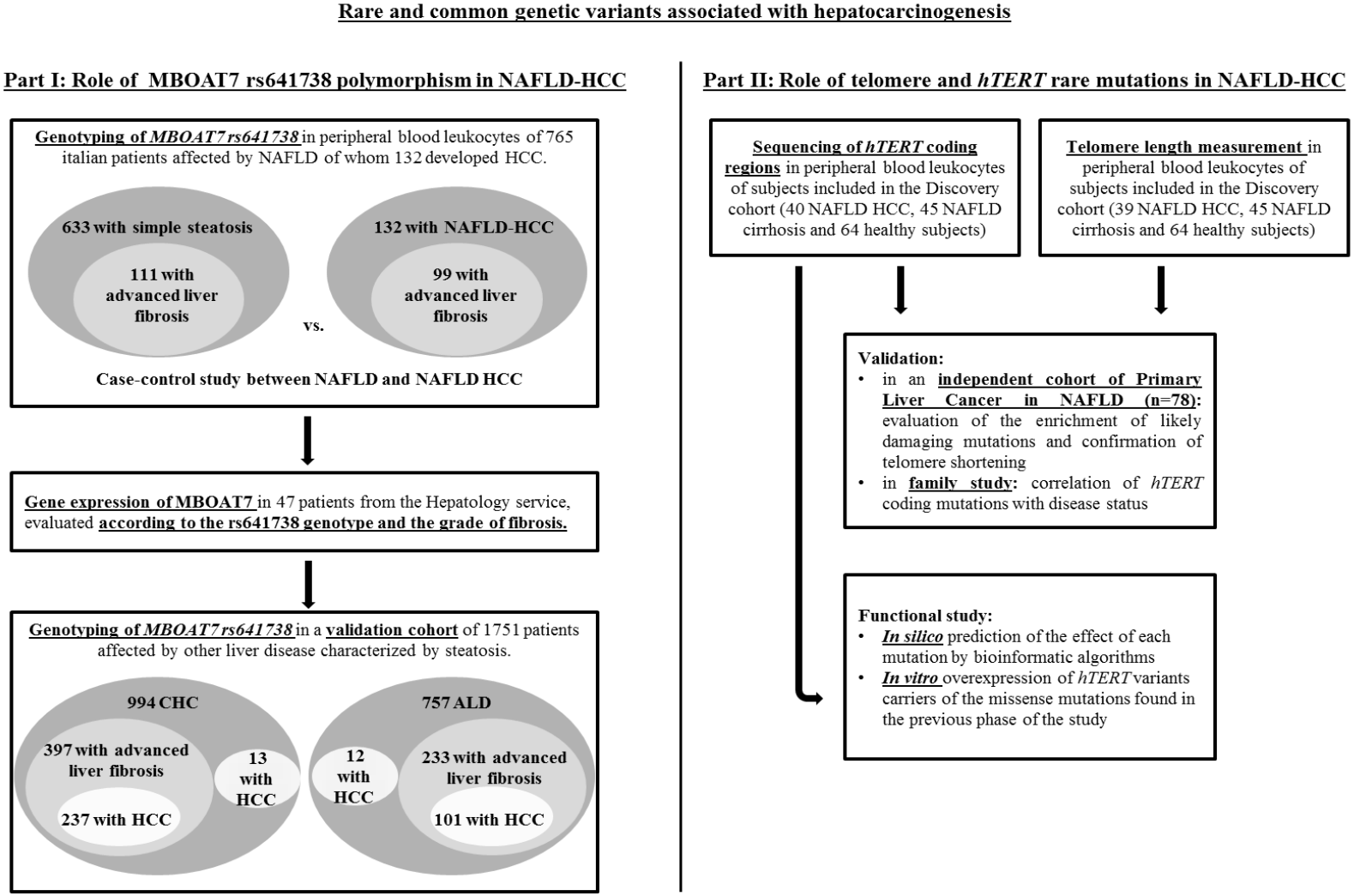

Figure 1. Study design

3.2 Patients

In this study, we collected samples of patients enrolled in several Italian and a UK centers, which joined the "Epidemic study group investigators", founded for the study of genetic factors influencing the progression of NAFLD towards HCC. The cases under study have been collected along several years from 2008 to 2015. For this reason the two different part of this study (investigating common and rare genetic variants respectively) are based on different numbers of NAFLD-HCC patients according to the period in which the projects have been developed. Moreover, differences in the number of patients involved may be related to the different quantity and quality of the samples collected not always suitable for each technique that we performed (for example sequencing required better quality and higher quantity of DNA compared to genotyping). No selection in NAFLD-HCC group collection has been made according to the clinical features of the patients enrolled.

All patients included in this study were tested for secondary causes of steatosis including alcohol abuse (≥ 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. Diagnosis of HCC was based on the EASL–EORTC Clinical Practice Guidelines for the management of hepatocellular carcinoma [83].

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.

3.2.1 Patients investigated for the association of MBOAT7 rs641738 common genetic variant with NAFLD-HCC

We enrolled 132 consecutive unrelated patients with NAFLD-HCC of Italian descent, referred between January 2008 and January 2015 to the Fondazione IRCCS Ca' Granda Policlinico Hospital of Milan (n=35), S. Maria della Misericordia Hospital of Udine (n=16), Città della Salute e della Scienza University Hospital of Torino (n=46), Policlinico Gemelli Roma (n=34), University of Palermo (n=1), for whom DNA samples were available. 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 (n=633; n=351 for Milan and n=282 from Palermo)[48, 84], from a recently published database [51], who did not develop HCC during follow-up.

Advanced fibrosis was defined in the presence of fibrosis stage F3-F4 [85], 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 ≥ 8.4 kPa evaluated by elastometry (n=3), or by a positive NAFLD fibrosis score (n=5) [86]. Obesity was defined when BMI>30 Kg/m^2. 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.

Table 1. Clinical features of 765 NAFLD patients stratified by HCC diagnosis.

	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
PNPLA3, I148M			<0.001
I/I	31 (23)	231 (36)	
I/M	55 (42)	283 (45)	
M/M	46 (35)	120 (19)	
TM6SF2, E167K			0.30
E/E	109 (83)	539 (85)	
E/K	19 (14)	88 (14)	
K/K	4 (3)	7 (1)	
MBOAT7, rs641738 C>T			0.003
C/C	26 (20)	213 (34)	
C/T	69 (52)	285 (45)	
T/T	37 (28)	135 (21)	

( ): % 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.

We next evaluated the impact of the rs641738 variant on HCC risk in an Italian multicenter cohort of patients with chronic hepatitis C (CHC; n=994) and alcoholic liver disease (ALD; n=757), stratified by the presence of cirrhosis. Of CHC patients, 815 were from the well-described histological Milan cohort [87, 88], whereas the remaining 179 were HCC cases developed in CHC patients, which were previously described by our group (Milan HCC cohort, where presence of cirrhosis was carefully assessed) [89].

ALD patients with and without HCC included 171 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 [89]. We also considered 586 consecutive individuals, who were admitted to the Outpatient Clinic at the Department of Clinical Medicine, Policlinico Umberto I, Rome (Italy) for alcohol abuse or dependence between 2005 and 2014, for whom DNA samples were still available and genotyping was successful. This series comprised 384 previously described individuals [90], and 202 additional patients subsequently enrolled using the same criteria [90]. At-risk alcohol consumption was defined as ≥3/2 alcohol units per day for M/F, respectively. Diagnosis of cirrhosis was 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 2.

Table 2. Clinical and genetic features of 994 patients with chronic hepatitis C (CHC) and 757 with alcoholic liver disease (ALD).

	CHC	ALD
n=	994	757
Age, years	60±13	50±12
Sex, Female	390 (39)	176 (23)
Cirrhosis, yes	397 (40)	233 (31)
HCC, yes	250 (25)	113 (15)
PNPLA3, I148M	505/386/103	309/330/117
TM6SF2, E167K	910/84	639/113/5
MBOAT7/TMC4, rs641738 C>T	307/452/235	217/378/161

( ): % values; HCC: hepatocellular carcinoma.

3.2.2 Patients investigated to study the role of telomere and hTERT rare coding mutations in NAFLD progression towards HCC

In phase 1, we enrolled 40 patients with NAFLD-HCC: 20 from Policlinico Hospital of Milan, 4 from S. Giovanni Battista Hospital of Turin, one from the Gastroenterology unit of Palermo and the remaining 15 from the S. Maria della Misericordia Hospital of Udine. Additionally, 45 patients affected by NAFLD cirrhosis were enrolled in order to confirm the association of the mutations eventually found with the carcinogenic phenotype (20 from Palermo and 25 from Milan). For all the patients complete clinical data and follow up are available in Table 3. Missing data (representing less than 5% for each category) have been replaced by the median for each category. These samples, belonging to the Discovery cohort, were collected from January 2012 until December 2013.

Finally, we analyzed a local ethnically matched control group of comparable sex distribution including sixty-four healthy blood donors without clinical and biochemical evidence of liver disease and no alcohol abuse [91].

In the second phase of the study we examined a validation cohort collected after January 2014, including 78 patients affected by NAFLD-PLC, two of whom had intrahepatic cholangiocarcinoma (28 from the Freeman Hospital of Newcastle upon Tyne, 24 from Turin, 25 from the Policlinico Gemelli of Rome and one from Milan). We did not exclude NAFLD-associated intrahepatic cholangiocarcinoma in the validation cohort, since we could not rule out that germline mutations in hepatic stem cells might give rise to cancers with different phenotypes. Clinical features are shown in Table 3.

Telomere length in intra-tumoral liver specimens was also evaluated in 5 Italian HCC patients, whose clinical features are listed in Table 4.

Table 3. Clinical features of subjects included in the study.

	Discovery cohort				Validation cohort	
	Healthy	Cirrhosis	HCC	p	PLC	p
Age, years	59.1±6.6	58.8±8.7	66.3±9.5	<0.0001	67.5±8.4	<0.0001
Sex, F	19 (30)	15 (33)	13 (33)	0.606	11 (14)	0.0239
BMI	25.5±2.6	30.5±4.3	28.6±4.0	0.0011	30.3±5.4	<0.0001
T2DM, y	0	29 (64)	24 (60)	<0.0001	45 (58)	<0.0001
Fibrosis, F3-4	0	45 (100)	32 (80)	<0.0001	61 (78)	<0.0001
Italian	64 (100)	45 (100)	40 (100)		50 (64)	
PNPLA3,				0.0086		<0.0001
I/I	36 (56)	9 (20)	15 (37)		16 (21)	
I/M	24 (38)	23 (51)	14 (35)		36 (46)	
M/M	4 (6)	13 (29)	11 (28)		26 (33)	

PLC: primary liver cancer; ( ): % values; y: yes; T2DM: type 2 diabetes mellitus; p: p value calculated as HCC versus Healthy subjects.

Table 4. Clinical features for characterization of HCC patients for whom biopsy samples have been analyzed.

	HCC (n=5)
Age, years	69.8±4.4
Sex, F	1 (20)
BMI	33.7±8.3
T2DM, y	5 (100)
Fibrosis, F3-4	5 (100)
Italian Ancestry	5 (100)

( ): % values; y: yes; T2DM: type 2 diabetes mellitus.

3.3 Genotyping

Patients were genotyped for rs738409 (*PNPLA3* I148M) and rs58542926 (*TM6SF2* E167K), as previously described [48]. 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 (Life Technologies, Carlsbad, CA). Genotyping success rate was >98%. The duplicate genotype concordance rate was 100%.

3.4 *hTERT* sequencing

The presence of coding mutations in *hTERT* sequence was evaluated by Sanger Sequencing on DNA previously extracted from peripheral blood leukocytes by phenol-chloroform extraction. The primers used for amplification and sequencing of the whole *hTERT* sequence, including the 16 exons and intron-exon boundaries, are listed in Table 5.

In order to evaluate the presence of known acquired somatic activating mutations, *hTERT* promoter sequencing was performed as previously described on DNA extracted from liver specimens of 5 NAFLD-HCC patients [77, 92].

Table 5. Primers used for PCR and Sanger sequencing in *hTERT* sequencing analysis.

Exon	Sequence 5'→ 3'
1-2 Forward	GAGTTTCAGGCAGCGCTGCGT
1-2 Reverse	CTTGTCGCCTGAGGAGTAGAG
1-2 Seq Forward*	CAACACGGTGACCGACGCACT
1-2 Seq Reverse*	CAGGTGAACCAGCACGTCGTC
2 Forward	CAGGACGCGTGGACCGAGTGA
2 Reverse	GTGAACCTCGTAAGTTTATGC
3 Forward	GTGATCTGGATGTGGCATGT
3 Reverse	GGTGTTCCAGGACTTCGAGA
4 Forward	GTCTGTTGTCTGGCTGAGCA
4 Reverse	GCTCAAACGCACTTCTGTTTA
5 Forward	ACTTGGCCGGATCCACTT
5 Reverse	CACTCCCAAGGTCCAGCA
6 Forward	CGTGGCCACTGTCAGTCT
6 Reverse	CAGAGACACACATCCTGGACA
7 Forward	CCACATTTGTGGCTCATGC
7 Reverse	TCATGAGCCCAGTGATTGC
8 Forward	GTCCTGCCTGTCTCAGCAC
8 Reverse	GAAGGGGCAGGAGAGAGGT
9 Forward	AGTGTACGCATGTCCAGCAC
9 Reverse	AATCAACCCCCACCCAAG
10 Forward	CCGCTTTGGAGAATGTTACTT
10 Reverse	TGGAGGTCCCCACAGACA
11 Forward	TCAGGTTACCTCCTGGGTGA
11 Reverse	CCACACGGAAGCAGAGGT
12 Forward	GCAGGAGGCTCTTTGGAG
12 Reverse	TCCTGAACTCTGAACTCTGTG
13 Forward	CCAGAGAGGTTTCTACCGTTT
13 Reverse	GGTCAGAGGTGAGCAGAGC
14 Forward	ACGAGCACCGTCTGATTAGG
14 Reverse	AGGCACTGCTGCCACTGA
15 Forward	CAGCTTTCCGGTGTCTCCT
15 Reverse	GGGCGTTCAAGGATGACC
16 Forward	CACCTCTGGCCTCTTCTGG
16 Reverse	GACAGGGCTGCTGGTGTCT

*Primer used only for sequencing

3.5 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 [51]. 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 6. 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 [51]. Association analysis between rs641738 variant (additive model) and gene expression, and linkage with rs8736 were conducted by the PLINK v1.07 genetic analysis software.

Table 6. Clinical and genetic features of 47 patients from the Milan Hepatology service, for whom hepatic gene expression analysis was available.

	Mean±SD; n (%)
Age, years	48±10
Sex, F	13 (28)
BMI, Kg/m^2	30.3±5.1
T2DM or IFG	18 (38)
NASH, yes	39 (83)
Fibrosis, stage F2-F4	26 (55)
PNPLA3, I148M	11/23/13 (23/50/28)
TM6SF2, E167K	35/12 (74/26)
MBOAT7/TMC4, rs641738 C>T	22/19/6 (47/40/13)

T2DM: type 2 diabetes mellitus.

3.6 Telomere length measurement

DNA was obtained from peripheral blood leukocytes or liver biopsies by phenol-chloroform extraction. Quality control was performed by evaluating 260/280 nM absorbance ratio and by polyacrylamide gel electrophoresis. Mean telomere length was measured by quantitative real-time polymerase chain reaction (qPCR), as previously described [93-96]. Briefly, PCR was conducted in triplicate in a 7500Fast Real Time PCR System (Life Technologies, Foster City, CA) and results are presented as ratio of telomere repeat copy number to 36B4 single gene copy number, calculated considering the relative quantity of the two distinct PCR products.

3.7 Mutagenesis, cloning and overexpression of *hTERT* variants

Wild type *hTERT* cDNA was synthesized and cloned in the pcDNA 3.1 vector with a V5 epitope tag at the C-terminus by GeneArt Gene Synthesis (Thermo Fisher Scientific, Rockford, IL, USA). Single base-pair changes resulting in an alanine to valine substitution (Ala67Val) or a glutamate to aspartate substitution (Glu668Asp) were introduced by overlap extension PCR cloning. Single base-pair changes resulting into a proline to leucine substitution (Pro193Leu) or a histidine to proline substitution (His296Pro) were introduced by site-directed mutagenesis. The PCR products were cloned in pcDNA 3.1 vector (pcDNA 3.1 Directional TOPO Expression Kit; Invitrogen, Carlsbad, CA). The presence of the *hTERT* mutations and fidelity of each construct were confirmed by DNA sequencing. Human embryonic kidney cells (HEK-293) and immortalized human hepatic stellate cells (LX-2, kindly provided by Professor Scott L. Friedman, [97]) were cultured in DMEM (Dulbecco's Modified Eagle Medium) containing 10% FBS (Fetal Bovine Serum). Expression plasmids (30 μg/T-75 flask) containing the human wild type hTERT or mutants, were used to transfect HEK-293 or LX-2 cells using Lipofectamine 3000 (Thermo Fisher Scientific) reagent according to the manufacturer's protocol. After 48 h cells were collected. Cells were lysed in

M-PER® (Mammalian Protein Extraction Reagent, Thermo Fisher Scientific) containing complete protease inhibitors cocktail (Sigma-Aldrich, Saint Louis, Missouri, USA) and analyzed by Western blotting. The intensity of the Western blotting bands was measured by Image Lab Software (Bio-Rad) and expressed as arbitrary unit (AU). The highest value obtained was assigned as 1.

3.8 Bioinformatics and statistical analysis

For descriptive statistics, continuous variables are shown as mean and standard deviation, while categorical variables are presented as number and proportion.

The odds of HCC for *MBOAT7* rs641738 T alleles were estimated by logistic regression models, assuming an additive effect of rs641738, and adjusted for ethnicity, age, gender, BMI, IFG/T2DM, *PNPLA3* I148M and *TM6SF2* E167K genotypes (clinical and genetic factors previously associated with or candidate for liver disease evolution to HCC) [43, 47]. All genetic analyses in the first part of the study were performed assuming an additive model. Interactions among genes and between genes and environmental factors were tested by adding the corresponding gene*gene and gene*environment interaction terms to the multivariable-adjusted regression model. Population attributable risk (PAR) of genetic variants on HCC has been estimated as previously described [98]. A NAFLD-HCC risk score was developed as previously described [86].

Telomere lengths comparisons have been conducted by generalized linear model corrected for sex and age.

All the variants found in the *hTERT* sequencing experiment were functionally annotated using the *hTERT* RefSeq reference transcript NM_001193376. We defined rare coding mutations as variations not present in dbSNP (release 147) and 1000Genomes Project (Phase 3) or described with a minor allele frequency (MAF) <0.001 according to ExAC-NFE and ESP-EA

populations. The rare variants identified were submitted to the ClinVar database (www.ncbi.nlm.nih.gov/clinvar/; submission ID: SUB2041085 (MDI-7607)).

Burden test was performed using the collapsing method CAST (Cohort Allelic Sum Test) available in the R package AssotesteR (http://CRAN.R-project.org/package=AssotesteR). Briefly, the genotype of the rare coding mutations for the individual of each group of patients and controls was summarized and collapsed into a single genetic score, taking into account both the number of mutated alleles in each group and the mutation frequency in controls. The association of this score with the trait was tested using Fisher test. The association was considered statistically significant with p-values lower than 0.05.

The impact of rare coding mutations on protein activity was predicted *in silico* by different bioinformatics algorithms including Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/), SIFT (http://sift.jcvi.org) and PROVEAN (http://provean.jcvi.org).

Statistical analyses were carried out using the JMP 12.0 Statistical analysis software (SAS Institute, Cary, NC) and PLINK v1.07 [99]. 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.

4.Results

4.1 The *MBOAT7* rs641738 variant increases the risk of hepatocellular carcinoma in European individuals without cirrhosis

4.1.1 The NAFLD cohort

The clinical features of NAFLD patients stratified by HCC diagnosis are presented in Table 1 (Material and methods section). Patients who developed HCC were older ($p<0.001$), and had higher prevalence of type 2 diabetes (T2DM; ($p<0.001$)), and had more frequently advanced fibrosis (stage F3-F4; $p<0.001$) than those without who did not. In contrast, we did not detect significant differences in sex distribution and prevalence of obesity (p=NS).

Concerning genetic risk factors, at univariate analysis development of HCC was associated with the I148M PNPLA3 variant ($p<0.001$).

The clinical features of NAFLD-HCC patients according to the presence of advanced fibrosis are reported in Table 7. Patients who developed HCC in the absence of advance fibrosis (n=21, 17%) were more frequently males (p=0.040), and carriers of the MBOAT7/TMC4 rs641738 risk T allele (p=0.006).

Table 7. Clinical and genetic features of 132 Italian patients with NAFLD-HCC stratified by the presence of severe hepatic fibrosis at diagnosis.

	Severe fibrosis		p
	Yes (n=111, 83%)	No (n=21, 17%)	
Age, years	66.3±8.3	68.3±12.3	0.49
Sex, F	24 (22)	1 (5)	0.040
Obesity, n	31 (28)	8 (38)	0.43
T2DM, y	70 (63)	14 (67)	0.97
PNPLA3, I148M			0.052
I/I	22 (20)	9 (43)	
I/M	49 (44)	6 (29)	
M/M	40 (36)	6 (27)	
TM6SF2, E167K			0.27
E/E	94(84)	15 (71)	
E/K	14 (14)	5 (24)	
K/K	3 (3)	1 (5)	
MBOAT7, rs641738			0.006
C/C	25 (23)	1 (5)	
C/T	60 (53)	9 (43)	
T/T	26 (24)	11 (52)	

( ): % values; T2DM: type 2 diabetes mellitus.

4.1.2 The MBOAT7 locus rs641738 sequence variant is associated with HCC

The frequency distribution of the *MBOAT7* rs641738 C>T polymorphism in NAFLD patients stratified by the presence of HCC is shown in Figure 2A. There was a significant over-representation of the rs641738 T allele in HCC vs. non-HCC NAFLD patients (p=0.003, Figure 2A and Table 1). In the NAFLD population, which was selected due to referral for suspected steatohepatitis and/or 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 distribution of the rs641738 variant did not violate Hardy-Weinberg equilibrium in 243 unselected healthy control subjects (p=NS; Table 8).

Table 8. Frequency distribution of the rs641738 C>T variant in 243 healthy individuals with available DNA samples, normal liver enzymes with low probability of steatosis*. HWE p=0.36.

rs61738 genotype	N=	%
C/C	78	32
C/T	113	45
T/T	52	23

* The full cohort was already described [100].

The clinical features of NAFLD patients stratified by the rs641738 genotype are shown in Table 9. 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 advanced fibrosis F3-F4 (p=0.006). As expected, the rs641738 T allele was nearly associated with advanced fibrosis (stage F3-F4; OR 1.24, 95% c.i. 1.00-1.54; p=0.052). After stratification of patients for fibrosis severity (stage F0-F2 vs. F3-F4), the rs641738 T allele was associated with HCC in patients without (p<0.001; Figure 2B), but not in those with (p=0.55; Figure 2C) advanced fibrosis.

Table 9. Clinical features in 765 Italian patients with NAFLD stratified by MBOAT7 rs641738 polymorphism according to HCC diagnosis.

	Hepatocellular carcinoma								Overall			
	Yes				No							
		rs641738		p		rs641738		p		rs641738		p
	C/C	C/T	T/T		C/C	C/T	T/T		C/C	C/T	T/T	
n=	26	69	37		213	285	135		239	354	172	
Age, years	68±10	66±9	68±8	0.36	47±12	48±12	48±14	0.84	49±14	51±13	52±15	0.10
Sex, Female	5 (19)	14 (20)	6 (16)	0.72	52 (24)	68 (24)	43 (32)	0.17	57 (24)	82 (23)	49 (28)	0.33
Obesity, yes	4 (15)	24 (35)	12 (33)	0.035	66 (31)	94 (33)	42 (31)	0.92	70 (29)	117 (33)	54 (31)	0.36
T2DM, yes	15 (58)	43 (62)	26 (70)	0.29	40 (19)	58 (20)	26 (19)	0.92	55 (23)	101 (28)	52 (30)	0.050
Severe fibrosis, F3-F4	25 (96)	59 (85)	27 (73)	0.006	30 (14)	41 (14)	27 (20)	0.17	55 (23)	100 (28)	54 (31)	0.052

( ): % values; T2DM: type 2 diabetes mellitus. Comparisons were performed by logistic regression setting HCC as dependent variable, and the association with the *MBOAT7* variant was analyzed assuming an additive model.

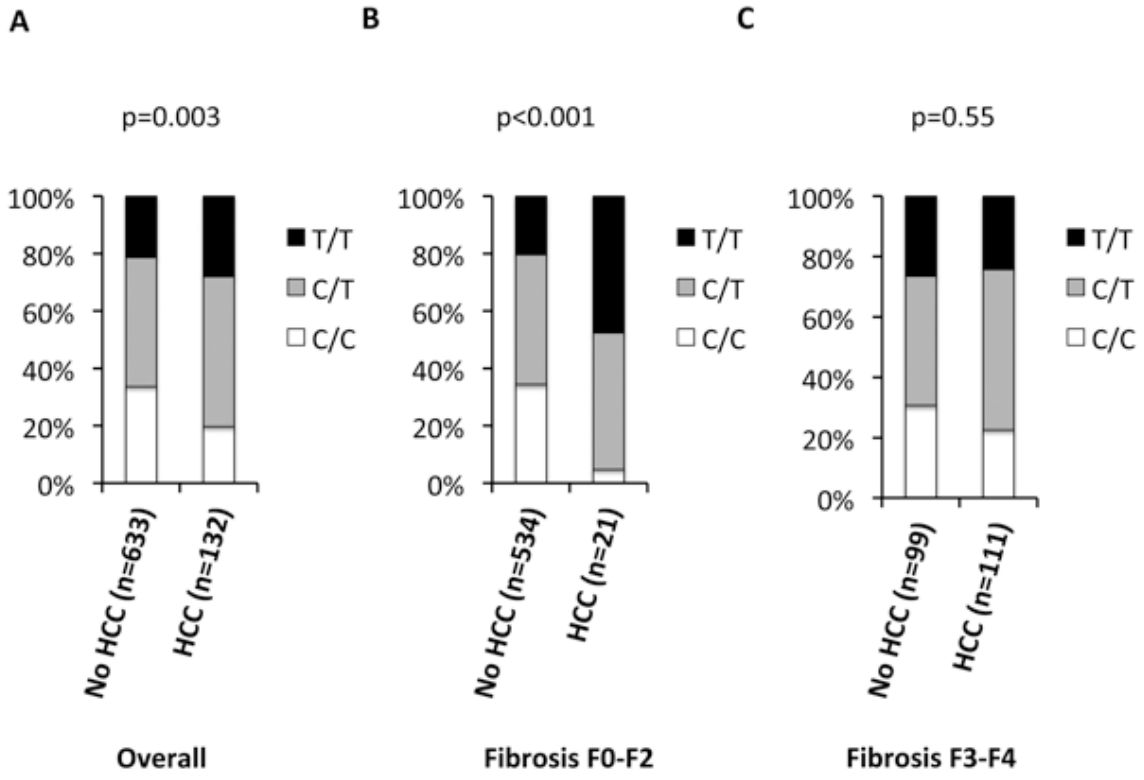

Figure 2. Frequency distribution of the *MBOAT7* locus rs641738 T allele in 765 Italian NAFLD patients stratified by the presence of hepatocellular carcinoma (HCC). A) Overall cohort; B) patients with stage F0-F2 fibrosis; C) patients with stage F3-F4 fibrosis. Comparisons were performed by logistic regression setting HCC as dependent variable, and the association with the *MBOAT7* variant was analyzed assuming an additive model.

4.1.3 Independent predictors of NAFLD-HCC

The independent predictors of NAFLD-HCC are presented in Table 10. At univariate analysis (left panel), development of HCC was associated with older age, T2DM, and presence of severe fibrosis (p<0.001 for all), whereas among the genetic factors with *PNPLA3* I148M alleles (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) such as age, sex, presence of obesity, T2DM, number of *PNPLA3* I148M, *TM6SF2* E167K, and *MBOAT7* rs641738 T alleles (middle panel), HCC 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 *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. A pooled estimate of the impact of the I148M variant on NAFLD-HCC risk in this cohort and a UK cohort [43] is presented in Table 11. This meta-analysis confirmed that the I148M variant predisposes to HCC independently of the severity of fibrosis (OR 1.60, 95% c.i. 1.14-2.25; p=0.007). 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).

Table 10. Independent predictors of HCC in 765 Italian patients with NAFLD

	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
PNPLA3, 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
TM6SF2, 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
MBOAT7, 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

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 the presence of advanced fibrosis; NA: not addressed.

Table 11. Pooled estimates of HCC risk in NAFLD patients for the presence of the *PNPLA3* I148M variant.

Genetic variant	HCC	No-HCC	OR	95% c.i.	p value
Italian cohort	132	633	1.31	0.86-2.03	0.24
UK cohort	100	275	2.26	1.23-4.14	0.008
Pooled estimate	232	908	1.60	1.14-2.25	0.007

Odds ratios (OR) and 95% confidence intervals (c.i.) were calculated for carriage of each genetic variant in the fully adjusted model (including for both studies age, sex, obesity/BMI, presence of type 2 diabetes, and severe fibrosis/cirrhosis). Pooled estimates (OR, 95% c.i.) were calculated by the inverse variance method (fixed effect model) by the Review Manager v.5 software (Cochrane collaboration). p=0.13 for heterogeneity between studies.

4.1.4 Combined effect of acquired and common genetic risk factors for HCC

The PAR of genetic risk factors of NAFLD-HCC, estimated based on Model 1 presented in Table 10, is presented in Table 12. These data suggest that the *MBOAT7* T allele has a sizable impact on HCC risk in this specific cohort (estimate PAR 27%, 95% c.i. 2-49%), which was comparable to that of the I148M *PNPLA3* variant.

Table 12. Population attributable risk of HCC in 765 Italian patients with NAFLD for *PNPLA3, TM6SF2,* and *MBOAT7* variants.

	PAR	95% c.i.
PNPLA3, I148M alleles	22%	5-38%
TM6SF2, E167K alleles	8%	1-18%
MBOAT7, rs641738 T alleles	27%	2-49%

PAR: population attributable risk; 95% c.i.: 95% confidence interval.

In both multivariate models (Model 1 and 2), there was no significant interaction among genetic risk factors in determining HCC risk (p>0.1 for all). However, there was a significant interaction between the number of rs641738 T alleles carried and the presence of obesity in determining HCC risk (p=0.035 in Model 2, shown in Table 13).

Table 13. Full model reporting the independent predictors of HCC in 765 Italian patients with NAFLD.

	Full model (Model 3)		
	OR	95%c .i.	p
Age, years	1.17	1.13-1.22	<0.001
Sex, Female	0.40	0.18-0.84	0.008
Obesity, yes	2.99	1.56-5.90	0.008
T2DM, yes	3.79	2.07-7.10	<0.001
Advanced fibrosis, stage F3-F4	16.50	8.63-33.3	<0.001
PNPLA3, number of I148M alleles	1.19	0.62-3.19	0.41
TM6SF2, number of E167K alleles	3.07	1.48-6.59	0.002
MBOAT7, number of T alleles	1.74	1.13-2.71	0.011
MBOAT7, T alleles * Obesity, yes	-	-	0.035

OR: Odd Ratio; 95% c.i.: 95% confidence interval.

The relationship between the number of risk alleles among *PNPLA3* I148M, *TM6SF2* E167K, and *MBOAT7* rs641738 T and HCC risk is presented in Figure 3. There was a significant association between the cumulative number of risk alleles and HCC (OR per allele 1.56, 95% c.i. 1.31-1.86; OR 9.25, 95% c.i. 3.83-22.8 between the extremes of the distribution; p<0.001). 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).

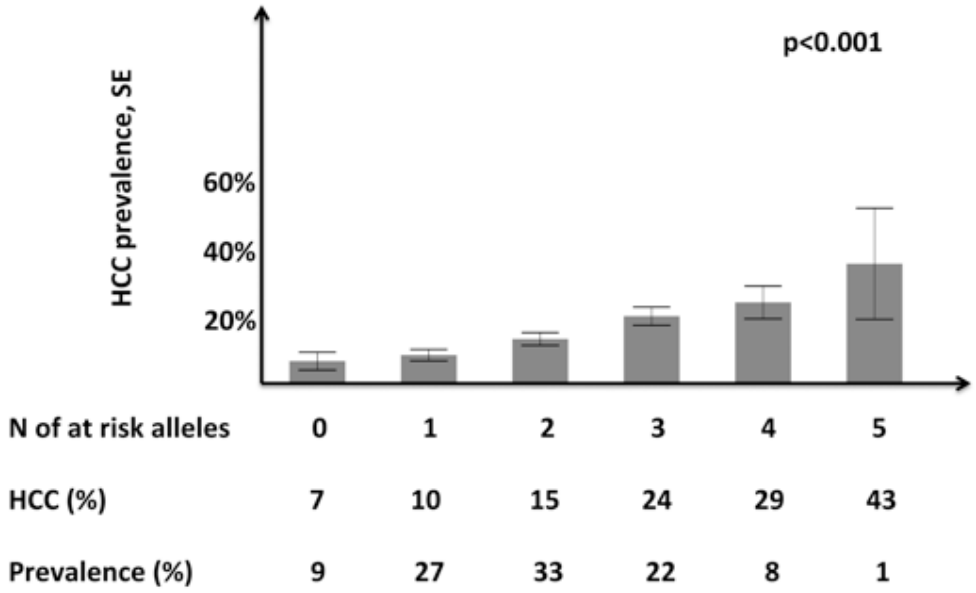

Figure 3. Risk of hepatocellular carcinoma according to the number of *PNPLA3* I148M, *TM6SF2* E167K, and *MBOAT7* rs641738 C>T risk variants in 765 Italian patients with NAFLD. HCC: hepatocellular carcinoma; SE: standard error. Comparisons were performed by a multivariate logistic regression setting 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.

A combined risk score considering acquired and common genetic risk factors was developed to predict HCC: 1 / (1 + e ^ - ((-12.588 + (0.162 * age) + (0.404 * Sex: 1 if male, -1 if female) + (0.259 * Obesity: 1 present, -1 absent) + (0.587 * T2DM: 1 present, -1 absent) + (1.299 * Severe Fibrosis: 1 yes, -1 no) + (0.442 * number of risk alleles))). The model had a 0.96 area under the receiving operating characteristic curve (AUROC). The optimal cutoff had 96%

sensitivity and 89% specificity for HCC in the present cohort (Figure 4). The corresponding AUROC of a model taking into consideration only clinical factors was slightly lower (0.94). In the subgroup of patients without severe fibrosis, the AUROC for clinical factors alone was 0.91, whereas the full model incorporating genetic risk factors maintained an AUROC of 0.96.

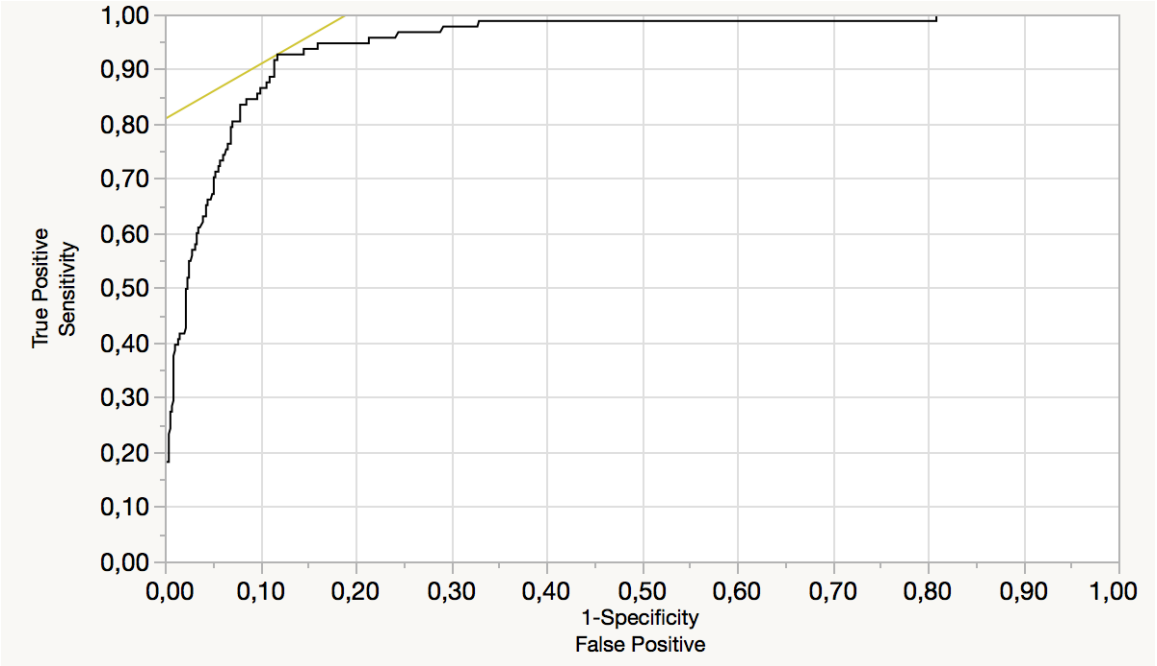

Figure 4. Receiving operating characteristic (ROC) curve for the combined risk score for predicting hepatocellular carcinoma in NAFLD patients. A combined risk score considering acquired and common genetic risk factors was developed to predict HCC: 1 / (1 + e ^ - ((-12.588 + (0.162 * age) + (0.404 * Sex: 1 if male, -1 if female) + (0.259 * Obesity: 1 present, -1 absent) + (0.587 * T2DM: 1 present, -1 absent) + (1.299 * Severe Fibrosis: 1 yes, -1 no) + (0.442 * number of risk alleles))). The model had a 0.96 area under the ROC curve. The optimal cutoff had 96% sensitivity and 89% specificity for HCC in the present cohort.

4.1.5 Relationship between MBOAT7/TMC4 locus variants and gene expression

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). Interestingly, in this subgroup rs8736 was more closely associated with NAFLD (p=0.048 vs. p=0.057) and MBOAT7 expression (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.

In an experiment conducted by a colleague in our lab, gene expression of MBOAT7 was evaluated in 47 patients from the Hepatology service characterized by more severe liver damage (Table 6), results are shown in Figure 5. The rs641738 T allele was associated with reduced hepatic MBOAT7 expression in patients without moderate-severe fibrosis (stage F0-F1; p=0.02), but not in those with significant fibrosis (stage F2-F4; p=0.1).

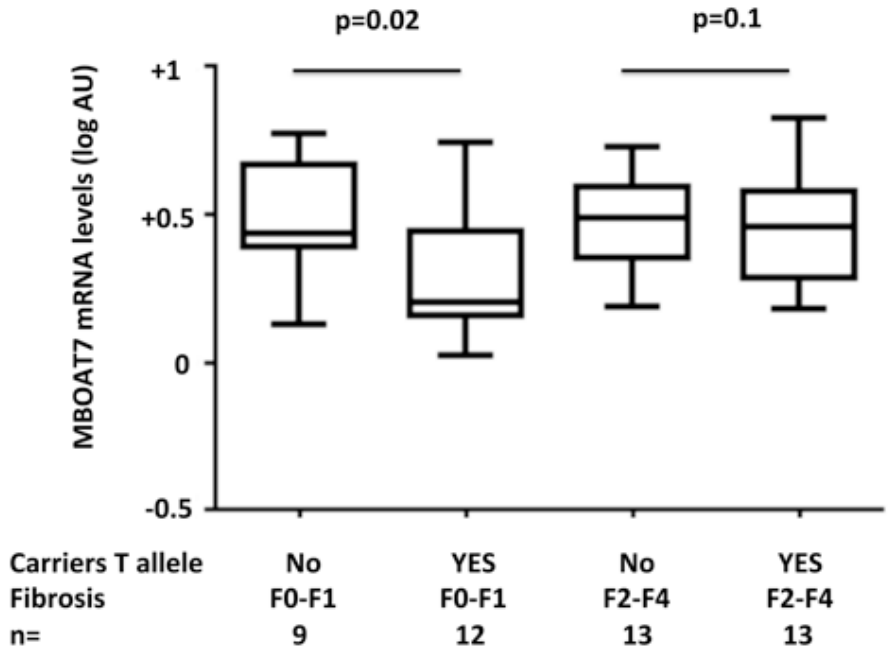

Figure 5. **Impact of the presence of rs643718 risk T allele on MBOAT7 expression**. MBOAT7 expression was shown as log mRNA levels in 47 patients with NAFLD from the Milan Hepatology service stratified by the presence of clinically significant hepatic fibrosis (stage F2-F4). Data were compared by Student's t-test.

4.1.6 Impact of rs641738 variant 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 non-cirrhotic patients with CHC and ALD. Results are presented in Table 14. We started evaluating a larger cohort of 1751 individuals with CHC and ALD, of whom 630 (36%) had cirrhosis and 363 (21%) had HCC (Table 2). In 25 of 363 cases (7%), HCC was detected in the absence of cirrhosis. The rs641738 T allele was associated with increased risk of HCC in patients without cirrhosis, independently of age, sex, and the etiology of liver disease (OR 1.93, 95% c.i. 1.07-3.58; p=0.035), but not in those with cirrhosis (p=NS).

We observed a similar trend for association of the T allele with non-cirrhotic HCC in patients with CHC and ALD analyzed separately (Table 15). The *PNPLA3* I148M variant was also associated with HCC development outside cirrhosis (Table 14; p=0.021).

Table 14. Independent predictors of hepatocellular carcinoma (HCC) in 1751 patients with chronic liver disease (994 with chronic hepatitis C and 757 with alcoholic liver disease) stratified by the presence of cirrhosis (Validation cohort).

	Cirrhosis NO				
	HCC				
	No (n=1,096)	Yes (n=25)	p value	Allelic OR, 95% c.i.*	p value*
PNPLA3 I148M	558/437/101 (51/40/9)	8/10/7 (32/40/28)	0.016	1.92, 1.07-3.45	0.021
TM6SF2 E167K	973/122/1 (89/11/0)	20/5 (80/20)	0.16	1.96, 0.61-5.27	0.16
MBOAT7/TMC4 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
	Cirrhosis YES				
	HCC				
	No (n=291)	Yes (n=338)	p value	Allelic OR, 95% c.i.*	p value*
PNPLA3 I148M	111/126/54 (38/43/19)	137/143/58 (41/42/17)	0.91	1.04, 0.81-1.34	0.88
TM6SF2 E167K	252/36/3 (87/12/1)	303/34/1 (89/10/1)	0.31	0.79, 0.47-1.32	0.32
MBOAT7/TMC4 rs641738 C>T	78/155/58 (28/51/21)	117/150/71 (35/44/21)	0.96	1.01, 0.78-1.30	0.96

( ): % 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.

Table 15. Independent predictors of hepatocellular carcinoma (HCC) in 994 with chronic hepatitis C (CHC) and 757 with alcoholic liver disease (ALD) stratified by the presence of cirrhosis.

	HCC			
Chronic hepatitis C: Cirrhosis NO				
	No (n=584)	Yes (n=13)	p value	p value*
PNPLA3 I148M	313/219/52 (54/37/9)	5/5/3 (38/39/24)	0.18	0.22
TM6SF2 E167K	538/46 (92/8)	9/4 (69/31)	0.032	0.008
MBOAT7/TMC4 rs641738 C>T	177/259/148 (30/45/25)	1/7/5 (8/54/38)	0.087	0.071
Chronic hepatitis C: Cirrhosis YES				
	No (n=160)	Yes (n=237)	p value	p value*
PNPLA3 I148M	74/65/21 (46/41/13)	113/97/27 (48/41/11)	0.74	0.76
TM6SF2 E167K	142/18 (89/11)	221/16 (93/7)	0.18	0.18
MBOAT7/TMC4 rs641738 C>T	45/84/31 (28/53/19)	84/102/51 (35/43/22)	0.96	0.94
Alcoholic liver disease: Cirrhosis NO				
	No (n=512)	Yes (n=12)	p value	p value*
PNPLA3 I148M	245/218/49 (48/43/9)	3/5/4 (25/42/33)	0.054	0.041
TM6SF2 E167K	435/76/1 (85/15/0)	11/1 (92/8)	0.58	0.41
MBOAT7/TMC4 rs641738 C>T	150/251/111 (29/49/22)	1/8/3 (8/67/25)	0.12	0.090
Alcoholic liver disease: Cirrhosis YES				
	No (n=131)	Yes (n=101)	p value	p value*
PNPLA3 I148M	37/61/33 (28/47/25)	24/46/31 (24/45/31)	0.35	0.33
TM6SF2 E167K	110/18/3 (84/14/2)	82/18/1 (81/18/1)	0.97	0.98
MBOAT7/TMC4 rs641738 C>T	33/71/27 (25/54/21)	33/48/20 (33/47/20)	0.68	0.59

( ): % values; * Adjusted for age, sex, and PNPLA3, TM6SF2 and MBOAT7/TMC4 genetic variants.

4.2 Telomerase reverse transcriptase rare germline mutations and hepatocellular carcinoma in patients with nonalcoholic fatty liver disease

4.2.1 Telomere length is reduced in peripheral blood leukocytes of NAFLD HCC patients

Telomere length in peripheral blood leukocytes decreased with liver disease progression from healthy controls to NAFLD-cirrhosis to NAFLD-HCC (p=0.0003; ANOVA), even after adjustment for sex and age (p=0.0002; Figure 6A). In particular, telomere length was lower in HCC patients vs. cirrhosis (median: 0.93 IQR: 0.66-1.24 vs. 1.12 IQR: 0.82-1.69; p= 0.014) and vs. healthy subjects (median: 1.38 IQR: 0.99-1.75; p=0.0001). When we added an additional independent cohort of 50 Italian NAFLD-HCC, we observed the same shortening (p for trend=0.0008), with NAFLD-HCC patients having shorter telomeres as compared to those with uncomplicated cirrhosis and to healthy individuals (median: 0.98 IQR: 0.77-1.52; p= 0.048 and p= 0.0006, respectively; Figure 6B).

In n=5 patients for whom intra-tumoral specimens were available, we investigated telomere length in the intra-tumoral tissue. As expected, we observed telomere elongation in HCC lesions compared to peripheral blood (p=0.05; Figure 6C). In these HCC specimens, telomeres elongation was not always due to somatic mutations in the *hTERT* promoter. Indeed, we analyzed the known hot spots of mutations at -124 and -146 from the *hTERT* origin and we found only one carrier of a variation in heterozygosity (-124 G>A).

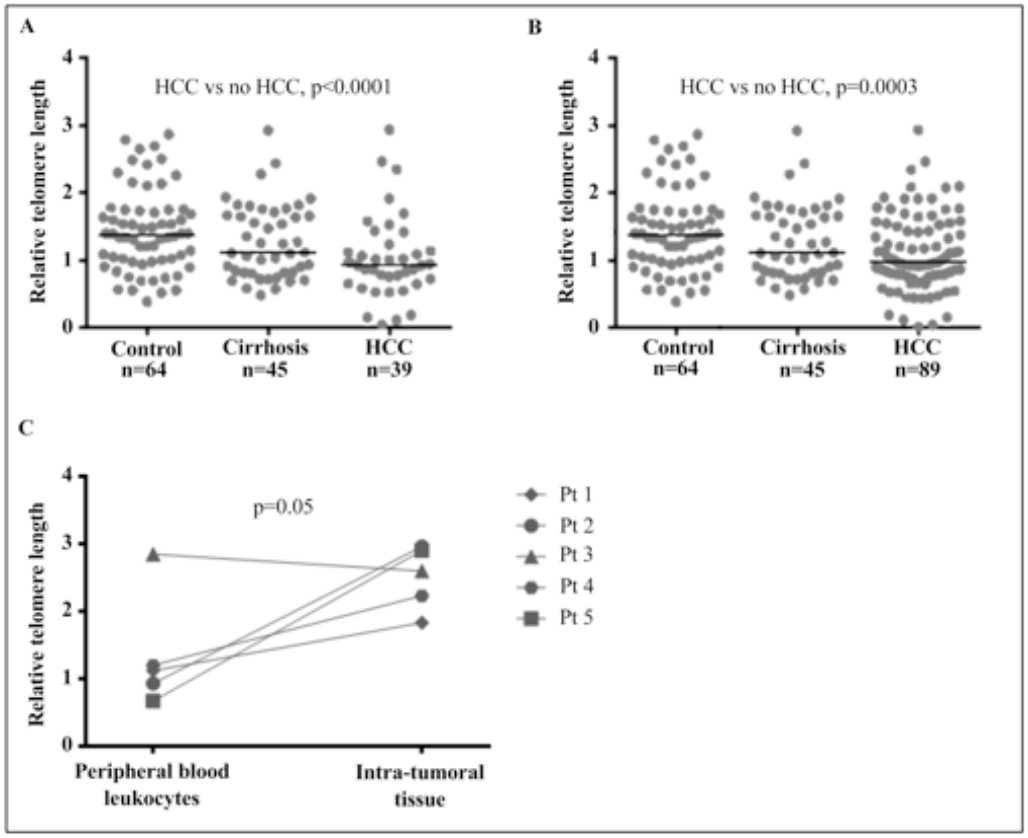

Figure 6. Telomere length is reduced in peripheral blood leukocytes of NAFLD HCC patients. Relative telomere length of patients included in the Discovery cohort (**panel A**) and in the Overall cohort (**panel B**) have been reported in the figure. Overall cohort includes all the Italian NAFLD-HCC patients (see table 3). For the different comparison p values were calculated by linear generalized model corrected for sex and age considering log-transformed data (table 16). Comparison of telomere length between peripheral blood leukocytes and intra-tumoral tissue has been shown in the figure (**panel C**). p value was calculated by Fisher paired T-test. For all the specimens, telomere length was evaluated by qPCR as ratio of the relative quantity of the telomere PCR product and the reference gene (36B4).

Table 16. Values of significance relative to telomere length comparisons.

	P value	
	Discovery	Validation
HCC vs Healthy	0.0001	0.0006
HCC vs Cirrhosis	0.0144	0.048
Cirrhosis vs Healthy	0.2364	-
Liver disease vs Healthy	0.0067	0.0049

4.2.2 Rare hTERT coding mutations are enriched in NAFLD HCC patients

We next evaluated the presence of rare *hTERT* coding variants in the discovery Italian cohort. We found three novel mutations and one mutation previously described in dbSNP with rare frequency. In particular, one patient had a heterozygous frameshift mutation, Glu113Arg_fs*79, in the second exon of the sequence. Three HCC patients carried missense mutations: Ala67Val, Pro193Leu and Glu668Asp. The homozygous Ala67Val and the heterozygous Pro193Leu are both located in the N-terminal (template binding domain). The Glu668Asp, found in heterozygosity, is located in the catalytic domain (Figure 7).

Enrichment of these rare coding mutations was detected in NAFLD-HCC patients rather than in subjects with cirrhosis or healthy controls (prevalence 10%; p=0.022 vs. cirrhosis, p=0.008 vs. healthy subjects; p=0.001 vs. controls overall). Burden test analysis confirmed the enrichment of rare *hTERT* coding mutations with HCC (p=0.020 vs healthy controls; p=0.045 vs cirrhosis; Table 17). When we extended the control group to the 503 healthy European subjects from the 1000Genomes Project database (http://www.internationalgenome.org/faq/can-i-get-phenotype-geneder-and-family-relationship-information-samples), in which only one individual of Italian origin was a heterozygous carrier of a rare coding mutation in *hTERT*, we confirmed a strong enrichment of missense variations in NAFLD-HCC (p=0.0001; Table 17).

Interestingly, considering the distribution of known rare coding mutations in *hTERT* (see Telomere Database - http://telomerase.asu.edu), we observed an enrichment of variations in the N-terminal of the gene (p=0.037) as compared to catalytic and C-terminal domains. Besides rare variants, common missense mutations in *hTERT* were found in our cohort both in HCC patients, cirrhotic patients and healthy controls, at the expected frequencies (Ala279Thr, His412Tyr and Ala1062Thr; Table 18).

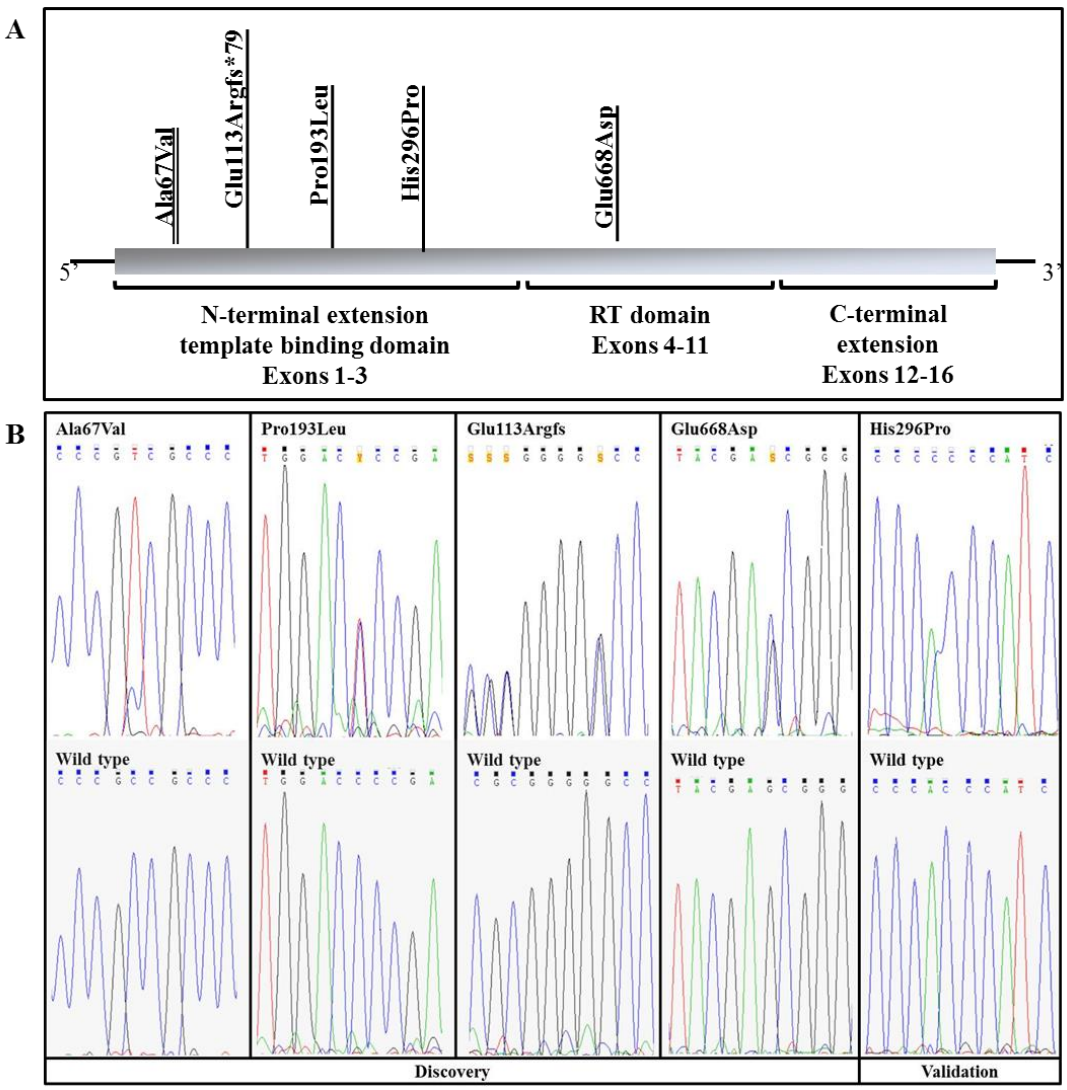

Figure 7. Rare coding mutations found in *hTERT* gene. Schematic representation of *hTERT* sequence showing the distribution of the rare coding mutations identified in the different domains (**panel A**). Electropherograms related to the identified rare coding mutations (**panel B**). The panel shows the mutated sequences in the upper part, while the corresponding normal sequences are represented below.

Table 17. Description of rare non-synonymous variations in *hTERT* gene found in subjects analyzed. The non-synonymous variations found in subjects sequenced are listed in the table, together with the only mutation found with frequency <0.001 in "European 1000G". All the mutations have been described with a frequency lower than <0.001 in both the ExAC NFE and the ESP EA. The table indicates the number of mutated alleles for each variation in the different groups analyzed and in the public database of "European 1000G".

					Discovery				Validation	
Mutation	SNP_ID (dbSNP147)	Chr:Position (GRCh37)	Ref/Alt	European 1000G (n.a.=1006)	Healthy controls (n.a.=128)	Cirrhosis (n.a.=90)	HCC (n.a.=80)	Discovery p value	PLC (n.a.=156)	Overall p value
								0.005° 0.0001^		0.0005
Ala67Val*	-	5:1294905	G/A	-	-	-	2		-	
Glu113Argfs	-	5:1294664	G/GG	-	-	-	1		-	
Pro193Leu	rs751762765	5:1294423	G/A	-	-	-	1		-	
His296Pro	rs778187343	5:1294114	T/G	-	-	-	-		1	
Glu668Asp	-	5:1279532	C/G	-	-	-	1		-	
Val897Met	rs559028617	5:1264558	C/T	1	-	-	-		-	

SNP (single nucleotide polymorphism); GRCh37 (human genome assembly release 37); Ref/Alt (Reference/Alternative allele); PLC (primary liver cancer); n.a. (number of alleles); *mutation in homozygosity; p values were calculated by Burden test considering HCCs vs Controls (healthy subjects and cirrhosis; °) and vs European 1000G (^); Overall p value was calculated by Burden test considering overall HCCs vs Controls (cirrhosis, healthy subjects and European 1000G). Variants were annotated according to hg19/GRCh37, using the *hTERT* RefSeq reference transcript NM_001193376.

Table 18. Description of common non-synonymous variations in *hTERT* gene found in subjects analyzed. The non-synonymous variations found in subjects sequenced with a frequency higher than 0.1% in both the ExAC NFE and the ESP EA are listed in the table. The table indicates the number of mutated alleles for each variation in the different groups and in the public database of "European 1000G".

Mutation	SNP_ID (dbSNP147)	Chr:Position (GRCh37)	Ref/Alt	European 1000G (n.a.=1006)	Discovery			Discovery p value	Validation	Overall p value
					Healthy controls (n.a.=128)	Cirrhosis (n.a.=90)	HCC (n.a.=80)		PLC (n.a.=156)	
Ala279Thr	rs61748181	5:1294166	C/T	36	4	3	2	0.613* 0.751§	6	0.925
His412Tyr	rs34094720	5:1294166	C/T	3	1	-	-	0.625* 0.544§	-	0.842
Ala1062Thr	rs35719940	5:1254594	G/A	22	1	-	-	0.544* 0.518§	3	0.518

SNP (single nucleotide polymorphism); GRCh37 (human genome assembly release 37); Ref/Alt (Reference/Alternative allele); PLC (primary liver cancer); n.a. (number of alleles); p values were calculated by Chi Square test considering HCCs vs Controls (healthy subjects and cirrhosis; §) and vs European 1000G (*); Overall p value was calculated by Chi Square test considering overall HCCs vs Controls (cirrhosis, healthy subjects and European 1000G. Variants were annotated according to hg19/GRCh37, using the *hTERT* RefSeq reference transcript NM_001193376.

4.2.3 hTERT mutations in a European cohort of NAFLD-PLC patients

We further sequenced *hTERT* in an independent cohort comprising 78 patients affected by PLC in NAFLD, 64% of Italian ancestry (Table 3). Here, we found one heterozygous carrier of a rare missense mutation (His296Pro; Figure 7) affected by intrahepatic cholangiocarcinoma. There was no significant enrichment in *hTERT* mutations between the validation cohort and controls of the Discovery cohort and healthy European subjects (p=0.42 and p=0.25 respectively). However, when we considered the overall cohort of cancer patients (n=118), as compared to all available controls (from discovery cohort and the 1000Genomes database, n=612), we confirmed a significant enrichment of rare germline *hTERT* variants in subjects affected by tumor developed in NAFLD (p=0.0005; Table 17), which remained significant after the exclusion of cholangiocarcinoma patients (p=0.003).

4.2.4 Clinical features of patients carrying rare hTERT mutations

The clinical features of patients positive and negative for the presence of the rare *hTERT* mutations are shown in Table 19. We found that the prevalence of female sex was higher in carriers of *hTERT* mutations (p=0.03). We did not observe any significant difference in the distribution of age, BMI, diabetes and the presence of cirrhosis between the two groups (p=NS). The genetic risk factor PNPLA3 I148M was equally distributed between the two groups (p=0.885). No significant differences were detected in peripheral blood telomere length between patients positive and those negative for *hTERT* mutations (not shown).

Table 19. Clinical features of 118 patients who developed primary liver cancer (PLC) in NAFLD stratified by carriage of rare *hTERT* mutations.

	PLC cohort overall		
hTERT mutation	yes	no	p
Age, years	70.4±10.8	67.0±8.8	0.52
Sex, F	3 (60)	21 (19)	0.03
BMI, Kg/m^2	28.8±5.0	29.9±5.3	0.65
T2DM, y	3 (60)	66 (58)	0.94
Fibrosis, F3-4	4 (80)	89 (79)	0.93
PNPLA3, I148M			0.90
I/I	1 (20)	30 (27)	
I/M	2 (40)	48 (42)	
M/M	2 (40)	35 (31)	

PLC: primary liver cancer; ( ): % values; y: yes; T2DM: type 2 diabetes mellitus; p: p value calculated as patients carriers of *hTERT* mutations versus not carriers.

4.2.5 Family study

To investigate the causal effect of the mutations found, we examined whether these co-segregated with liver disease or other pathological phenotypes. We were able to genotype and phenotype some relatives of the probands carrying the Glu668Asp (Family A, Figure 8) and the Glu113Argfs (Family B, Figure 8) mutations. In Family A, one son of the proband carried the mutation in heterozygosity and already showed traits of liver damage (NAFLD and increased liver enzymes) despite a relatively young age, while the nephew had a phenotype possibly related to mutations in *hTERT*, namely obstructive lung disease. The mother and the sister of the proband were affected by a phenotype likely related to telomere disease (cryptogenic cirrhosis and idiopathic pulmonary fibrosis, respectively). In Family B, the proband's brother, who also carries the Glu113Argfs mutation, had altered liver function tests. One of her younger daughters and a nephew were carriers of the mutations in heterozygosity, but with no phenotypic signs, possibly due to the young age.

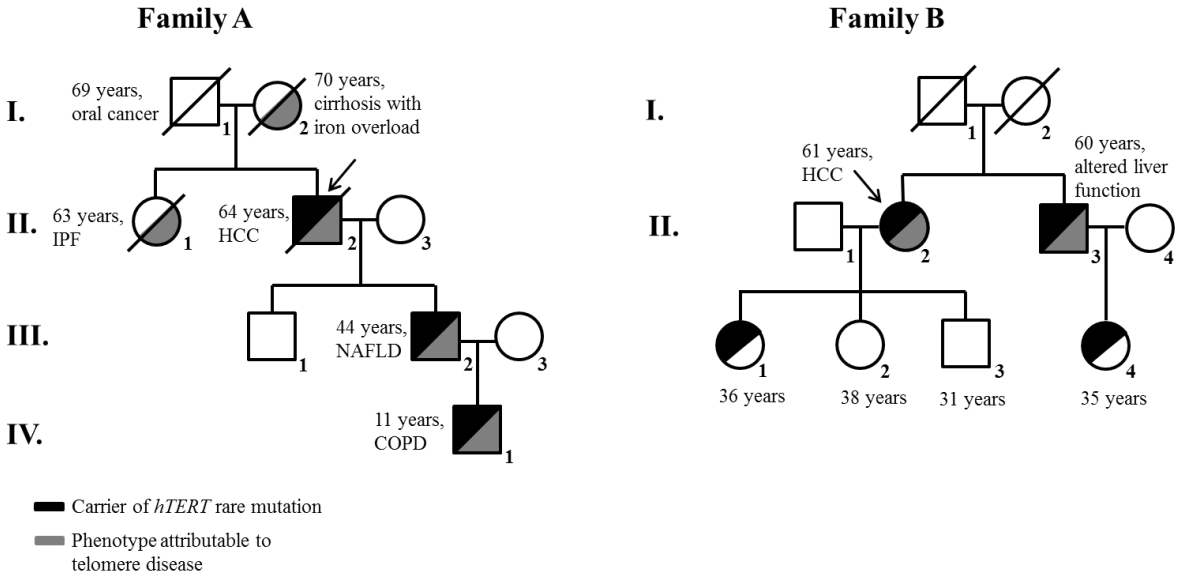

Figure 8. Family study. Family trees of patients carriers of Glu668Asp and Glu113Argfs mutations (Family A and Family B respectively).

4.2.6 Functional evaluation of hTERT mutations

In order to investigate whether the mutations found could affect the catalytic activity of hTERT, their effect was modelled *in silico* using predictive bioinformatics algorithms. In addition to the Glu113Argfs frameshift mutation, which causes a premature termination of protein synthesis and is frankly damaging, the Glu668Asp variant was also predicted to be deleterious by two out of three damage prediction tools (Table 20). Conversely, the amino acid substitutions Ala67Val, Pro193Leu and His296Pro were not predicted to influence the activity of the protein (Table 20), but we could not establish their impact on the DNA-binding of TERT.

Table 20. In silico prediction of functional impact of *hTERT* mutations according to bioinformatics algorithms.

	hTERT *variants*				
	Ala67Val	Glu113Argfs	Pro193Leu	His296Pro	Glu668Asp
Polyphen2	Benign	Damaging	Benign	Benign	Probably damaging
score	0.011		0.000	0.160	1.000
SIFT	Tolerated	Damaging	Tolerated	Tolerated	Damaging
score	0.16		0.38	0.21	0.04
Provean	Neutral	Deleterious	Neutral	Neutral	Neutral
score	-1.210		-1.08	-1.025	-1.724

To gain further insight into the consequences of *hTERT* mutations, the wild type and the missense mutants, which were generated by in situ mutagenesis or overlap extension PCR cloning, were transiently overexpressed in human HEK-293 cells by our collaborators from Professor Romeo's Lab in Sweden. It was observed a substantial reduction in the intracellular

protein synthesis of the Ala67Val and Glu668Asp mutations as compared to the wild type protein (Figure 9). There was virtually no change in the protein synthesis of the Pro193Leu and His296Pro mutants when compared to the wild type protein (Figure 9).

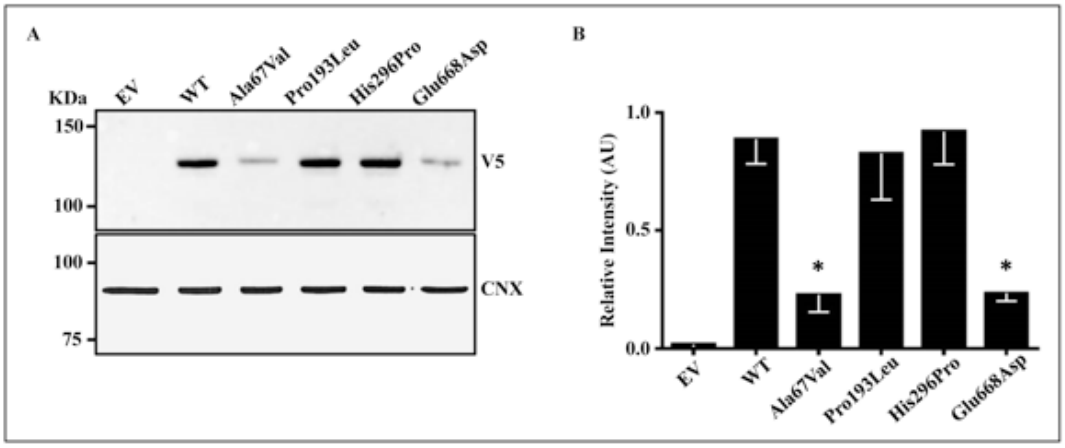

Figure 9. The hTERT Ala67Val and Glu668Asp variants result in a reduction of the intracellular protein levels. Effect of missense mutations on hTERT protein levels in HEK-293 cells **(panel A)**. Human TERT was transiently overexpressed in HEK-293 cells. HEK-293 cells were transfected with hTERT wild type and mutant forms cloned in pcDNA 3.1 vector for 48h; recombinant hTERT protein levels were examined on cell lysate by Western blotting analysis by using a V5 antibody. Empty vector was used as negative and calnexin was used as loading control. Quantification of Western blotting bands **(panel B)**. The graph bar represents intracellular protein levels expressed as mean and standard deviation (whisker) of three independent experiments. Protein levels was quantified by Image Lab Software (Bio-Rad).

5. Discussion

5.1 The *MBOAT7* rs641738 variant associates with NAFLD-HCC

5.1.1 Speculated mechanisms and clinical implications

In the first part of this study, we evaluated whether the rs641738 C>T *MBOAT7* locus sequence variant, associated with the development and progression of NAFLD [51], influences susceptibility to NAFLD-HCC in patients stratified by the severity of liver fibrosis. We found that each *MBOAT7* rs641738 T allele conferred an approximately 80% increased risk of HCC (65% when taking into consideration also the impact of 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 [51]. The rs641738 T allele is associated with reduced MBOAT7 expression and altered phosphatidyl-inositol plasma and hepatic composition [51, 62]. Of note, we detected a significant interaction between *MBOAT7* genotype and the presence of obesity in determining HCC risk, in line with the hypothesis that the phenotypic expression of the mutation is triggered by the increased flux of fatty acids to the liver. Therefore, the mechanism underpinning the association of the *MBOAT7* rs641738 sequence variant with NAFLD-HCC may be mediated by changes in phosphatidyl-inositol and arachidonic acid metabolism, favoring hepatocellular fat accumulation and the production of inflammatory mediators [51]. 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 the 3'-UTR of *MBOAT7*, which may be more closely related to the phenotype and are potentially involved in the regulation of MBOAT7 mRNA stability.

An important finding is that the effect size of rs641738 on HCC risk may have been 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. We also observed that

the rs641738 T allele is associated with reduced hepatic expression of MBOAT7 only in NAFLD patients without severe fibrosis. It could therefore be speculated that the MBOAT7 variant exerts 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 [51, 101]. In keeping with this hypothesis, we also showed that the rs641738 T allele was associated with HCC development in non-cirrhotic patients with ALD or CHC, although further studies are necessary to clarify whether genotyping of the MBOAT7 variant may be useful to stratify disease risk in patients cured from chronic hepatitis C. Consistently, the rs641738 T allele was also associated with development of early stages, but not severe fibrosis in patients at risk of NASH [51], and in a large cohort of CHC patient [102]. Therefore, it could be speculated that MBOAT7 variation 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, which however subsequently hampers the impact of this genetic variant.

5.1.2 Common genetic variants in PNPLA3, TM6SF2 and MBOAT7 are risk factors for HCC

In the NAFLD cohort, the overall impact of the *MBOAT7* rs641738 on HCC risk was similar to that of the I148M PNPLA3 variant. Indeed, *MBOAT7* and *PNPLA3* variants explained about 27% and 22%, respectively, of HCC phenotype variability. 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 [34, 39, 103]. 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 [41]. This difference may be due to lifestyle factors, and to the

higher prevalence of clinical cofactors (obesity, T2DM, and cirrhosis) as opposed to genetic risk variants (lower frequency of the I148M variant) in the UK cohort. Notwithstanding, a pooled estimate of the effect of the I148M variant in the UK and Italian cohort confirmed an increased HCC susceptibility independently of fibrosis. In addition, we also report for the first time an association between the TM6SF2 E167K variant and NAFLD-HCC, which was independent of severe fibrosis. This was not observed at univariate analysis due to an interaction of the TM6SF2 variant with clinical factors, and in the UK NAFLD cohort [47], whereas a predisposing effect on HCC was reported in a Italian cohort of patients with alcoholic cirrhosis [49]. Further studies are therefore required to confirm whether the E167K variant is an independent risk factor for HCC.

However, we found that, independently of any acquired factor, the number of genetic risk variants in *PNPLA3*, *TM6SF2* and *MBOAT7* was able to classify NAFLD patients in three groups with very different HCC risk: 9% in the 36% with 0-1 risk alleles, l9% in the 55% with 2-3 risk alleles, and 31% in those with more than 3 risk alleles.

5.1.3 Study' s limitations

Limitations include the cross-sectional retrospective nature of the study resulting, very similarly to what was observed in independent studies [41], in an uneven representation of clinical risk factors (age, sex, T2DM, severe fibrosis) between HCC cases and controls with histological NAFLD. However, the majority of NAFLD-HCC patients are still diagnosed incidentally outside regular follow-up [104], 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 the evaluation of genetic risk variants can be considered for implementation in clinical practice. Finally, these results may not be applicable to other ethnic groups.

5.2 *hTERT* rare germline mutations and telomere attrition are risk factor for NAFLD-HCC

5.2.1 Role of telomere in HCC pathogenesis

In the second part of the study we firstly assessed whether telomere attrition, a known feature of liver fibrosis, predisposes to NAFLD-HCC development.

We observed a progressive shortening of peripheral telomere length, from healthy controls to patients with cirrhosis related to NAFLD and in particular to those who developed HCC. In keeping with these observations, previous studies demonstrated that telomere attrition is involved in the progression of liver disease, as well as in other chronic degenerative conditions [65, 66, 70, 74, 105]. Indeed, shorter telomeres lead to the exhaustion of tissue staminal compartments, senescence, and fibrosis, which may involve the lung, liver and blood, and predispose to a wide spectrum of cancers by favoring genomic instability [106]. Not only genetic alterations in Telomerase complex genes, but also environmental risk factors for degenerative diseases, such as smoke and pollution, have been associated with peripheral blood telomere shortening, reflecting whole body exposure [105, 107]. It is therefore possible that shorter telomeres may reflect both genetic predisposition and environmental factors exposure in NAFLD-HCC patients.

Since reactivation of hTERT plays a key role in the immortalization of hepatocytes during malignant transformation [55, 73, 74], as expected we observed a re-elongation of telomeres within neoplastic lesions in a subgroup of patients. This was not frequently accounted for by *hTERT* promoter mutations, found in only one patient. It could be speculated that alternative

mechanisms of telomere elongation, such as insertional mutagenesis or duplication, or activation of the alternative lengthening of telomeres or the increase expression of other proteins of the telomerase complex may be involved [108-110].

5.2.2 hTERT rare germline mutations predispose to HCC

NAFLD and HCC have a strong heritable component and rare genetic mutations may play a role in triggering disease development [10, 30, 61]. Therefore, in this part of the study we evaluated whether mutations in *hTERT*, a key player in hepatic fibrosis progression and carcinogenesis [57, 59, 65, 111], are associated with NAFLD-HCC development.

We focused on rare germline mutations determining an alteration of protein sequence because these are more likely to alter protein function [112, 113]. Consistently, in the NAFLD-HCC discovery cohort, we observed a strong enrichment in rare and novel *hTERT* mutations, almost 76-fold as compared to the frequency of analogous mutations in local controls and healthy individuals of matched ethnicity in public databases. In further analysis, we considered as population controls only apparently healthy Europeans individuals included in the 1000Genomes database. These data are consistent with the hypothesis that rare *hTERT* mutations predispose to NAFLD-HCC.

The frequency of *hTERT* mutations was lower in a larger European validation cohort, but amongst NAFLD-PLC patients overall there still was a significant 31-fold enrichment in *hTERT* mutations. The higher prevalence of mutations in the discovery cohort may possibly be ascribed to the higher proportion of individuals without strong cofactors for progressive NAFLD such as obesity and male sex, so that genetic factors may have played a larger role. Interestingly, despite HCC more frequently developing in males, we observed a strikingly higher prevalence of *hTERT* mutations in female patients with NAFLD-PLC, so that the lower proportion of females in the validation cohort may have reduced the probability of finding mutations carriers. As one mutation carrier was actually diagnosed with intrahepatic

cholangiocarcinoma, we cannot exclude the possibility that *hTERT* genetic variations represent a common pathogenic risk factor for PLC.

On the other hand, common missense variants in hTERT, including Ala279Thr, His412Tyr and Ala1062Thr, previously reported to confer increased risk of telomere disease [114-118], were not differently represented between HCC cases and healthy controls. Even if these genetic variants may possibly confer subtle alterations in hTERT activity, we can reasonably exclude that they have a strong impact on telomere dysfunction; otherwise they would have undergone a strong negative selection (purification) during evolution. Our study was not sufficiently powered to detect a moderate effect on the risk of progressive NAFLD.

5.2.3 Effects of the rare germline mutations identified in hTERT

Several clues, besides the fact itself that are rare or novel mutations, suggest that the genetic variants identified may have played a role in the pathogenesis of NAFLD-HCC. The first one is that mutations tended to co-segregate within the N-terminal domain of *hTERT*, which is involved in telomere binding. Indeed, it is know that mutations in specific *hTERT* domains tend to determine specific pathological phenotypes [106]. Secondly, mutations co-segregated with liver disease and other phenotypes shared by telomere diseases in older individuals. As telomeropathies are age-dependent degenerative conditions, a longer follow-up will be necessary to better characterize the penetrance of these genetic variants in younger carriers. Furthermore, *in silico* analysis predicted functional consequences for Glu113Arg_fs, determining an early termination of the protein, and the Glu668Asp mutation in the catalytic domain. However, the ability of bioinformatics tools to predict the interaction of the other N-terminal domain mutations with telomeres and interacting proteins is quite limited.

To gain further insight into the functional consequences of *hTERT* mutations, we transiently overexpressed the wild type and the missense variants. We did not test *in vitro* the

Glu113Argfs*79 mutation because of the high likelihood to undergo mRNA decay due to the severe damage caused by the frameshift [119]. We observed that the Ala67Val and the Glu668Asp mutations resulted in a severe reduction in the intracellular protein levels as compared to the wild type, suggesting that the Ala67Val, carried in homozygosity by one NAFLD-HCC patient, may induce a reduction in telomerase activity due to altered protein translation or stability. On the other hand, the Pro193Leu and the His296Pro mutations in the N-terminal template binding domain, do not likely influence the catalytic activity of hTERT, but may result in a reduction of the binding ability of hTERT, inducing novel telomerase functions in tumorigenesis independently of hTERC [120]. For example, hTERT can act as a transcription factor in the Wnt-β-catenin signaling pathway, regulating the expression of pro-carcinogenic Wnt target genes [121, 122]. Our bioinformatics and *in vitro* approach was not suitable to test these hypotheses, therefore further studies are required to better investigate their possible causal role and mechanism in determining HCC predisposition.

6. Conclusion

Family history and genetic factors play an important roles in the pathogenesis of progressive NAFLD and of HCC [10, 30], however there is a lack of a common prognostic marker able to identify NAFLD patients at risk to progress towards cancer.

In this study, we demonstrated that common and rare genetic variants may have a role in influencing the pathogenesis of this rare disease characterized by increasing incidence in Western countries.

An important finding is that the *MBOAT7* rs641738 C>T polymorphism increased HCC risk in European NAFLD specifically in patients without severe fibrosis and it was associated with reduced MBOAT7 expression. This data was confirmed also in non-cirrhotic patients affected by other liver diseases (ALD and CHC). This genetic polymorphism may thus represent a useful biomarker to identify individuals at HCC risk in patients presenting without advanced liver damage, which is an important clinical problem in the management of patients with NAFLD.

Furthermore, the number of genetic risk variants in *PNPLA3*, *TM6SF2*, and *MBOAT7* carried by patients with NAFLD was associated with HCC risk independently of cofactors, with 13.4-fold higher risk in those carrying the maximum number of variants as compared to none. These data suggest that genetic variants predisposing to hepatic fat accumulation promote hepatic carcinogenesis. Indeed, hepatocellular fat accumulation seems to represent a key feature of hepatic carcinogenesis [123, 124]. Therefore, all in all they might represent useful biomarkers for risk stratification, performing better than the PNPLA3 I148M variant alone, which was proposed by the EASL-EASD-EASO NAFLD guidelines [41, 83]. However, in the present retrospective cohort genetic risk variants did not significantly improve the predictive accuracy of clinical factors. Large prospective studies are necessary to evaluate

whether combined evaluation of genetic and acquired risk factors may improve HCC risk stratification for patients with NAFLD and non-cirrhotic liver disease.

For what concern the second part of the study, our results suggest that telomere attrition is associated with NAFLD-HCC development, and that rare germline mutations in *hTERT* predispose to cirrhosis progression towards HCC in some familial cases, potentially assisting the identification of high-risk individuals that may warrant closer surveillance. Our hypothesis for the role of telomere and telomerase in HCC pathogenesis is the following one. Triggering factors, such as obesity and insulin resistance in the case of NAFLD, induce a condition of chronic hepatic damage and regeneration characterized by progressive hepatocytes telomere shortening and senescence. When hepatocytes reach senescence, liver regeneration decreases, but chronic damage remains. Concomitantly, other cell types, such as HSCs, become activated and form fibrotic tissue in area of hepatocytes loss. In this context, germline *hTERT* loss-of-function mutations accelerate telomere shortening favoring fibrosis development and thus creating a favorable microenvironment for cancer onset. Moreover telomeres attrition and germline *hTERT* loss-of-function mutations may exert a direct pro-carcinogenic effect by promoting genomic instability both inducing telomere shortening and impairing Telomerase activity in DNA repair and chromatin organization [72]. Within this context, the presence of heterozygous mutations does not prevent the reactivation of Telomerase wild type allele at later stages of carcinogenesis, which is necessary for the indefinite replication of the neoplastic clone (Figure 10)[55]. As many other proteins are included in Telomerase complex and participate to telomere elongation, we cannot exclude that mutations in other genes involved in telomere regulation play a role in telomere attrition in NAFLD-HCC. Indeed, telomere length in peripheral blood was shortened irrespective of the presence of *hTERT* mutations in NAFLD-HCC patients. Additional genetic factors should be investigated by next generation sequencing. Indeed, our approach suggests that rare germline mutations altering

the sequence of protein known to be involved in the pathogenesis of NAFLD and chronic liver disease may play an important role in NAFLD-HCC predisposition.

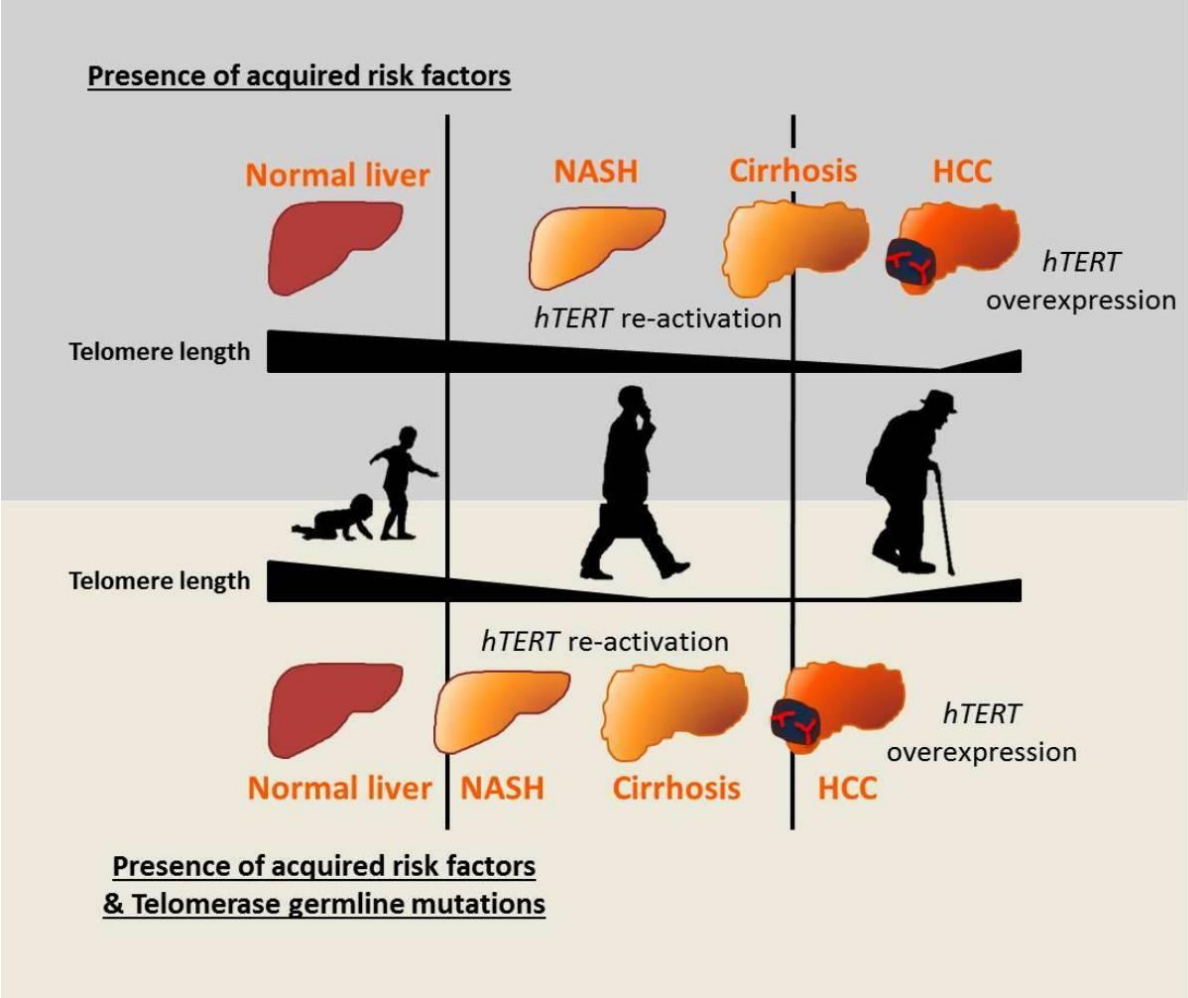

Figure 10. Hypothesis for telomeres role in pathogenesis of NAFLD progression toward cirrhosis and HCC.

In conclusion, these findings suggest that common genetic variants, such as *MBOAT7* rs641738 and the other ones associated with hepatic fat accumulation, are biomarkers able to improve HCC screening by assisting in the identification of subjects at risk even in the absence of severe fibrosis, most likely in different etiologies of liver disease. Moreover, these data suggest that rare mutations, such as those in *hTERT* gene, and strong heritable tracts, such as telomere length, influence HCC development and may play an important role for risk stratification particularly in family members of affected individuals. All in all, these results

pave the way for new potential therapeutic targets. At this purpose in future, functional studies are needed in order to identify the precise molecular mechanisms that underlie the pathogenesis of NAFLD-HCC and to understand the exact role and the interplay of the several common and rare genetic variants that contribute to the onset and the progression of this rare and complex disease.

7. References

1. Dongiovanni, P., S. Romeo, and L. Valenti, *Hepatocellular carcinoma in nonalcoholic fatty liver: role of environmental and genetic factors.* World J Gastroenterol, 2014. **20**(36): p. 12945-55.
2. Bugianesi, E., et al., *Expanding the natural history of nonalcoholic steatohepatitis: from cryptogenic cirrhosis to hepatocellular carcinoma.* Gastroenterology, 2002. **123**(1): p. 134-40.
3. Baffy, G., E.M. Brunt, and S.H. Caldwell, *Hepatocellular carcinoma in non-alcoholic fatty liver disease: an emerging menace.* J Hepatol, 2012. **56**(6): p. 1384-91.
4. Dyson, J., et al., *Hepatocellular cancer: the impact of obesity, type 2 diabetes and a multidisciplinary team.* J Hepatol, 2014. **60**(1): p. 110-7.
5. Younossi, Z. and L. Henry, *Contribution of Alcoholic and Nonalcoholic Fatty Liver Disease to the Burden of Liver-Related Morbidity and Mortality.* Gastroenterology, 2016. **150**(8): p. 1778-85.
6. Torres, D.M. and S.A. Harrison, *Nonalcoholic steatohepatitis and noncirrhotic hepatocellular carcinoma: fertile soil.* Semin Liver Dis, 2012. **32**(1): p. 30-8.
7. Kawada, N., et al., *Hepatocellular carcinoma arising from non-cirrhotic nonalcoholic steatohepatitis.* J Gastroenterol, 2009. **44**(12): p. 1190-4.
8. Ertle, J., et al., *Non-alcoholic fatty liver disease progresses to hepatocellular carcinoma in the absence of apparent cirrhosis.* Int J Cancer, 2011. **128**(10): p. 2436-43.
9. Chagas, A.L., et al., *Does hepatocellular carcinoma in non-alcoholic steatohepatitis exist in cirrhotic and non-cirrhotic patients?* Braz J Med Biol Res, 2009. **42**(10): p. 958-62.
10. Dongiovanni, P. and L. Valenti, *Genetics of nonalcoholic fatty liver disease.* Metabolism, 2015.
11. Paradis, V., et al., *Hepatocellular carcinomas in patients with metabolic syndrome often develop without significant liver fibrosis: a pathological analysis.* Hepatology, 2009. **49**(3): p. 851-9.
12. El-Serag, H.B. and K.L. Rudolph, *Hepatocellular carcinoma: epidemiology and molecular carcinogenesis.* Gastroenterology, 2007. **132**(7): p. 2557-76.

13. Poonawala, A., S.P. Nair, and P.J. Thuluvath, *Prevalence of obesity and diabetes in patients with cryptogenic cirrhosis: a case-control study.* Hepatology, 2000. **32**(4 Pt 1): p. 689-92.
14. Calle, E.E., et al., *Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults.* N Engl J Med, 2003. **348**(17): p. 1625-38.
15. Larsson, S.C. and A. Wolk, *Overweight, obesity and risk of liver cancer: a meta-analysis of cohort studies.* Br J Cancer, 2007. **97**(7): p. 1005-8.
16. Davila, J.A., et al., *Diabetes increases the risk of hepatocellular carcinoma in the United States: a population based case control study.* Gut, 2005. **54**(4): p. 533-9.
17. Calle, E.E. and R. Kaaks, *Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms.* Nat Rev Cancer, 2004. **4**(8): p. 579-91.
18. Starley, B.Q., C.J. Calcagno, and S.A. Harrison, *Nonalcoholic fatty liver disease and hepatocellular carcinoma: a weighty connection.* Hepatology, 2010. **51**(5): p. 1820-32.
19. Stickel, F. and C. Hellerbrand, *Non-alcoholic fatty liver disease as a risk factor for hepatocellular carcinoma: mechanisms and implications.* Gut, 2010. **59**(10): p. 1303-7.
20. Park, E.J., et al., *Dietary and genetic obesity promote liver inflammation and tumorigenesis by enhancing IL-6 and TNF expression.* Cell, 2010. **140**(2): p. 197-208.
21. Marra, F. and C. Bertolani, *Adipokines in liver diseases.* Hepatology, 2009. **50**(3): p. 957-69.
22. Ikejima, K., et al., *Role of adipocytokines in hepatic fibrogenesis.* J Gastroenterol Hepatol, 2007. **22 Suppl 1**: p. S87-92.
23. Menendez, J.A. and R. Lupu, *Fatty acid synthase and the lipogenic phenotype in cancer pathogenesis.* Nat Rev Cancer, 2007. **7**(10): p. 763-77.
24. Yamashita, T., et al., *Activation of lipogenic pathway correlates with cell proliferation and poor prognosis in hepatocellular carcinoma.* J Hepatol, 2009. **50**(1): p. 100-10.
25. Unger, R.H., et al., *Lipid homeostasis, lipotoxicity and the metabolic syndrome.* Biochim Biophys Acta, 2010. **1801**(3): p. 209-14.
26. Wei, Y., et al., *Saturated fatty acids induce endoplasmic reticulum stress and apoptosis independently of ceramide in liver cells.* Am J Physiol Endocrinol Metab, 2006. **291**(2): p. E275-81.
27. Gallagher, E.J. and D. LeRoith, *Minireview: IGF, Insulin, and Cancer.* Endocrinology, 2011. **152**(7): p. 2546-51.

28. Friedman, S.L., *Mechanisms of hepatic fibrogenesis.* Gastroenterology, 2008. **134**(6): p. 1655-69.
29. Schnabl, B., et al., *Replicative senescence of activated human hepatic stellate cells is accompanied by a pronounced inflammatory but less fibrogenic phenotype.* Hepatology, 2003. **37**(3): p. 653-64.
30. Turati, F., et al., *Family history of liver cancer and hepatocellular carcinoma.* Hepatology, 2012. **55**(5): p. 1416-25.
31. Romeo, S., et al., *Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease.* Nat Genet, 2008. **40**: p. 1461-1465.
32. Valenti, L., et al., *Homozygosity for the PNPLA3 / adiponutrin I148M polymorphism influences liver fibrosis in patients with nonalcoholic fatty liver disease.* Hepatology, 2010. **51**: p. 1209-1217.
33. Sookoian, S. and C.J. Pirola, *Meta-analysis of the influence of I148M variant of patatin-like phospholipase domain containing 3 gene (PNPLA3) on the susceptibility and histological severity of nonalcoholic fatty liver disease.* Hepatology, 2011. **53**(6): p. 1883-94.
34. Dongiovanni, P., et al., *PNPLA3 I148M polymorphism and progressive liver disease.* World J Gastroenterol, 2013. **19**(41): p. 6969-78.
35. Yuan, X., et al., *Population-based genome-wide association studies reveal six loci influencing plasma levels of liver enzymes.* Am J Hum Genet, 2008. **83**(4): p. 520-8.
36. Donati, B., et al., *The rs2294918 E434K Variant Modulates Patatin-Like Phospholipase Domain-Containing 3 Expression and Liver Damage.* Hepatology, 2016. **63**(3): p. 787-798.
37. Ruhanen, H., et al., *PNPLA3 mediates hepatocyte triacylglycerol remodeling.* J Lipid Res, 2014. **55**(4): p. 739-46.
38. Smagris, E., et al., *Pnpla3I148M knockin mice accumulate PNPLA3 on lipid droplets and develop hepatic steatosis.* Hepatology, 2015. **61**: p. 108-18.
39. Pirazzi, C., et al., *PNPLA3 has retinyl-palmitate lipase activity in human hepatic stellate cells.* Hum Mol Genet, 2014. **23**(15): p. 4077-85.
40. Mondul, A., et al., *PNPLA3 I148M Variant Influences Circulating Retinol in Adults with Nonalcoholic Fatty Liver Disease or Obesity.* J Nutr, 2015. **145**(8): p. 1687-91.
41. 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, 2014. **61**(1): p. 75-81.

42. 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, 2014. **59**(6): p. 2170-7.
43. 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, 2013. **61**(1): p. 75-81.
44. Valenti, L., et al., *PNPLA3 I148M variant and hepatocellular carcinoma: a common genetic variant for a rare disease.* Dig Liver Dis, 2013. **45**(8): p. 619-24.
45. Anstee, Q.M., et al., *Reply to: HCC and liver disease risk in homozygous PNPLA3 p.I148M carriers approach monogenic inheritance.* J Hepatol, 2015. **62**(4): p. 982-3.
46. Kozlitina, J., et al., *Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease.* Nat Genet, 2014. **46**(4): p. 352-6.
47. Liu, Y.L., et al., *TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease.* Nat Commun, 2014. **5**: p. 4309.
48. Dongiovanni, P., et al., *Transmembrane 6 superfamily member 2 gene variant disentangles nonalcoholic steatohepatitis from cardiovascular disease.* Hepatology, 2015. **61**(2): p. 506-14.
49. Falleti, E., et al., *PNPLA3 rs738409 and TM6SF2 rs58542926 variants increase the risk of hepatocellular carcinoma in alcoholic cirrhosis.* Dig Liver Dis, 2016. **48**(1): p. 69-75.
50. Buch, S., F. Stickel, and E. Trepo, *A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis.* 2015. **47**(12): p. 1443-8.
51. Mancina, R.M., et al., *The MBOAT7-TMC4 Variant rs641738 Increases Risk of Nonalcoholic Fatty Liver Disease in Individuals of European Descent.* Gastroenterology, 2016. **150**(5): p. 1219-1230.e6.
52. Aravinthan, A., et al., *Hepatocyte senescence predicts progression in non-alcohol-related fatty liver disease.* J Hepatol, 2013. **58**(3): p. 549-56.
53. Aravinthan, A., et al., *Gene polymorphisms of cellular senescence marker p21 and disease progression in non-alcohol-related fatty liver disease.* Cell Cycle, 2014. **13**(9): p. 1489-94.
54. Korthagen, N.M., et al., *Association between variations in cell cycle genes and idiopathic pulmonary fibrosis.* PLoS One, 2012. **7**(1): p. e30442.

55. Donati, B. and L. Valenti, *Telomeres, NAFLD and Chronic Liver Disease.* Int J Mol Sci, 2016. **17**(3): p. 383.
56. Calado, R.T., et al., *A spectrum of severe familial liver disorders associate with telomerase mutations.* PLoS One, 2009. **4**(11): p. e7926.
57. Rudolph, K.L., et al., *Inhibition of experimental liver cirrhosis in mice by telomerase gene delivery.* Science, 2000. **287**(5456): p. 1253-8.
58. Calado, R.T., et al., *Constitutional telomerase mutations are genetic risk factors for cirrhosis.* Hepatology, 2011. **53**(5): p. 1600-7.
59. Hartmann, D., et al., *Telomerase gene mutations are associated with cirrhosis formation.* Hepatology, 2011. **53**(5): p. 1608-17.
60. Valenti, L., et al., *Liver transplantation for hepatocellular carcinoma in a patient with a novel telomerase mutation and steatosis.* J Hepatol, 2013. **58**(2): p. 399-401.
61. Di Filippo, M., et al., *Homozygous MTTP and APOB mutations may lead to hepatic steatosis and fibrosis despite metabolic differences in congenital hypocholesterolemia.* J Hepatol, 2014. **61**(4): p. 891-902.
62. 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. **65**(6): p. 1263-1265.
63. Blackburn, E.H., *Structure and function of telomeres.* Nature, 1991. **350**(6319): p. 569-73.
64. Chaiteerakij, R. and L.R. Roberts, *Telomerase mutation: a genetic risk factor for cirrhosis.* Hepatology, 2011. **53**(5): p. 1430-2.
65. Kitada, T., et al., *Telomere shortening in chronic liver diseases.* Biochem Biophys Res Commun, 1995. **211**(1): p. 33-9.
66. Urabe, Y., et al., *Telomere length in human liver diseases.* Liver, 1996. **16**(5): p. 293-7.
67. Aikata, H., et al., *Telomere reduction in human liver tissues with age and chronic inflammation.* Exp Cell Res, 2000. **256**(2): p. 578-82.
68. Wiemann, S.U., et al., *Hepatocyte telomere shortening and senescence are general markers of human liver cirrhosis.* Faseb j, 2002. **16**(9): p. 935-42.
69. Satyanarayana, A., et al., *Telomere shortening impairs organ regeneration by inhibiting cell cycle re-entry of a subpopulation of cells.* EMBO J, 2003. **22**(15): p. 4003-13.

70. Plentz, R.R., et al., *Telomere shortening correlates with increasing aneuploidy of chromosome 8 in human hepatocellular carcinoma.* Hepatology, 2005. **42**(3): p. 522-6.
71. Farazi, P.A., et al., *Differential impact of telomere dysfunction on initiation and progression of hepatocellular carcinoma.* Cancer Res, 2003. **63**(16): p. 5021-7.
72. Masutomi, K., et al., *The telomerase reverse transcriptase regulates chromatin state and DNA damage responses.* Proc Natl Acad Sci U S A, 2005. **102**(23): p. 8222-7.
73. Oh, B.K., et al., *High telomerase activity and long telomeres in advanced hepatocellular carcinomas with poor prognosis.* Lab Invest, 2008. **88**(2): p. 144-52.
74. Oh, B.K., et al., *Telomere shortening and telomerase reactivation in dysplastic nodules of human hepatocarcinogenesis.* J Hepatol, 2003. **39**(5): p. 786-92.
75. Ju, Z. and K.L. Rudolph, *Telomeres and telomerase in cancer stem cells.* Eur J Cancer, 2006. **42**(9): p. 1197-203.
76. Cevik, D., G. Yildiz, and M. Ozturk, *Common telomerase reverse transcriptase promoter mutations in hepatocellular carcinomas from different geographical locations.* World J Gastroenterol, 2015. **21**(1): p. 311-7.
77. Nault, J.C., et al., *High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions.* Nat Commun, 2013. **4**: p. 2218.
78. Horn, S., et al., *TERT promoter mutations in familial and sporadic melanoma.* Science, 2013. **339**(6122): p. 959-61.
79. Huang, F.W., et al., *Highly recurrent TERT promoter mutations in human melanoma.* Science, 2013. **339**(6122): p. 957-9.
80. Armanios, M.Y., et al., *Telomerase Mutations in Families with Idiopathic Pulmonary Fibrosis.* New England Journal of Medicine, 2007. **356**(13): p. 1317-1326.
81. Pinzani, M., *Pathophysiology of Liver Fibrosis.* Dig Dis, 2015. **33**(4): p. 492-7.
82. Poynard, T., et al., *A comparison of fibrosis progression in chronic liver diseases.* J Hepatol, 2003. **38**(3): p. 257-65.
83. European Association For The Study Of The, L., R. European Organisation For, and C. Treatment Of, *EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma.* J Hepatol, 2012. **56**(4): p. 908-43.
84. Dongiovanni, P., et al., *Statin use and nonalcoholic steatohepatitis in at risk individuals.* J Hepatol, 2015. **63**: p. 705-12.
85. Kleiner, D.E., et al., *Design and validation of a histological scoring system for nonalcoholic fatty liver disease.* Hepatology, 2005. **41**(6): p. 1313-21.

86. Angulo, P., et al., *The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD.* Hepatology, 2007. **45**(4): p. 846-54.
87. Valenti, L., et al., *Patatin-Like phospholipase domain-containing 3 I148M polymorphism, steatosis, and liver damage in chronic hepatitis C.* Hepatology, 2011. **53**(3): p. 791-9.
88. Milano, M., et al., *Transmembrane 6 superfamily member 2 gene E167K variant impacts on steatosis and liver damage in chronic hepatitis C patients.* Hepatology, 2015. **62**(1): p. 111-7.
89. Valenti, L., et al., *PNPLA3 I148M polymorphism, clinical presentation, and survival in patients with hepatocellular carcinoma.* PLoS One, 2013. **8**(10): p. e75982.
90. 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, 2014. **34**: p. 514-20.
91. Valenti, L., et al., *The I148M PNPLA3 polymorphism influences serum adiponectin in patients with fatty liver and healthy controls.* BMC Gastroenterol, 2012. **12**: p. 111.
92. Chen, Y.L., et al., *TERT promoter mutation in resectable hepatocellular carcinomas: a strong association with hepatitis C infection and absence of hepatitis B infection.* Int J Surg, 2014. **12**(7): p. 659-65.
93. Cawthon, R.M., *Telomere length measurement by a novel monochrome multiplex quantitative PCR method.* Nucleic Acids Res, 2009. **37**(3): p. e21.
94. Brouilette, S.W., et al., *Telomere length, risk of coronary heart disease, and statin treatment in the West of Scotland Primary Prevention Study: a nested case-control study.* Lancet, 2007. **369**(9556): p. 107-14.
95. Winkler, T., et al., *Defective telomere elongation and hematopoiesis from telomerase-mutant aplastic anemia iPSCs.* J Clin Invest, 2013. **123**(5): p. 1952-63.
96. Townsley, D.M., et al., *Danazol Treatment for Telomere Diseases.* N Engl J Med, 2016. **374**(20): p. 1922-31.
97. Xu, L., et al., *Human hepatic stellate cell lines, LX-1 and LX-2: new tools for analysis of hepatic fibrosis.* Gut, 2005. **54**(1): p. 142-51.
98. Stickel, F., et al., *Genetic variation in the PNPLA3 gene is associated with alcoholic liver injury in caucasians.* Hepatology, 2011. **53**(1): p. 86-95.
99. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.

100. Rametta, R., et al., *Hepatic steatosis and PNPLA3 I148M variant are associated with serum Fetuin-A independently of insulin resistance.* Eur J Clin Invest, 2014. **44**(7): p. 627-33.
101. Gijon, M.A., et al., *Lysophospholipid acyltransferases and arachidonate recycling in human neutrophils.* J Biol Chem, 2008. **283**(44): p. 30235-45.
102. Thabet, K., et al., *MBOAT7 rs641738 increases risk of liver inflammation and transition to fibrosis in chronic hepatitis C.* Nat Commun, 2016. **7**: p. 12757.
103. Pingitore, P., et al., *PNPLA3 overexpression results in reduction of proteins predisposing to fibrosis.* Hum Mol Genet, 2016.
104. Piscaglia, F., et al., *Clinical patterns of hepatocellular carcinoma in nonalcoholic fatty liver disease: A multicenter prospective study.* Hepatology, 2016. **63**(3): p. 827-38.
105. Birch, J., et al., *DNA damage response at telomeres contributes to lung aging and chronic obstructive pulmonary disease.* Am J Physiol Lung Cell Mol Physiol, 2015. **309**(10): p. L1124-37.
106. Garcia, C.K., W.E. Wright, and J.W. Shay, *Human diseases of telomerase dysfunction: insights into tissue aging.* Nucleic Acids Res, 2007. **35**(22): p. 7406-16.
107. Guzzardi, M.A., et al., *Exposure to Persistent Organic Pollutants Predicts Telomere Length in Older Age: Results from the Helsinki Birth Cohort Study.* Aging Dis, 2016. **7**(5): p. 540-552.
108. Liu, B., et al., *Dyskerin overexpression in human hepatocellular carcinoma is associated with advanced clinical stage and poor patient prognosis.* PLoS One, 2012. **7**(8): p. e43147.
109. Kim, H., et al., *Telomere length, TERT and shelterin complex proteins in hepatocellular carcinomas expressing "stemness"-related markers.* J Hepatol, 2013. **59**(4): p. 746-52.
110. Heaphy, C.M., et al., *Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes.* Am J Pathol, 2011. **179**(4): p. 1608-15.
111. Nault, J.C. and J. Zucman-Rossi, *TERT promoter mutations in primary liver tumors.* Clin Res Hepatol Gastroenterol, 2016. **40**(1): p. 9-14.
112. Lek, M., et al., *Analysis of protein-coding genetic variation in 60,706 humans.* Nature, 2016. **536**(7616): p. 285-91.
113. Kryukov, G.V., L.A. Pennacchio, and S.R. Sunyaev, *Most rare missense alleles are deleterious in humans: implications for complex disease and association studies.* Am J Hum Genet, 2007. **80**(4): p. 727-39.

114. Alder, J.K., et al., *Short telomeres are a risk factor for idiopathic pulmonary fibrosis.* Proc Natl Acad Sci U S A, 2008. **105**(35): p. 13051-6.
115. Du, H.Y., et al., *TERC and TERT gene mutations in patients with bone marrow failure and the significance of telomere length measurements.* Blood, 2009. **113**(2): p. 309-16.
116. Calado, R.T., et al., *Constitutional hypomorphic telomerase mutations in patients with acute myeloid leukemia.* Proc Natl Acad Sci U S A, 2009. **106**(4): p. 1187-92.
117. Vulliamy, T.J., et al., *Mutations in dyskeratosis congenita: their impact on telomere length and the diversity of clinical presentation.* Blood, 2006. **107**(7): p. 2680-5.
118. Yamaguchi, H., et al., *Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia.* N Engl J Med, 2005. **352**(14): p. 1413-24.
119. Chang, Y.F., J.S. Imam, and M.F. Wilkinson, *The nonsense-mediated decay RNA surveillance pathway.* Annu Rev Biochem, 2007. **76**: p. 51-74.
120. Chiodi, I. and C. Mondello, *Telomere-independent functions of telomerase in nuclei, cytoplasm, and mitochondria.* Front Oncol, 2012. **2**: p. 133.
121. Park, J.I., et al., *Telomerase modulates Wnt signalling by association with target gene chromatin.* Nature, 2009. **460**(7251): p. 66-72.
122. Zhang, Y., et al., *Human telomerase reverse transcriptase (hTERT) is a novel target of the Wnt/beta-catenin pathway in human cancer*. J Biol Chem, 2012. **287**(39): p. 32494-511.
123. Bjornson, E., et al., *Stratification of Hepatocellular Carcinoma Patients Based on Acetate Utilization.* Cell Rep, 2015. **13**(9): p. 2014-26.
124. Li, L., et al., *Differential requirement for de novo lipogenesis in cholangiocarcinoma and hepatocellular carcinoma of mice and humans.* Hepatology, 2016. **63**(6): p. 1900-13.