UNIVERSITÀ DEGLI STUDI DI MILANO

CORSO DI DOTTORATO
Medicina Traslazionale

DIPARTIMENTO DI AFFERENZA DEL CORSO
Dipartimento di Fisiopatologia Medico-Chirurgica e dei Trapianti

Genome-wide association study of non-invasive assessment of liver fibrosis in MASLD: evidence from an Italian population

Settore scientifico disciplinare MED/09

FRANCESCO MALVESTITI
MATRICOLA R13827

TUTOR:
Prof. LUCA VITTORIO CARLO VALENTI

CO-TUTOR:
Prof. VALENTINA VAIRA

COORDINATORE DEL DOTTORATO:
Prof. MARIO SALVATORE CLERICI

A.A. 2024-2025

Abstract

Metabolic Associated Steatotic Liver Disease (MASLD) is the most prevalent cause of chronic liver disease worldwide and a significant risk factor for cirrhosis and hepatocellular carcinoma (HCC). Despite its pathogenesis being determined by both environmental and genetic determinants, the genetic component of fibrosis progression is still only partially known, particularly in the Italian population. This Ph.D. project aimed to detect genetic variants implicated in advanced liver fibrosis in MASLD by performing a cross-sectional genome-wide association study (GWAS) using non-invasive fibrosis markers in a large biobank environment. The analysis was performed in the Milan Biobank, a multicenter cohort of 7,025 individuals encompassing healthy blood donors, metabolic dysfunction patients, and liver disease subjects across the MASLD spectrum. Clinical phenotyping was based on validated non-invasive indices: the Fibrosis-4 index (FIB-4) was used as the primary outcome, whereas the Fibrotic NASH Index (FNI) and serum liver enzymes, namely alanine transaminase (ALT), aspartate transaminase (AST), gamma-glutamyl transferase (GGT), were investigated as secondary phenotypes. The GWAS confirmed the strong effects of *PNPLA3* and *TM6SF2* on MASLD progression and fibrosis. Novel associations were identified at the *MRC1* and *MAGEE2* loci for FIB-4, and at *IFI27* and *DIAPH2* for FNI, implicating immune regulation and cytoskeletal remodeling as additional pathways driving fibrosis susceptibility. Replication analyses in the UK Biobank (UKBB, N>400,000) validated associations at *MRC1*, reinforcing the reliability and generalizability of these findings. Functional follow-up analyses, including transcriptomics and phenome-wide association study (PheWAS), highlighted downregulation of *MRC1* in patients with advanced fibrosis and hepatocellular ballooning, further linking this locus to inflammatory mechanisms of liver injury. Along with locus discovery, polygenic risk scores (PRSs) were derived from GWAS summary statistics and contrasted with pre-existing PRSs. These investigations revealed that PRSs including fibrosis-associated loci enhanced fibrosis, cirrhosis, and HCC

prediction, with performance differing between traits. Notably, partitioned PRSs (pPRSs) discriminated against liver-specific vs. systemic MASLD pathways, providing insight into disease heterogeneity and the basis for divergent progression patterns in patients. In summary, this project represents the first comprehensive GWAS of non-invasively assessed liver fibrosis in an Italian population. It identified both known and novel susceptibility loci, replicated significant findings in independent cohorts, and integrated functional genomics to support biological interpretation. Clinically, the findings underscore the potential of combining genetic information with non-invasive biomarkers to refine risk stratification, enable earlier detection of high-risk individuals, and inform the development of precision medicine strategies for MASLD. These findings substantially advance the understanding of MASLD genetics and set the stage for targeted therapeutic approaches.

Introduction

MASLD definition and epidemiology

Metabolic Associated Steatotic Liver Disease (MASLD), which was previously known as non-alcoholic fatty liver disease (NAFLD) before the introduction of the new nomenclature, represents the leading cause of liver disease and the most prevalent chronic liver condition worldwide. The introduction in this new and revised nomenclature of the term “steatotic liver disease” (SLD) was intended to enhances the understanding of the pathophysiology of the disease by incorporating its underlying metabolic components, in the attempt to eliminate the stigma associated with the word “fatty”; beyond that this revised nomenclature also improves patient classification, thus enabling an even more accurate risk prediction for both liver-related and cardiovascular complications[1,2]. MASLD is characterized by increased hepatic fat, or steatosis (>5.5% of liver mass) linked to insulin resistance (IR) and not related to harmful alcohol intake or other liver diseases. MASLD is associated with increased risk of liver with cardiometabolic clinical events. MASLD encompasses a spectrum of disease which includes steatosis, intended as the deposition of fat in the liver, steatohepatitis, which is intended as inflammation associated with that fat (Metabolic Dysfunction–Associated Steatohepatitis [MASH]), fibrosis, intended as deposition in the liver of connective tissue or extracellular matrix which gives rise to buildup of scar tissue, and cirrhosis, which represents an extensive fibrosis with nodular regeneration[3,4]. Cirrhosis can progress to hepatic decompensation and hepatocellular carcinoma (HCC) and was the ninth leading cause of death in the US in 2022[5]. More precisely, MASLD consists in a gradual and integrated transition between comorbidities such as type 2 diabetes (T2D), obesity, hypertriglyceridemia, hyperglycemia and low high-density lipoprotein (HDL) cholesterol. Thus, the term MASLD was recently introduced for a recapitulation of the crosstalk between intrahepatic fat accumulation and extra-hepatic metabolic manifestations.

Indeed, the term MASLD encompasses the mutual contribution of IR, gut dysbiosis and adipose tissue dysfunction in enduring one another and carrying forward a vicious cycle with systemic effects.

Importantly, simple steatosis can progress to more severe pathological phenotypes upon the occurrence of an inflammatory trigger. For this reason, about one third of MASLD patients progress to MASH, previously known as non-alcoholic steatohepatitis (NASH)[1,6], characterized by hepatocyte ballooning and cell death, infiltration of immune cells and deposition of fibrotic tissue. Up to 29% of individuals afflicted by MASH develops hepatic cirrhosis, which occurs when both architecture and vasculature of the liver are disrupted by fibrotic insults, leading to progressive loss of hepatic functionality and portal hypertension. In 4-27% of patients, chronic inflammation and fibrosis, combined with the development of somatic mutations and the self-renewal capabilities of liver cells, pave the way to the development of HCC[7,8], which is the first cause of primary liver cancer and the third cause of cancer-related mortality globally[9].

As the progression to MASH leads to hepatocyte death, the deposition of fibrotic tissue starts and consequently the amount of intracellular lipids is reduced. This event is even more conspicuous in cirrhotic livers where liver organization is shattered by fibrotic septa and hepatocytes are replaced with collagen fibers. Finally, with the development of HCC the liver parenchyma turns into a completely disorganized and hampered structure.

The prevalence of MASLD rose above 30%[10] in the last decade, increasing by 50% from 2009-2016 to 2016-2019. Prevalence is increasing jointly with the ongoing obesity pandemic and endorsed by the increasingly common hypercaloric diet and sedentary lifestyle. Furthermore, the global population is alarmingly developing MASLD at a younger age[10,11] and, most importantly, HCC etiology is shifting from viral to non-viral related causes, partly because of the reduce incidence of viral hepatitis, but mostly due to a higher MASLD and MASH incidence worldwide[12].

All in all, MASLD is a complex, multifactorial disease which poses a worrying threat to global health, and therefore a thriving field of both clinical and translational research. Notably, both MASLD and MASH are likely to be starring contributors to the development of T2D, kidney disease and cardiovascular disease independently of canonical risk factors, thus contributing to overall morbidity and mortality.

Inflammation and liver fibrosis

The most common hallmark of liver diseases, occurring in all stages regardless of the etiology, is chronic hepatic inflammation which represents one of the strongest and persistent trigger for the development of fibrosis, cirrhosis and HCC[13].

The complications in MASLD induced by immunity involve both low-grade inflammation induced by the adipose tissue and other downstream processes following cell damage. In fact, hepatocyte death by apoptosis or necrosis is followed by the release of damage-associated molecular patterns (DAMPs) which in turn activate inflammatory signaling cascades. In this way, damaged parenchymal cells and apoptotic bodies derived from dying hepatocytes can activate quiescent hepatic stellate cells (HSCs) and Kupffer cells (KCs), triggering sterile inflammation via several pathways. The clearest example of this process is the activation of TLR4/9 pathway[14,15], which in turn can activate activating NF-kβ and NLRP3-inflammasome pathways, thus promoting fibrogenic responses.

Inflammasomes are triggered by DAMPs, pathogen-associated molecular patterns (PAMPs) such as lipopolysaccharide (LPS), toxic lipids, and cholesterol crystals, which all contribute to liver inflammation in metabolic disease contexts[16]. Inflammasome activation, especially NLRP3, in KCs and damaged hepatocytes leads to release of IL-1β and IL-18, sustaining inflammation and promoting progression of NASH and fibrosis. In hepatocytes, IL-1 and IL-18 secretion signal converges on NF-kβ, besides inducing pyroptosis and further cellular damage[17–19]. Moreover, these inflammatory cytokines and pyroptotic process recruit KCs

via IL-6 and TNF-α, while also activating HSCs via TGF-β, hence causing fibrosis[20–22]. Liver sinusoidal endothelial cells (LSECs) as well play a key role in modulating the inflammatory process since in physiological conditions they prevent KCs and HSCs activation. Indeed, during MASLD pathogenesis they express, beyond several adhesion molecules, cytokines (IL-1, IL-6 and TNF-α), which enables the attraction of more immune cells which the subsequent failure in the repression of HSCs activation[23].

As resident KCs become activated, they either become congested with lipids and proinflammatory molecules or lose self-renewal capability and encounter cell death. Thus, native KCs are replaced by more pyrogenic monocyte-derived KCs (m-KCs), driving the progression to MASH through excessive inflammation[24].

Ultimately, the inflammatory process triggers aberrant deposition of fibrotic tissue, driving the onset and the development of liver fibrosis. The latter, together fueled by lipotoxicity, IR or chronic inflammation, by destroying the physiological architecture of the liver, limits liver perfusion, leading to loss of hepatic function[25]. For these reasons, the main cause of liver failure in MASH and cirrhotic patients is fibrosis and portal hypertension.

Various etiologies may contribute to damaging hepatocytes such as pathogenic conditions, toxic, metabolic and viral diseases, triggering the infiltration of immune cells that activate trans-differentiation of HSCs into collagen-producing myofibroblasts[26]. Once activated, HSCs begin to deposit excessive amounts of type I, III and IV collagen, fibronectin, and hyaluronic acid. Moreover, during fibrosis development, metalloproteinases' function is also compromised, therefore the novel extracellular matrix (ECM) presents composition and stiffness different from the physiological ECM[27].

While environmental and metabolic factors contributing to MASLD are well known, the genetic determinants of this condition are only partially understood, especially in the Italian population.

Genetics of MASLD and Genome wide association studies

MASLD is a complex and multifactorial pathology which is 25%–50% heritable[28].

First, retrospective studies demonstrated that relatives of patients with severe MASLD displayed dysglycemia and IR as well as predisposition to liver fibrosis. Then, consequently to the advent of genomic studies, the genetic architecture of MASLD contributing to disease onset, progression and outcome has been explored through unbiased approaches over the years. Genome-wide and exome-wide association studies (GWAS and EWAS) and later next generation sequencing (NGS) studies, for common and rare variants respectively, have also shown the importance of large and carefully defined cohorts to fully dissect the heritable component of MASLD. GWAS allowed the unbiased association of carriage of thousands of single-nucleotide polymorphisms (SNPs) to phenotypical traits and varying degrees of disease severity, and the main genetic factors predisposing to the disease have been described in several populations worldwide[28–30].

The first and major genetic variant associated with SLD is rs738409 C>G polymorphism, encoding for the p.I148M aminoacidic substitution of patatin like phospholipase domain containing 3 (PNPLA3) protein, a lipase expressed in the human liver, retina and adipose tissue. This variant favors MASLD development, onset and progression to MASH and fibrosis, indeed it is now accounted as an overall major determinant for MASLD susceptibility. Furthermore, the carriage in homozygosity increases the risk of SLD-derived HCC by almost 10 folds[31,32].

A second genetic determinant predisposing to MASLD is rs58542926 C>T, causing the E167K substitution in Transmembrane 6 superfamily member 2 (TM6SF2)[33]. This protein is involved in the regulation of lipid metabolism by loading TGs to the apolipoprotein B100 (ApoB-100), a key component of very-low density lipoproteins (VLDL)[34]. This loss of function mutation hampers lipid clearance from hepatocytes, thus predisposing to steatosis yet being protective against cardiovascular morbidities[35].

The rs641738 C>T variant has been shown to reduce the mRNA expression of Membrane bound O-acyltransferase domain-containing 7 (MBOAT7), a lysophosphatidylinositol acyltransferase specific for arachidonoyl-CoA involved in the Lands cycle which transfers arachidonic acid to lysophospholipids[33]. Hampered MBOAT7 expression results in higher intracellular arachidonic acid levels that are converted into pro-inflammatory eicosanoids, while the non-acylated lysophosphatidylinositol is redirected to triglycerides (TGs) synthesis, driving lipid accumulation thus predisposing to hepatic steatosis[36].

Glucose metabolism plays a pivotal role in the predisposition to MASLD as well; in fact, glucokinase regulator (GCKR) is a master regulator of hepatic glucose intake, and it is highly involved in regulating intracellular lipids, being the glucose the starting substrate for the novo lipogenesis. The alteration of the regulatory activity of GCKR on glucokinase is mainly determined by the rs1260326 C>T mutation, which encodes for the P446L missense substitution. This results in the hindering in response to fructose-6-phosphate, leading to uncontrolled uptake of glucose from hepatocytes, addressed to de novo lipogenesis[37].

Finally, the rs72613567 T>TA insertion in the splice region of intron 6 of hydroxysteroid 17-beta Dehydrogenase (HSD17B13), results in a frameshift that disrupts the function of HSD17B13, a lipid droplet-associated protein required for phosphatidylcholines and phosphatidylethanolamines remodeling on lipid droplets, besides steroid hormones, and retinoids metabolism and its activity is associated with MASH development[38]. The carriage of this mutation was demonstrated to be protective against MASH[39]. While the specific role of HSD17B13 is still being fully defined, its inhibition is an enticing novel therapeutical approach aimed at mimicking the loss of function variant in wild type patients[40].

Non-invasive biomarkers of liver fibrosis

Timely detection and staging of liver fibrosis are essential to prevent complications and to guide therapeutic decisions. In this optics, several advancements have been made toward

the development of non-invasive diagnostic tools and techniques. In recent years, the clinical utility of non-invasive diagnostic modalities has been steadily increasing, and many efforts have been made to find an alternative solution to liver biopsy, although it represents an invasive, costly, and time-consuming procedure with potential complications for the patient[41]. Several methodologies and a wide range of non-invasive markers and composite scores have been developed, validated and usefully employed to evaluate liver fibrosis, including serological markers, multi-omics-based methods and imaging-based modalities such as liver elastography[42].

In patients management, Fibrosis-4 Index (FIB-4) is recommended by the top-standard clinical guidelines as a first-line fibrosis assessment in patients with metabolic co-factors, followed by Fibroscan[43] (transient elastography) - a device that measures liver stiffness in a non-invasive imaging, providing rapid, reproducible, and operator-independent quantification of liver stiffness, which correlates with fibrosis severity - in those who need referral to a specialist liver clinic[3]. Among the alternative tools the most widely used and studied are the Fibrotic NASH Index (FNI), FibroScan-AST (FAST) score and Hepatic Steatosis Index (HSI). These tools have been integrated into clinical practice and are endorsed by hepatology guidelines for initial fibrosis assessment.

First of all, the FIB-4 index is a simple and a cost-effective score which was extensively validated. Initially developed to assess fibrosis in HIV/HCV co-infected individuals, later on has been widely adopted across different chronic liver diseases[44]. This score is particularly useful for ruling out advanced fibrosis in primary care and low-resource settings since it can guarantee accessibility and reproducibility and it is calculated as follows:

$$FIB-4 = \frac{Age\ (years)\ \times AST\ \left(\frac{U}{L}\right)}{Platelet\ count\ \left(\frac{10^9}{L}\right) \times \sqrt{ALT}\left(\frac{U}{L}\right)}$$

Then, the FNI, which incorporates multiple clinical and biochemical parameters, is a non-invasive composite score specifically developed to identify fibrotic MASH with reasonable

accuracy[45]. FNI offers a valuable screening tool to select patients who may benefit from more intensive evaluation or therapy and is gaining traction in both clinical and research settings, and it is defined as follows:

$$FNI = \frac{e^{\alpha+\beta+\gamma+\delta}}{1 + e^{\alpha+\beta+\gamma+\delta}}$$

where α = -10.33, β = 2.54 * ln(AST), γ = 3.86 * ln(HbA1c), and δ = -1.66 * ln(HDL).

All in all, FNI represents an accurate noninvasive score to screen for fibrotic MASH in individuals with dysmetabolism in primary health care, even if it lacks genetic validation so far.

From the introduction of FibroScan, the FAST score was specifically developed to identify patients with active MASH and significant fibrosis, who may be candidates for pharmacological intervention[46]. This is a composite index derived from FibroScan measurements, including both Liver Stiffness Measurement (LSM, kPa) and Controlled Attenuation Parameter (CAP, dB/m), and AST levels (U/L). Here the advantage is that the inclusion of both liver stiffness and biochemical markers allows for improved risk stratification and targeted patient management. FAST score is defined as follows:

$$FAST\ score = \frac{e^{(-1.65 + 1.07 \times \ln(LSM) + 2.66 \times \ln(AST) - 0.87 \times CAP)}}{1 + e^{(-1.65 + 1.07 \times \ln(LSM) + 2.66 \times \ln(AST) - 0.87 \times CAP)}}$$

Last, the HSI[47] is a non-invasive score developed to detect hepatic steatosis, a precursor of fibrosis, particularly in MAFLD patients, which couples also anthropometric parameters, such as BMI and biological sex, with T2D status and serological markers. It is calculated as follows:

$$HSI = 8 \times (ALT\ /\ AST) + BMI\ (+2\ if\ female;\ +2\ if\ diabetes)$$

HSI is relevant in identifying at-risk populations, despite the fact that it is not a direct fibrosis marker, since it serves as a screening tool in epidemiological and clinical settings in at-risk patients.

In this optics, using non-invasive markers of liver fibrosis as phenotypic traits to perform a

GWAS represents a novel and promising approach for several reasons.

First of all, due to the fact that liver biopsies are seldom available, especially for large cohorts, due to ethical and logistical constraints, non-invasive scores can be exploited thanks to their easy calculation from routine clinical data, enabling the inclusion of thousands of participants, raising the statistical power for genetic discovery. In the second place, this approach allows the identification of genetic variants associated with varying degrees of fibrosis severity; in fact, non-invasive markers are better suited for GWAS than binary outcome since they provide continuous traits.

Furthermore, they may help in uncovering new genes and pathways involved in fibrosis onset and progression, which are independent of known loci associated with histological fibrosis, unveiling mechanisms linked to metabolism, inflammation, and extracellular matrix remodeling.

The last reason resides in the clinical relevance of non-invasive marker; since being already used in clinical practice, identifying genetic determinants associated with them enhances their interpretability and could lead to personalized risk scores.

In addition, a deeper genetic understanding of non-invasive markers could be a perfectly suited instrument to provide a tool for a public health through the improvement of the stratification of the risk in the general population and support the development of polygenic risk scores (PRSs)[48], shedding light on the heterogeneity of MASLD and for the early identification of individuals at high risk of liver fibrosis.

Polygenic Risk Scores to improve risk stratification

In this setting, the construction of PRSs is a powerful strategy to quantify the cumulative impact of multiple genetic variants on an individual's susceptibility to liver fibrosis. PRSs incorporate the effect sizes of risk alleles from genome-wide association studies to estimate an individual's inherited risk. This strategy is especially useful in complex, multifactorial

diseases like MASLD and MASH, where genetic predisposition is interrelated with environmental and metabolic factors.

A first attempt to provide a predictive tool based on a PRS was provide by Dongiovanni and colleagues[49] who built a PRS based on known variants that influence hepatic fat accumulation (*PNPLA3*, *TM6SF2*, *GCKR*, *MBOAT7*), the PRS-HFC, and used it in Mendelian randomization analyses to demonstrate a causal role of hepatic fat in the development of liver damage and insulin resistance. Later, Bianco and collaborators[50] drawn on a robust polygenic risk scores (PRS-5), adding the contribution of *HSD13B17* protective genetic variants to the PRS-HFC, to be evaluated in the clinic to gain insight into the causal relationship between MASLD and HCC and to improve HCC risk stratification.

In 2023, Chen and colleagues carried out a GWAS meta-analysis of imaging and diagnostic code measured MAFLD across diverse ancestries, identifying 17 unique loci associated with the disease[30], shedding light on several potential subtypes of MASLD and developing a PRS to predict the risk to develop cirrhosis and HCC.

A main turning case in point is the study offered by Jamialahmadi et al. (Nat Med, 2024)[48], in which the authors developed and validated two different partitioned PRSs (pPRSs) for MASLD using histologically confirmed cases and showed its predictability for advanced liver disease, cardiovascular events, and all-cause mortality. Since the homeostasis of intrahepatocyte TGs is governed by either TGs synthesis, lipoprotein secretion or energy substrate utilization, they developed two pPRSs: one composed of variants in which the association between liver TGs content and circulating TGs were discordant and one in which they were concordant. Thwarting lipoprotein secretion causes liver TGs accumulation by retention. In fact, loss-of-function variants in *TM6SF2* and *PNPLA3* cause liver TGs retention by reducing lipoprotein secretion[51,52]. In their work, the concordant pPRS predicted the entire spectrum of cardiometabolic disease while the discordant pPRS was associated with liver disease mirrored by protection from cardiovascular disease due to lipoprotein retention,

despite a marginal increase in the risk of T2D.

Most importantly, they proposed the presence of at least two distinct types of MASLD with specific disease-causing molecular mechanisms: one specific for the liver and one systemic and entwined with cardiometabolic syndrome. Clinically, these subtypes translated clinically in the presence of individuals rapidly progressing to later stages of MASLD and those with a slow-progressing MASLD associated with the entire cardiometabolic syndrome. These types of MASLD may account for the disease heterogeneity and help explain why several drugs have failed in clinical trials to treat MASLD.

By transferring PRS to non-invasive fibrosis biomarkers, individuals can be stratified by genetic risk even before there is histological damage, facilitating earlier surveillance and intervention.

Application of this paradigm to non-invasive fibrosis biomarkers could augment clinical utility further by allowing personalized risk assessment on a scale, especially in large biobank or population-based cohorts where liver biopsy is not possible. Furthermore, PRS derived from GWAS on these biomarkers may reveal shared and distinct genetic pathways operative in fibrosis progression, informing both diagnostic approaches and therapeutic targets.

Aim

The aim of this study was to identify the genetic determinants of severe MASLD progressing to MASH and significant liver fibrosis by examining non-invasive biomarkers in a population enriched in people with MASLD, namely the Milan Biobank cohort, and to characterize new hits. The main outcome was FIB-4, as the first-line test employed in clinical practice. Other biomarkers of liver damage were assessed as secondary outcomes.

This project aims to delight the genetic mechanisms underlying the development and progression to liver fibrosis of MASLD exploiting multi-omics approaches such as genome- and phenome-wide association studies (PheWAS), public repositories and both bulk (bulk RNA-Seq) and single cell RNA sequencing (scRNA-Seq), validating the main loci associated in previous studies[28,30,53–56] with principal phenotypes linked to MASLD and identifying new genetic variants – i.e. SNPs – associated with liver fibrosis, cirrhosis and HCC in the Italian population.

The ultimate goal is to both improve MASLD risk stratification in the Italian population and to pinpoint novel genetic loci for pharmacological treatment to ameliorate hepatic damage in carriers of at-risk variants which predispose to liver damage and end-stage liver disease.

Methods

Study design

The investigation was designed as a cross-sectional GWAS using the Milan Biobank, a multicenter Italian resource that includes 7,025 participants recruited with uniform consent and ethics oversight. The cohort combines three recruitment streams (apparently healthy blood donors, dysmetabolic clinic attendees and patients with established liver disease) to encompass the complete range of MASLD phenotypes within one analytic framework. This wide, single-time-point ascertainment provided scope for harmonized phenotyping and genotyping across the biobank while reducing between-study heterogeneity inherent in multi-study meta-analyses. Phenotyping at baseline exploited non-invasive markers of fibrosis and routine biochemistry. The FIB-4 index was the primary genetic association outcome, selected for clinical traction and availability across the entire cohort; the FNI was included as a secondary quantitative trait, and liver enzymes (ALT, AST, GGT) were analyzed as ancillary phenotypes to place fibrosis-related signals in context. This hierarchy was pre-specified to concentrate power on fibrosis-related biology while allowing triangulation with liver injury biomarkers.

For assessing external validity, genome-wide significant loci were propelled for replication in the UK Biobank (UKBB) using independent samples with standardized phenotype definitions. The two-stage strategy thus seamlessly blends discovery into a national biobank with large-scale validation, hence reducing false positives and enhancing the generalizability of results outside the original discovery setting.

Clinical cohort(s)

The Milan Biobank is the result of an ongoing multi-center effort which led to the genome genotyping of DNA, which at the time of database closure for the present analysis (July 1st, 2025), was made up of 7,025 individuals from the Italian population whose biological

samples, clinical and genetic data were collected at the Biological Resource Center of the Fondazione IRCCS Ca' Granda Policlinico Ospedale Maggiore, Milano. The cohort (Figure 1) is composed of N=3,211 consecutive blood donors at Fondazione IRCCS Cá Granda Ospedale Maggiore Policlinico, Milan, belonging to FoGS cohort[57]; N=1,659 individuals with metabolic dysfunction belonging to LIVER-BIBLE cohort[58,59], who underwent noninvasive assessment of liver damage/fibrosis, with at least three criteria among overweight [BMI > 25 kg/m2], arterial hypertension, hyperglycemia [>100 mg/ dL], low HDL cholesterol [150 mg/dL]), and age at risk [40-65 years] have been enrolled[60]; N=2,155 patients with several degree of liver disease collected at Fondazione IRCCS Ca' Granda Policlinico Ospedale Maggiore, Milano, and several other centers across Italy (Napoli, Padova, Palermo, Roma, Torino, Udine).

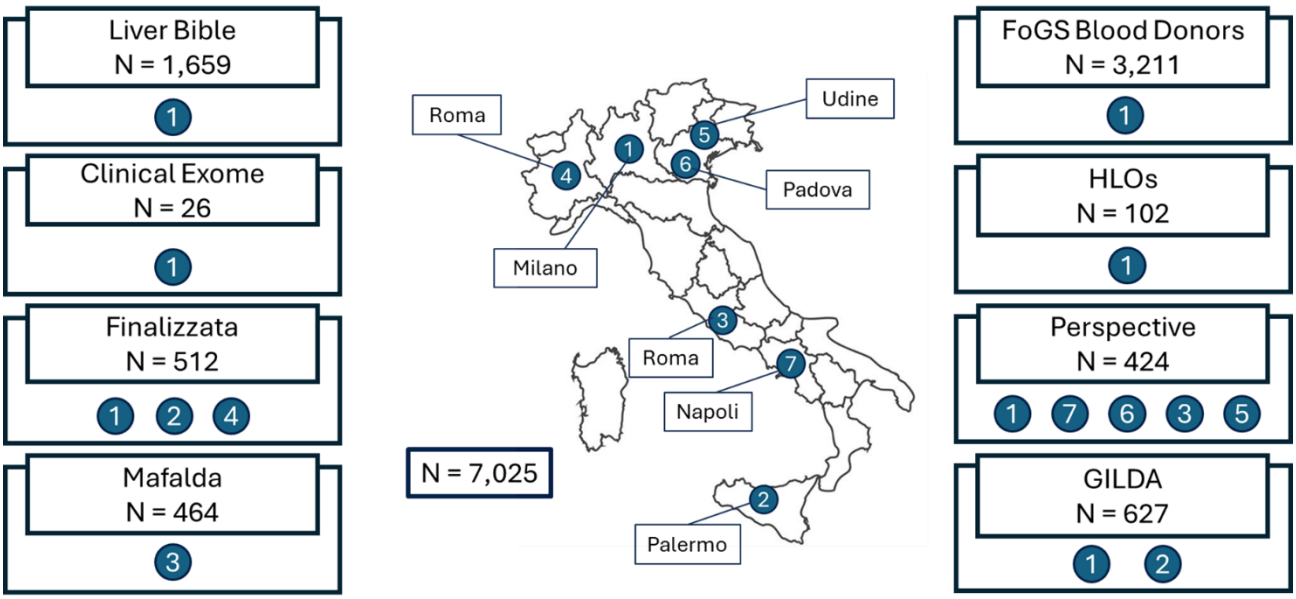

Figure 1. Milan Biobank cohort composition. This cohort represents a multicentric effort which enabled the collection of genetic material and phenotypic information of 7,025 individuals from the Italian population whose biological samples, clinical and genetic data were collected at the Biological Resource Center of the Fondazione IRCCS Ca' Granda Policlinico Ospedale Maggiore, Milano. The cohort covers a wide spectrum of MASLD phenotypes, and the samples were collected from various research centers across Italy.

The overall characteristics of the cohort are reported in the descriptive statistics of Table 1. Written informed consent was obtained, and the following approvals of the project were obtained from the relevant ethics committees of the Fondazione IRCCS Cá Granda Ospedale Maggiore Policlinico: reference numbers 91-2023bis, 00225305-2019-CS/21196,

485-2017, 96-2023 for patients and 334-2020 for control participants.

Trait	Overall	Trait	Overall
Sample Size	7,025	Fibrosis, Stage 0 (%)	269 (15.9)
Sex, Male (%)	4,582 (65.2)	Fibrosis, Stage 1 (%)	422 (24.9)
Age (Mean (SD))	47.88 (13.84)	Fibrosis, Stage 2 (%)	256 (15.1)
BMI (Mean (SD))	27.74 (6.22)	Fibrosis, Stage 3 (%)	191 (11.3)
AST (Mean (SD))	30.45 (32.56)	Fibrosis, Stage 4 (%)	554 (32.7)
ALT (Mean (SD))	32.83 (41.66)	Steatosis, Grade 0 (%)	96 (22.0)
GGT (Mean (SD))	40.92 (75.76)	Steatosis, Grade 1 (%)	150 (34.4)
CHO (Mean (SD))	189.43 (37.48)	Steatosis, Grade 2 (%)	112 (25.7)
PLTs (Mean (SD))	229.61 (65.03)	Steatosis, Grade 3 (%)	78 (17.9)
Glycemia (Mean (SD))	95.66 (25.23)	Necroinflammation, Grade 0 (%)	141 (31.8)
HDL (Mean (SD))	53.36 (15.63)	Necroinflammation, Grade 1 (%)	249 (56.1)
TGs (Mean (SD))	120.24 (76.83)	Necroinflammation, Grade 2 (%)	50 (11.3)
LSM (Mean (SD))	7.92 (8.14)	Necroinflammation, Grade 3 (%)	4 (0.9)
CAP (Mean (SD))	275.04 (51.22)	Ballooning, Grade 0 (%)	73 (16.7)
FIB-4 (Mean (SD))	1.35 (1.58)	Ballooning, Grade 1 (%)	283 (64.9)
FNI (Mean (SD))	0.22 (0.23)	Ballooning, Grade 2 (%)	80 (18.3)
Obesity, Yes (%)	1,974 (28.1)	Cirrhosis, Yes (%)	554 (8.2)
Diabetes, Yes (%)	777 (11.1)	HCC, Yes (%)	164 (2.3)

Table 1. Descriptive Statistics of the Milan Biobank.

Sample Processing, Genotyping, Quality Control (QC) and Imputation

DNA extraction was either performed at the Institute of Clinical Molecular Biology (IKMB, Christian-Albrechts-University of Kiel, Germany, for the FoGS cohort) or the respective study centers (Dysmetabolic patients and patients with liver disease) (Figure 1) from whole-blood or buffy coat samples, using a Chemagic 360 (PerkinElmer) with the use of the low-volume kit CMG-1491 and the buffy-coat kit CMG-714 (Chemagen), respectively.

Genome genotyping was performed either at the Institute of Clinical Molecular Biology (IKMB, Christian-Albrechts-University of Kiel, Germany, for the FoGS cohort) or at the Institute for Molecular Medicine Finland (FIMM) Technology Centre at the University of Helsinki (Dysmetabolic patients and patients with liver disease).

For genotyping, the Global Screening Array (GSA), version 3.0 (Illumina), containing 712,189 variants before quality control, was used. Details on genotyping and quality-control procedures are described in previous publications[61–63]. Briefly, samples with <90% call rate

were removed using Genome Analysis Tool Kits (GATK), based on initial genotype data, finally accounted for 465,044 SNPs. Additionally, individuals with non-matching genotypic and phenotypic sex were discarded from the genotyping sample set. Variants that had >5% missing data, an MAF < 0.1% in disease sets or in blood donors, different missing genotype rates in affected and unaffected individuals ($p_{Fisher} < 10^{-5}$) or deviated from Hardy–Weinberg equilibrium (with a false discovery rate (FDR) threshold of 10^{-5} in controls) were excluded. After genome genotyping, samples that had overall increased/decreased heterozygosity rates (i.e., ±5 SD away from the sample mean) were removed. For robust duplicate/relatedness testing (IBS/IBD estimation) and population structure analysis, a linkage disequilibrium (LD)-pruned subset of SNPs was used on the basis of a set of independent (MAF ≥ 5%) SNPs excluding X- and Y-chromosomes, SNPs in LD (leaving no pairs with $R^2 > 0.2$) and 11 high-LD regions as described by Price et al[64]. For ancestry stratification of the overall cohort, principal component analysis (PCA) for the genotyped subjects using PLINK v1.9 was performed, in order to correct the association analysis for the first 10 principal components (PCs).

To maximize genetic coverage, SNP imputation was performed on genome build GRCh38 using the Michigan Imputation Server and 194,512 haplotypes generated by the Trans-Omics for Precision Medicine (TOPMed) program (Freeze 5). A total of 8,609,091 SNPs were included in the cohort.

Genome-wide association study

For the phenotypic association genome-wide SNP information was used to test for phenotypic associations with allele dosage data, performing a logistic regression analysis for dichotomous traits or a linear regression analysis for continuous traits, assuming additive effects with REGENIE software for chromosomes 1–22 an X, adjusting the analysis for age, biological sex, T2D status and to correct for the ethnicity the first 10 PCs from PCA as

covariates, as previously described in Degenhardt et al[62].

Validation in UK Biobank

To validate the association of *MRC1* locus with FIB-4 and FNI, the publicly available genetic and phenotypic data gathered in the UKBB project[65], accounting for 488,377 participants whose genotyping arrays data were available, were used. Quality control and imputation processes followed the same rules used for the Milano Biobank cohort, in order to guarantee reproducibility and reliability of the results. A total of 9,361,425 variant were considered in the analysis after quality control steps. GWAS on FIB-4 accounted for 401,607 individuals while 359,545 individuals were considered for the GWAS on FNI. Again, for the phenotypic association genome-wide SNP information was used to test for phenotypic associations with allele dosage data, a linear regression analysis assuming additive effects with REGENIE software for chromosomes 1–22 an X, adjusting the analysis for age, biological sex, T2D status and to correct for the ethnicity the first 10 PCs from PCA as covariates.

Polygenic risk scores testing

The pPRS take the concept of PRS a step further by dividing the genetic variants into distinct categories or partitions based on certain criteria, allowing for more nuanced risk predictions based on functional annotation, biological pathways, genomic regions and heritability enrichment. To discriminate which aspect of MASLD pathophysiology is more closely associated with clinical events, a straightforward comparison among the performance of different PRSs – namely PRS-HFC[66], summarizing the risk of hepatic fat accumulation, PRS-5[50], summarizing the risk of steatohepatitis due to the main risk variants, pPRS-concordant[48], summarizing the processes leading to fat accumulation due to enhanced lipogenesis and therefore resulting in increased secretion of lipoproteins, and pPRS-discordant[48], summarizing the processes leading to fat accumulation due to reduced lipoproteins secretion and/or increased uptake was performed. Furthermore, their accuracy

in the prediction of phenotypes related to liver disease was assessed by testing their diagnostic ability through the calculation of the area under the receiver operating characteristic (AUROC) curve.

The two novel PRSs for FIB-4 and FNI used in the study of PRSs comparison were calculated using the summary statistics of GWAS in the Milan Biobank, considering the top variant for each significant genetic locus, setting the permissive statistical significance threshold at $p \leq 1*10^{-6}$. To verify how well the score distinguishes between individuals, the predicted probabilities were retrieved after fitting a binomial generalized linear model for each dichotomous trait. Fibrosis was defined as LSM≥8 and/or presence of fibrotic insults at the histological evaluation or at the enrolment. Two distinct model were considered in the analysis, the first unadjusted was performed in order to evaluate the predictivity of PRSs avoiding spurious associations due to the effect of comorbidities, which were indeed included in the second model which was adjusted for age, biological sex, BMI, T2D status and the first five PCs. To provide a statistics for heterogeneity for each PRS in the two models, the significance associated to the AUROC was presented separately, before the pairwise comparison made by De Long test.

Receiver operating characteristic (ROC), areas under the curve (AUCs), p-values and graphical representation have been calculated using the pROC package in RStudio.

Transcriptomic analysis

Bulk RNA-Seq public sequences were obtained from Gene Expression Omnibus (GEO) repository under accession numbers GSE13525117[67], including 216 snap frozen liver biopsies, comprising 206 NAFLD cases with different fibrosis stages and 10 controls, and GSE23942218[68], hepatic transcriptome of 125 obese individuals who underwent percutaneous liver biopsy performed during bariatric surgery. Histological features such as steatosis, ballooning, necroinflammation grades and NAFDL Activity Score (NAS) were

available only in the bariatric cohort while the fibrosis stage were analyzed in the two cohort joined together. scRNA-Seq dataset was obtained under accession number GSE18960019[69], which includes data from the single nuclei RNA-Seq (snRNA-Seq) of 3 human NASH livers and 3 healthy controls. Reads were mapped by a custom pipeline[68], encompassing reads quality check (FastQC software, Babraham Bioinformatics, Cambridge, UK), low-quality reads trimming and mapping on GRCh38 reference genome by STAR mapper. Reads count (ENSEMBL human transcript reference assembly v99) was performed using RSEM package. Counts were normalized using DESeq2 package and RUVg package was used for data harmonization to reduce biases due experimental procedures. For principal PCA, gene-level expression data were normalized under the null model through DESeq2 standard pipeline, and variance stabilizing transformation function was applied. To identify differentially expressed pathways, pre-ranked gene set enrichment analysis (GSEA) was performed on differentially expressed or significantly correlated genes. Genotype-Tissue Expression (GTEx) portal[70], which provides open access to data including gene expression, was used to asses possible expression quantitative trait loci (eQTL) between the rs2461143 C>G *MRC1* (top hit form GWAS) and MRC1 gene expression.

To compare the transcriptomic profile of patients with lower and higher *MRC1*, patients were divided into two groups setting the threshold of *MRC1* expression at the third tertile.

Statistical analysis

All statistical analyses were performed in accordance with established genome-wide association study standards. Continuous traits, including FIB-4, FNI, and circulating liver enzymes, were rank-based inverse normal transformed prior to association testing to approximate normal distributions.

Genome-wide association analyses were conducted using REGENIE, a two-step ridge regression–based approach optimized for large-scale biobank data. Additive genetic models were assumed, and regression models were adjusted for relevant covariates including age,

sex, T2D status, and the first ten principal components from principal component analysis to correct for population stratification. Variants with a minor allele frequency below 0.1%, call rate below 95%, deviation from Hardy–Weinberg equilibrium ($p < 1\times10^{-5}$ in controls), or differential missingness between cases and controls were excluded. Post-imputation analyses included only SNPs with an imputation quality score (INFO) > 0.1. To account for multiple testing, genome-wide significance was set at $p < 5\times10^{-8}$, while suggestive associations were reported at $p < 1\times10^{-6}$. Quantile–quantile plots and genomic inflation factors (λ_{GC}) were inspected to distinguish true polygenic signal from systematic bias. Replication analyses in the UKBB followed identical procedures, ensuring harmonization of quality control and association pipelines across datasets. For PRSs analyses, summary statistics from discovery GWAS were used to compute weighted allele scores, which were then tested for predictive performance against fibrosis, cirrhosis, and HCC using logistic regression and receiver operating characteristic (ROC) analysis. De Long test was used for PRSs pairwise comparison. All analyses were performed in PLINK v1.9, REGENIE, and RStudio using dedicated statistical packages, with correction for multiple testing applied where appropriate.

Results

Single nucleotide polymorphisms in genes involved in hepatic lipid metabolism are the main genetic determinants which are associated with main circulating biomarkers for hepatic damage

To investigate the genetic determinants of hepatic liver damage in the Italian population, in order to validate the Milan Biobank, first a GWAS was conducted on directly assessed biomarkers used to calculate liver damage scores, namely circulating alanine aminotransferase (ALT, N=7,025), aspartate aminotransferase (AST, N=7,025), γ-glutamyl transferase (GGT, N=7,025) levels, in the Milan Biobank which accounted for 7,025 subjects.

Fifteen unique genome-wide significant variants associated with these three phenotypes were identified (Table 2); among these loci, seven (*HAPLN4*/*TM6SF2*, *PNPLA3*, *MRC1*, *HNF1A*, *GGT1*, *RORA*, *EXOC3L4*) have been associated with these liver traits in previous GWASs.

The *PNPLA3* locus on chromosome 22 (22q13.31), is driven by rs3747207 G>A (β=0.161±0.016, p=3.83*10^{-23}, MAF=0.318), rs738408 C>T (β=0.177±0.017, p=1.06*10^{-26}, MAF=0.322), and rs3747207 G>A (β=0.086±0.015, p=8.30*10^{-9}) common variants, respectively for ALT, AST and GGT.

Trait	Variant	Locus	Band	Ref	ALT	Most severe consequence	MAF	β	SE	-log$_{10}$(p)
ALT	rs3747207	PNPLA3	22q13.31	G	A	intron variant	0.318	0.161	0.016	22.417
	rs56896489	VMA21	Xq28	C	T	intergenic variant	0.019	0.314	0.052	8.915
	rs16988467	IL1RAPL1	Xp21.3	A	G	intron variant	0.013	0.345	0.058	8.609
	rs150641967	HAPLN4/TM6SF2	19p13.11	TGACA	T	intron variant	0.059	0.185	0.032	8.192
	rs4571922	LRRIQ3	1p31.1	A	G	intergenic variant	0.021	0.693	0.123	7.708
AST	rs738408	PNPLA3	22q13.31	C	T	synonymous variant	0.322	0.177	0.017	25.975
	rs2477633	MRC1	10p12.33	C	A	intron variant	0.518	0.131	0.017	14.570
	rs10013613	NFKB1	4q24	G	T	intron variant	0.337	-0.102	0.017	8.408
GGT	rs2073397	GGT1	22q11.23	T	C	intron variant	0.443	0.142	0.014	21.800
	rs339969	RORA	15q22.2	C	A	intron variant	0.579	0.104	0.014	13.270
	rs3747207	PNPLA3	22q13.31	G	A	intron variant	0.318	0.086	0.015	8.081
	rs944002	EXOC3L4	14q32.32	A	G	intron variant	0.252	0.087	0.016	7.577
	rs1169299	HNF1A	12q24.31	T	C	intron variant	0.424	-0.077	0.014	7.375

Table 2. Genome-wide association study of circulating liver enzymes (ALT, AST and GGT) available for 7,025 individuals of the Milan Biobank (MBB) revealed 12 unique significant loci associated with these traits.

The *TM6SF2/HAPLN4* locus on chromosome 19 (19p13.11) associated with circulating ALT levels and it is indeed driven by the rs150641967 TGACA>T (β=0.185±0.032, p= $6.42*10^{-9}$, MAF=0.059) variant, which is in perfect linkage with the rs58542926 C>T (encoding for the E167K substitution). In addition, chromosome 1 (1p31.1 locus) circulating ALT levels were associated with rs6681577 A>G variant in the *LRRIQ3* gene (β=0.693±0.123, p=$2.57*10^{-8}$, MAF=0.021). Finally, the same trait associated with the rs56896489 C>T variant of the *VMA21* locus (β=0.314±0.052, p=$1.22*10^{-9}$, MAF=0.019) in the X chromosome (Xq28) and with the rs184838841 A>G variant in the *IL1RAPL1* locus (β=0.345±0.058, p=$3.40*10^{-9}$, MAF=0.013) in the X chromosome (Xp21.3).

Beyond the association with *PNPLA3* rs738408 C>T variants, in perfect linkage with the rs738409 I148M, the circulating AST levels associated with rs2477633 variant at the *MRC1* locus (β=0.131±0.017, p= $2.69*10^{-15}$, MAF=0.518) in the chromosome 10 (10p12.33) and with rs10013613 G>T variant at the *NFKB1* locus (β=-0.102±0.017, p=$3.91*10^{-9}$, MAF=0.337) in the chromosome 4 (4q24).

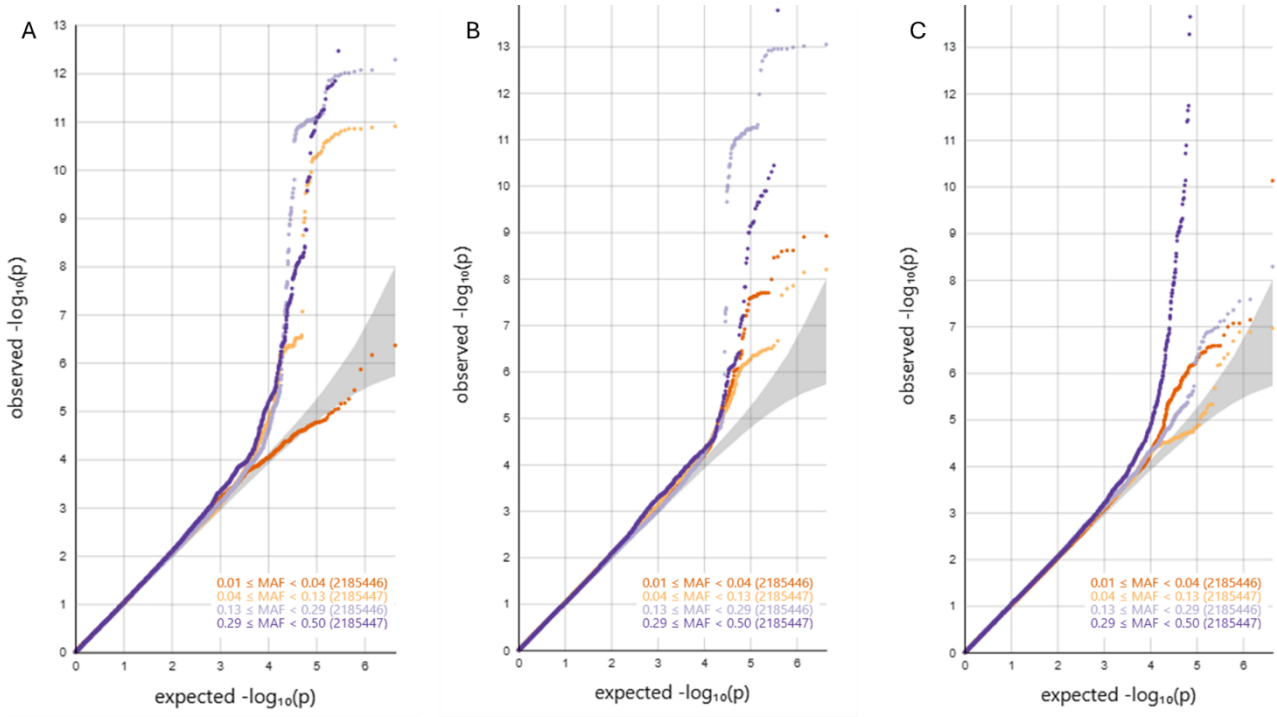

Figure 3. QQ Plot showing genomic inflation factor (λ_{GC}) for the gnome-wide analysis of circulating liver enzymes, *namely ALT (A), AST (B) and GGT (C).*

The analysis on GGT circulating levels revealed unique association also with *GGT1* (22q11.23, rs2073397 T>C, β=0.142±0.014, p=1.67*10^{-22}, MAF=0.443), *RORA* (15q22.2, rs339969 C>A, β=0.104±0.014, p=5.37*10^{-14}, MAF=0.579), *EXOC3L4* (14q32.32, rs944002 A>G, β=0.087±0.016, p=2.65*10^{-8}, MAF=0.252) and *HNF1A* (12q24.31, rs1169299 T>C, β=-0.077±0.014, p=4.21*10^{-8}, MAF=0.424) loci.

The genomic inflation factor (λ_{GC}) for ALT (Figure 3A, λ=1.022), AST (Figure 3B, λ=1.020) and GGT (Figure 3C, λ=1.024) respectively, suggests that the observed inflation is due to polygenicity of the traits rather than systematic error carried out during the analysis.

Next, a GWAS on MASLD-related endophenotypes was performed, namely total cholesterol (CHO, N=7,025), HDL (N=7,025), low-density lipoproteins (LDL, N=7,025) and TGs (N=7,025).

Coherently with the results obtained for the analysis of circulating liver enzyme levels, I highlighted 11 unique variants located in genes involved in fat/lipid metabolism, all of which have been reported in previous GWASs (Table 3).

The rs58542926 C>T variant in *TM6SF2* locus at the chromosome 19 (19p13.11) associated with a reduction of CHO (β=-0.208±0.033, p=8.25*10^{-12}, MAF=0.060), LDL (β=-0.208±0.033, p=2.30*10^{-10}) and TGs (β=-0.185±0.031, p=2.27*10^{-9}) circulating levels but not with HDL, which conversely was driven by *CETP* on chromosome 16 (16q13, rs247617 C>A, β=0.206±0.016, p=4.72*10^{-39}, MAF=0.275), LIPC/*ALDH1A2* on chromosome 15 (15q21.3; rs1800588 C>T, β=0.128±0.017, p=2.06*10^{-13}, MAF=0.204) and *LPL* on chromosome 8 (8p21.3, rs3208305 A>T, β=0.094±0.015, p=3.39*10^{-10}, MAF=0.338) loci.

The *APOE/APOC1* locus, driven by the rs7412 C>T and rs483082 G>T variants on chromosome 19 (19q13.32), were associated with a reduction of both circulating CHO (β=-0.237±0.033, p=1.02*10^{-12}, MAF=0.062) and LDL (β=-0.319±0.033, p=6.65*10^{-22}), and an increase in TGs levels (β=0.121±0.021, p=7.19*10^{-9}, MAF=0.156) respectively; similarly, *PSRC1/CELSR2* locus, which is driven by rs12740374 G>T variant on chromosome 1

(1p13.3), was inversely associated with CHO levels (β=0.125±0.020, p=2.34*10^{-10}, MAF=0.194) and with circulating LDL levels (β=-0.121±0.020, p=7.68*10^{-10}).

Circulating CHO levels were also negatively associated with *LDLR*/*SMARCA4* locus on chromosome 19 (19p13.2, rs113722226 T>C, β=-0.133±0.024, p=4.31*10^{-8}, MAF=0.126).

In addition, other 5 variants were found inversely associated with circulating TGs levels in *TBL2* (7q11.23; rs67644909 C>T, β=-0.105±0.018, p=3.16*10^{-9}, MAF=0.235), *LPL* (8p21.3; rs15285 C>T, β=-0.134±0.016, p=1.69*10^{-17}, MAF=0.338), *GCKR* (2p23.3; rs1260326 T>C, β=-0.095±0.015, p=1.84*10^{-10}, MAF=0.465), *ZPR1* (11q23.3; rs964184 G>C, β=-0.235±0.020, p=5.09*10^{-32}, MAF=0.829) and *TRIB1/TRIB1AL* (8q24.13; rs28601761 C>G, β=-0.091±0.016, p=5.41*10^{-9}, MAF=0.393) loci.

Circulating LDL levels were also positively associated with *APOB* locus on chromosome 2 (2p24.1, rs548145 T>C, β=0.123±0.020, p=3.16*10^{-10}, MAF=0.786).

The genomic inflation factor (λ_{GC}) for CHO, HDL, LDL and TGs were, respectively, λ=1.019 (Figure 2A), λ=1.020 (Figure 4B), λ=1.012 (Figure 4C) and λ=1.019 (Figure 4D). Again, this parameters indicate that the observed inflation is due to polygenicity of the traits rather than systematic error carried out during the analysis.

Trait	Variant	Locus	Band	Ref	ALT	Most severe consequence	MAF	β	SE	-log$_{10}$(p)
CHO	rs113722226	LDLR/SMARCA4	19p13.2	T	C	intron variant	0.126	-0.133	0.024	30.004
	rs12740374	PSRC1/CELSR2	1p13.3	G	T	3' UTR variant	0.194	-0.125	0.02	40.160
	rs58542926	TM6SF2	19p13.11	C	T	missense variant	0.060	-0.226	0.033	46.706
	rs7412	APOE	19q13.32	C	T	missense variant	0.062	-0.237	0.033	50.798
HDL	rs247617	CETP	16q13	C	A	intergenic variant	0.275	0.206	0.016	38.326
	rs1800588	LIPC/ALDH1A2	15q21.3	C	T	intergenic variant	0.204	0.128	0.017	12.686
	rs3208305	LPL	8p21.3	A	T	3' UTR variant	0.338	0.094	0.015	9.470
LDL	rs7412	APOE	19q13.32	C	T	missense variant	0.062	-0.319	0.033	21.177
	rs58542926	TM6SF2	19p13.11	C	T	missense variant	0.060	-0.208	0.033	9.639
	rs548145	APOB	2p24.1	T	C	intron variant	0.786	0.123	0.02	9.500
	rs12740374	PSRC1/CELSR2	1p13.3	G	T	3' UTR variant	0.194	-0.121	0.02	9.114
TGs	rs964184	ZPR1	11q23.3	G	C	3' UTR variant	0.829	-0.235	0.02	31.294
	rs15285	LPL	8p21.3	C	T	3' UTR variant	0.338	-0.134	0.016	16.773
	rs1260326	GCKR	2p23.3	T	C	missense variant	0.465	-0.095	0.015	9.735
	rs58542926	TM6SF2	19p13.11	C	T	missense variant	0.060	-0.185	0.031	8.643
	rs67644909	TBL2	7q11.23	C	T	upstream gene variant	0.235	-0.105	0.018	8.501
	rs28601761	TRIB1/TRIB1AL	8q24.13	C	G	intron variant	0.393	-0.091	0.016	8.267
	rs483082	APOE/APOC1	19q13.32	G	T	NC transcript exon variant	0.156	0.121	0.021	8.143

Table 3. Genome-wide association study of MASLD-related circulating endophenotypes (CHO, HDL, LDL and TGs) available for 7,025 individuals of the Milan Biobank (MBB) revealed 11 unique significant loci associated with these traits. CHO=Total cholesterol, HDL=High-density lipoproteins, LDL=Low-density lipoproteins, TGs=Triglycerides

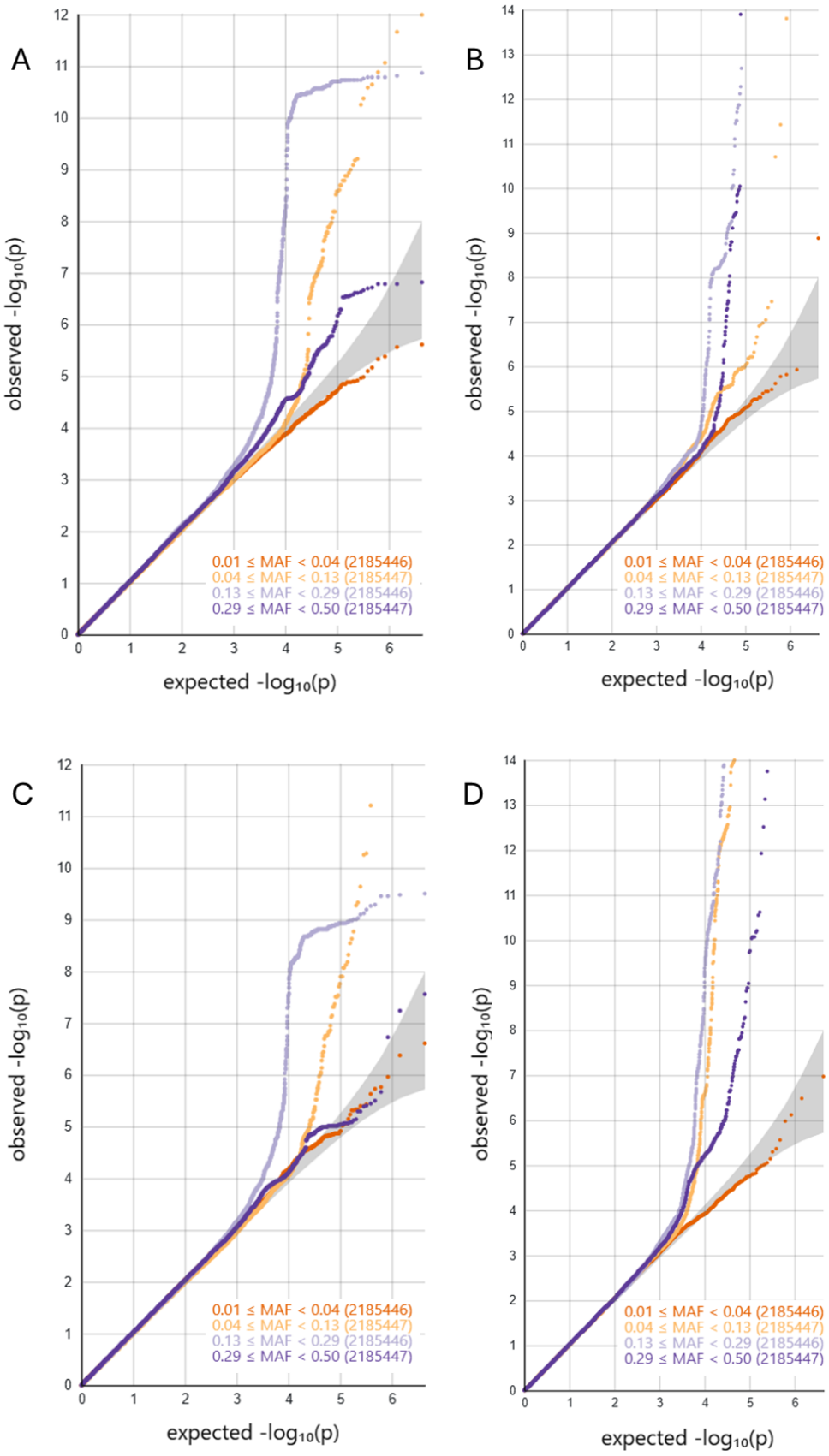

Figure 4. QQ Plot showing genomic inflation factor (λ_{GC}) for the gnome-wide analysis of circulating liver enzymes, *namely CHO (A), HDL (B), LDL (C) and TGs (D).*

As part of sensitivity analyses, genome-wide association studies were also performed on platelet counts (PLTs), one of the determinants of the FIB-4 score, and on T2D, which was included as a covariate in the main GWAS models.

The GWAS for PLTs identified four significant genome-wide loci (Table 4). The *PNPLA3* locus (22q13.31), driven by the missense variant rs738409 C>G (MAF=0.325, β=-0.127±0.016, $p=1.4*10^{-14}$), confirmed its pleiotropic role in MASLD-related traits. Additional associations were observed at *HBS1L* (6q23.3, rs9402685 T>C, MAF=0.240, β=0.128±0.018, $p=1.1*10^{-11}$), *VWA7* (6p21.33, rs6914412 T>C, MAF=0.085, β=-0.160±0.028, $p=8.1*10^{-9}$), and *AK3* (9p24.1, rs77798356 C>G, MAF=0.066, β=0.187±0.034, $p=7.3*10^{-8}$), pointing to genetic regulation of platelet production and turnover.

The GWAS for T2D highlighted two loci of interest (Table 4). The *PNPLA3* rs738409 C>G variant (MAF=0.325, β=0.367±0.064, $p=7.9*10^{-8}$) was again significantly associated, underscoring its contribution to both hepatic and metabolic phenotypes. Furthermore, a novel association was identified on X chromosome at *VMA21* (rs56896489 C>T, MAF=0.019, β=1.063±0.194, $p=7.4*10^{-8}$), an intergenic variant previously implicated in lysosomal function and cellular homeostasis. Together, these findings confirm that variants at *PNPLA3* exert a broad pleiotropic influence, while highlighting additional loci contributing to platelet biology and T2D risk that may indirectly affect liver fibrosis assessment through their impact on FIB-4 components and metabolic comorbidities.

Trait	Variant	Locus	Band	Ref	ALT	Most severe consequence	MAF	β	SE	$-\log_{10}(p)$
PLTs	rs738409	PNPLA3	22q13.31	C	G	missense variant	0.325	-0.127	0.016	14.367
	rs9402685	HBS1L	6q23.3	T	C	intergenic variant	0.240	0.128	0.018	11.534
	rs6914412	VWA7	6p21.33	T	C	intron variant	0.085	-0.16	0.028	8.058
	rs77798356	AK3	9p24.1	C	G	intergenic variant	0.066	0.187	0.034	7.325
T2D	rs738409	PNPLA3	22q13.31	C	G	missense variant	0.325	0.367	0.064	7.939
	rs56896489	VMA21	Xq28	C	T	intergenic variant	0.019	1.063	0.194	7.368

Table 4. Genome-wide association study, used as sensitivity analysis on PLTs (N=7,025) of the determinants of the FIB-4 score, and on T2D (N=7,025), which was included as a covariate in the main GWAS models, revealed 5 unique significant loci associated with these traits. PLTs=Platelet count, T2D=Type 2 diabetes.

Genome-wide association study of FIB-4 and FNI revealed *PNPLA3, MRC1, MAGEE2* and *TM6SF2* loci as the main genetic drivers of advanced liver disease

Given the consistency of the results previously obtained by the analyses in the Milan Biobank, to uncover new loci underpinning advanced liver damage, the associations of SNPs with validated non-invasive markers of liver fibrosis were investigated.

For the first time was performed a GWASs on FIB-4 (N=7,025), the recommended first-line screening test for hepatic fibrosis in people at risk of MASLD (main outcome), and on FNI (N=3,210), which represents an accurate noninvasive scores used to screen for fibrotic MASH in individuals with dysmetabolism in primary health care.

PNPLA3 rs738408 G>A variant which is almost in perfect linkage with the rs738409 I148M *PNPLA3* at the 22q13.31 locus, was the strongest genetic determinant associated with both traits, showing a positive association with higher FIB-4 score (Figure 5A, β=0.098±0.012, $p=5.32*10^{-16}$, MAF=0.322) and with FNI (Figure 5B, β=0.165±0.023, $p=9.57*10^{-13}$).

In addition, two novel associations for the FIB-4 were found: on chromosome 10, the intronic *MRC1* rs2461143 T>C SNP (Figure 5A, 10p12.33, β=0.068±0.012, $p=2.17*10^{-8}$), and on X chromosome the *MAGEE2* rs1813664 T>A SNP (Figure 5A, Xq13.3, β=0.462±0.076, $p=9.99*10^{-10}$, MAF=0.034).

The *MRC1* gene indeed encodes the mannose receptor C-type 1, also known as CD206. This receptor is primarily found on the surface of macrophages and dendritic cells and plays a crucial role in the innate immune system[71,72]. On the other hand, although the *MAGEE2* gene has been shown to be expressed in various tissues, including the liver, and to be associated with certain cancers such as melanoma and HCC, its specific function within the liver is not well-defined.

Additionally, beyond *PNPLA3* variant, three novel genome-wide significant associations were found for FNI. First, *TM6SF2* locus*,* coherently with altered MASLD endophenotypes

such as CHO, LDL, TGs circulating levels in the Milan Biobank showed in previous analysis, was associated, through the low-frequency rs56255430 A>C intron variant, with higher FNI (Figure 5B, β=0.238±0.043, $p=4.68*10^{-8}$, MAF=0.067, 99% LD with rs58542926 C>T variant). Conversely, inverse associations with FNI were found with *IFI27* rs10145931 (Figure 5B, β=-0.204±0.035, $p=8.71*10^{-9}$, MAF=0.117) and *DIAPH2* rs5920923 (Figure 5B, β=-1.144±0.161, $p=1.40*10^{-12}$, MAF=0.018) variants.

IFI27 (Alpha-Inducible Protein 27) is an interferon-stimulated gene which has been identified as a potential marker for liver cirrhosis, potentially influencing inflammation and macrophage activity[73]. The *DIAPH2* (diaphanous related formin 2) gene is a protein-coding gene that belongs to a family of proteins called formins, which are involved in regulating the cytoskeleton by controlling actin nucleation and polymerization. It is expressed in various tissues, including the liver, and it has been implicated in the development and metastasis of various cancers, with potential involvement in HCC since its alterations can increase cell motility, which is a factor in tumor metastasis[74].

The genomic inflation factor (λ_{GC}) for FIB-4 was λ=1.019 (Figure 5C) and for FNI was λ=1.019 (Figure 5D).

Finally, the validation of FIB-4 (N=401,607) and FNI (N=359,545) results obtained in the Milan Biobank were replicated in the UKBB. The *MRC1* locus was associated with both FIB-4 (rs56278466 T>G, β=0.096±0.002, $p=8.45*10^{-559}$, MAF=0.662) and FNI (rs56278466 T>G, β=0.102±0.002, $p=3.33*10^{-534}$, MAF=0.662), showing the consistency with the results in the Italian population (Table 5).

In addition, a LD analysis was performed in the Milan Biobank in order to investigate possible pattern of heritability between the *MRC1* significant SNP obtained from GWAS analysis. The intronic *MRC1* rs2461143 T>C SNP resulted in poor linkage with the rs56278466 T>G SNP, with R^2=0.53 (Figure 5E) and D'=0.96 (Figure 5F).

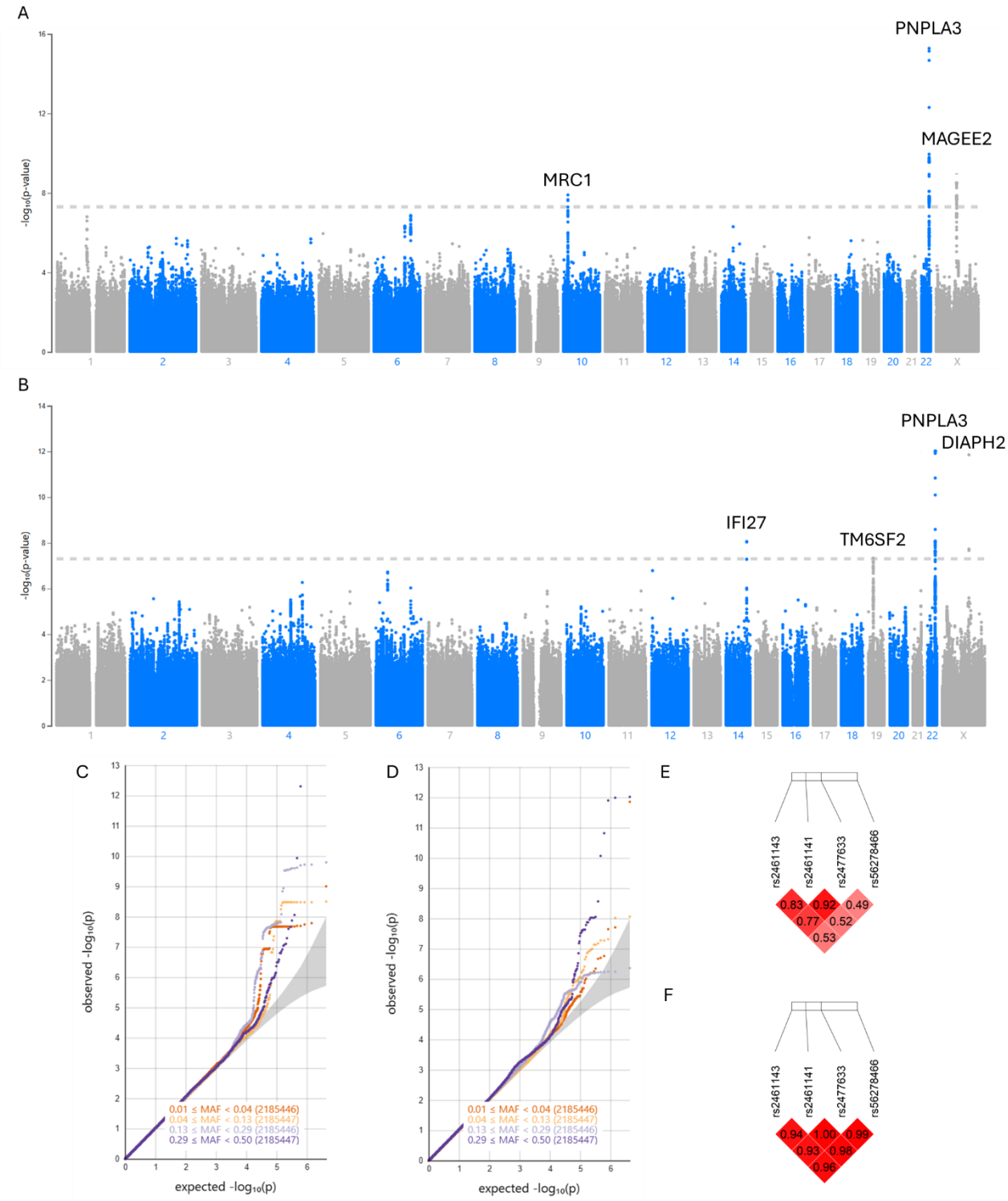

Figure 5. Genome-wide association study of noninvasive scores of fibrotic MASLD in the Milan Biobank (N=7,025). (A) GWAS on FIB-4 revealed novel associations with MRC1 and MAGEE2 loci while (B) GWAS on FNI revealed novel associations with IFI27 and DIAPH2 loci. (C-D) Q-Q plots of the analyses showing the observed inflation is due to polygenicity of the traits rather than systematic errors. (E-F) LD analysis showed poor heritability between intronic MRC1 rs2461143 T>C and rs56278466 T>G SNPs by their R^2 (E) and D' (F) scores.

Trait	Variant	Locus	Band	Ref	ALT	Most severe consequence	MAF	β	SE	-log$_{10}$(p)
FIB-4	rs56278466	MRC1	10p12.33	T	G	Intron variant	0.662	0.096	0.002	558.073
FNI	rs56278466	MRC1	10p12.34	T	G	Intron variant	0.662	0.102	0.002	533.477

Table 5. UKBB replication of MRC1 locus: the gene resulted associated with noninvasive biomarker of liver fibrosis.

Comparison of 7 polygenic risk scores of SLD for the prediction of advanced liver disease

Given the consistency of the results obtained from the GWAS of both liver enzymes, circulating MASLD endophenotypes and noninvasive marker of liver fibrosis and steatohepatitis in the Milan Biobank, which highlighted new genetic risk loci, the PRSs for FIB-4 and FNI were calculated based on summary statistics and then compared against the genetic scores available in the literature: PRS-HFC, PRS-5, pPRS-discordant, pPRS-concordant and the PRS developed by Chen and colleagues (named here PRS-Chen).

The comparison of the performance of these validated PRSs with the newly developed ones for noninvasive scores of liver fibrosis was done for 3 main phenotypes of advanced MASLD in the Milan Biobank: fibrosis (N=1609 vs 5402), cirrhosis (N=554 vs 6211) and HCC (N=164 vs 6861).

The analysis on liver fibrosis highlighted that the PRS-Chen (AUC=0.6624, $p=4.45*10^{-81}$) has the higher predictive performance for this trait in the unadjusted model (i.e. when comorbidities were not included in the model), followed by PRS-5 (AUC=0.6485, $p=2.51*10^{-70}$) and PRS-HFC (AUC=0.6416, $p=1.48*10^{-63}$) (Figure 6A). The PRS-Chen was the best predictor of liver fibrosis also in the model adjusted for the main confounders (AUC=0.8829, $p=1.16*10^{-64}$), followed by PRS-5 (AUC=0.8813, $p=9.71*10^{-60}$) and PRS-HFC (AUC=0.8806, $p=7.56*10^{-60}$) (Figure 6B). The pairwise comparison analysis of PRSs predictivity showed that pPRS-Concordant - i.e. the score which assume the primary cause may be an increase in uptake and synthesis of energy substrates - had a lower predictivity than pPRS-Discordant - i.e. the score which assume the primary cause of higher liver TGs content is likely to be retention - (unadjusted: AUC=0.5474 vs AUC=0.6374 respectively, $p_{DE\ LONG}=1.48*10^{-15}$; adjusted: AUC=0.8682 vs AUC=0.8805 respectively, $p_{DE\ LONG}=5.56*10^{-7}$) in line with phenotype of lipid retention described for the pPRS-Discordant (Figure 6). In the unadjusted model, the ability of PRS-5 in the prediction of liver fibrosis was higher than

the PRS-HFC ($p_{DE\ LONG}=1.07*10^{-3}$) (Figure 6A) while it became comparable when the major confounders were included in the model ($p_{DE\ LONG}=0.143$) (Figure 6B). The PRSs newly developed for FIB-4 and FNI showed an overall predictivity which was below the one of the other scores available in the literature, with the exception of pPRS-Concordant for PRS-FNI (unadjusted: $p_{DE\ LONG}=3.03*10^{-7}$, Figure 6A; adjusted: $p_{DE\ LONG}=1.47*10^{-3}$, Figure 6B); when compared, PRS-FNI showed an higher AUC than PRS-FIB-4 both in the unadjusted (AUC=0.6049 vs AUC=0.5349 respectively, $p_{DE\ LONG}=7.55*10^{-11}$, Figure 6A) and adjusted

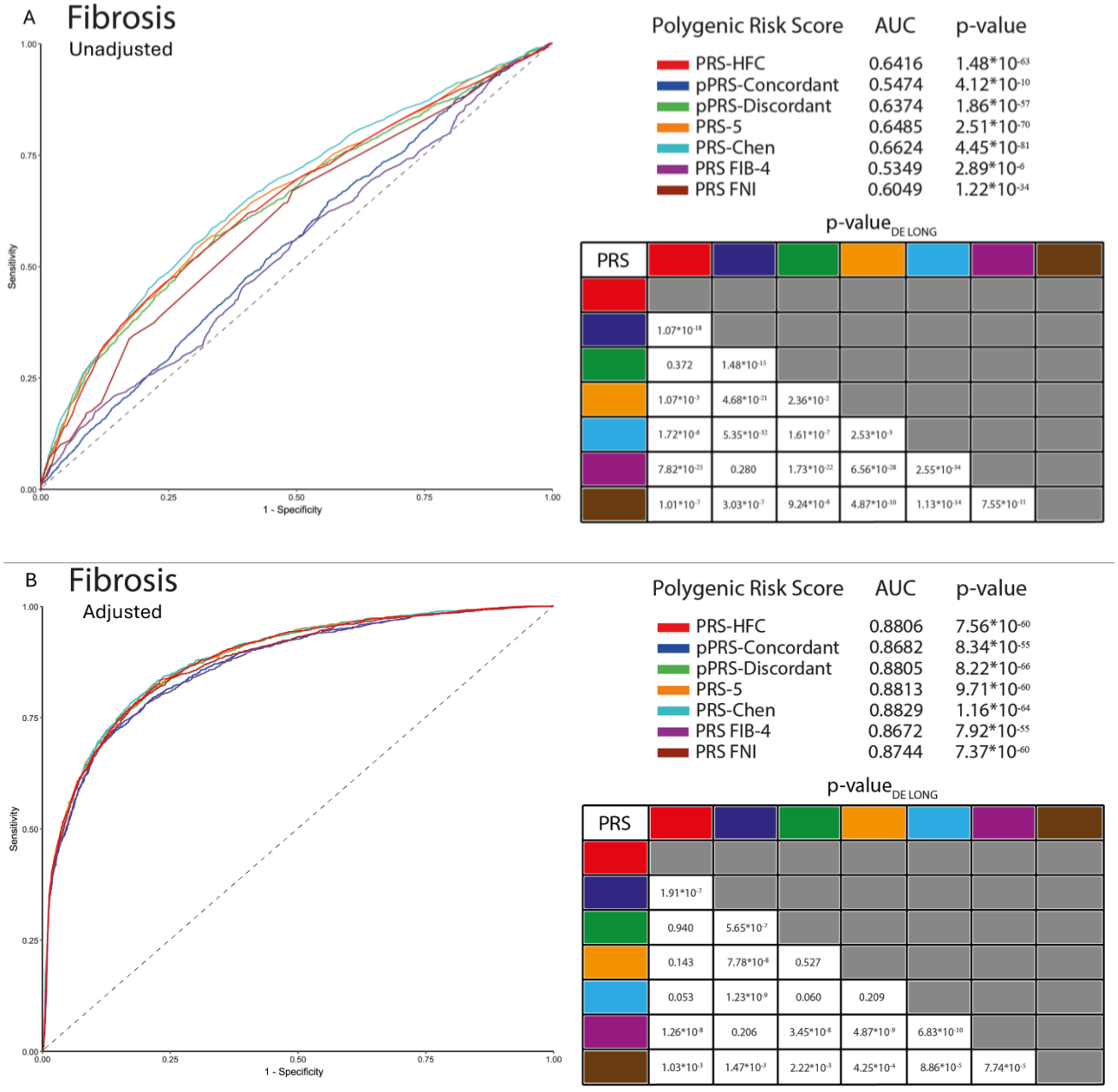

Figure 6. Comparison of the performance of the two newly developed PRS for FNI and FIB-4 with PRS available in literature considering both unadjusted (A) and adjusted (B) models for liver fibrosis (N=1609 vs 5402).

models (AUC=0.8744 vs AUC=0.8672 respectively, $p_{DE\ LONG}=7.74*10^{-5}$, Figure 6B). When cirrhosis was considered as outcome, PRS-Chen was shown to provide the higher AUC also for this trait in the unadjusted model (AUC=0.7098, $p=1.89*10^{-30}$), followed by PRS-5 (AUC=0.7035, $p=9.76*10^{-30}$) and pPRS-Discordant (AUC=0.7008, $p=1.81*10^{-30}$), even if all the three PRSs provided comparable performance in the prediction of the trait (PRS-Chen vs pPRS-Discordant: $p_{DE\ LONG}=0.187$; PRS-Chen vs pPRS-5: $p_{DE\ LONG}=0.356$; PRS-5 vs pPRS-Discordant: $p_{DE\ LONG}=0.710$) (Figure 7A). The adjusted model presented

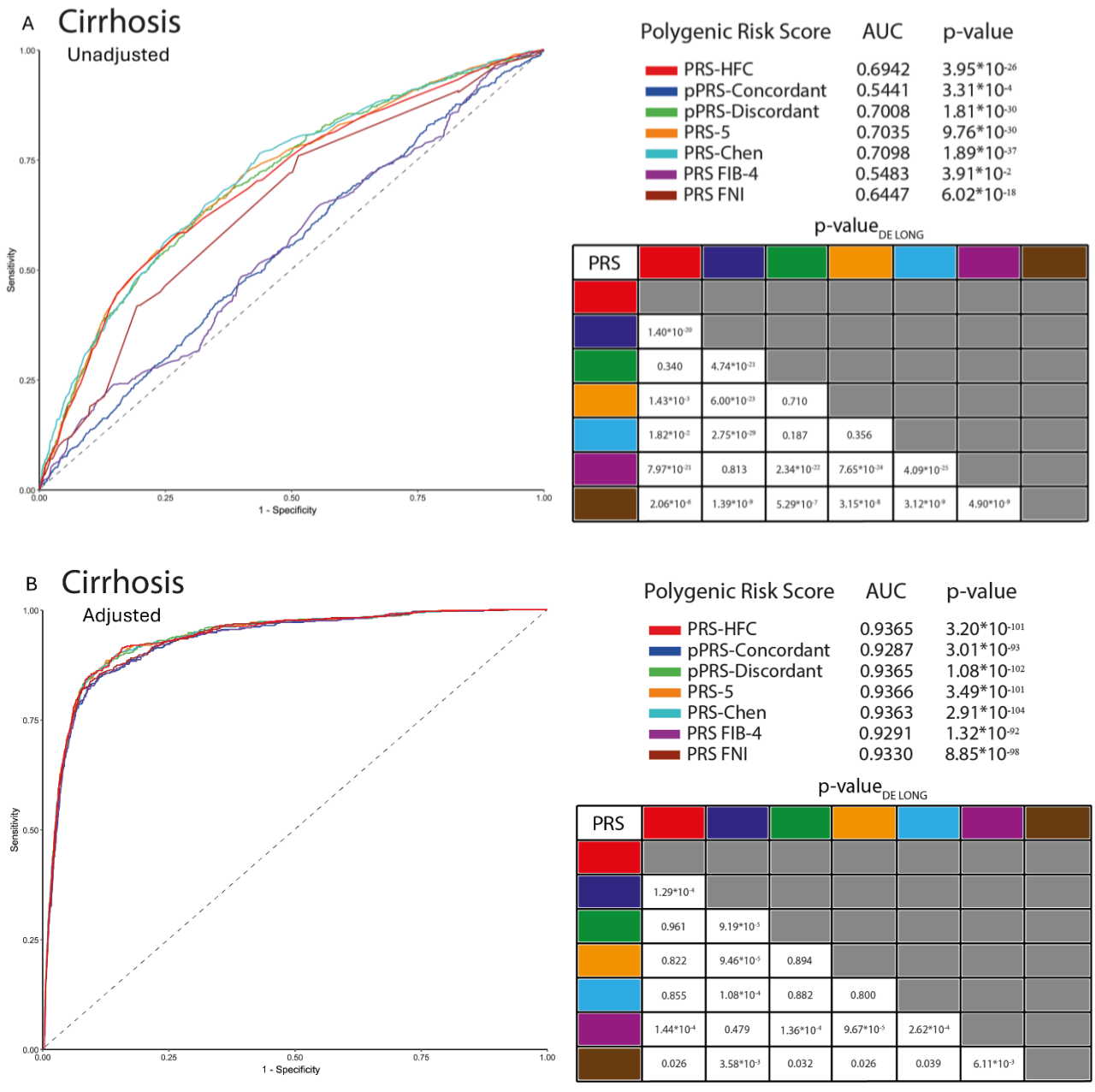

Figure 7. Comparison of the performance of the two newly developed PRS for FNI and FIB-4 with PRS available in literature considering both unadjusted (A) and adjusted (B) models for liver cirrhosis (N=554 vs 6211).

similar results but PRS-5 showing the higher AUC (PRS-5: AUC=0.9366, $p=3.49*10^{-101}$; pPRS-Discordant: AUC=0.9365, $p=1.08*10^{-102}$; PRS-HFC: AUC=0.9365, $p=3.20*10^{-101}$; PRS-Chen: AUC=0.9363, $p=2.91*10^{-104}$), with an overall matching performance for each of these PRS at the pairwise comparison analysis (Figure 7B). Even in this case, pPRS-Concordant had a lower predictivity than pPRS-Discordant (unadjusted: AUC=0.5441, $p_{DE\ LONG}=4.74*10^{-21}$; adjusted: AUC=0.9287, $p_{DE\ LONG}=9.19*10^{-5}$). Again, the PRSs for FIB-4 and FNI showed an overall predictivity which was below the one of the other scores available in the literature, with the exception of pPRS-Concordant for PRS-FNI (unadjusted: $p_{DE\ LONG}=1.39*10^{-9}$, Figure 7A; adjusted: $p_{DE\ LONG}=3.58*10^{-3}$, Figure 7B); when compared, PRS-FNI showed an higher AUC than PRS-FIB-4 both in the unadjusted (AUC=0.6447 vs AUC=0.5483 respectively, $p_{DE\ LONG}=4.90*10^{-9}$, Figure 7A) and adjusted (AUC=0.9330 vs AUC=0.9291 respectively, $p_{DE\ LONG}=6.11*10^{-3}$, Figure 7B) models; notably, PRS-FNI was significantly more performant in the prediction of cirrhosis likely due to the inclusion of *TM6SF2* SNP in the FNI score compared to PRS-FIB-4, which included only *PNPLA3* SNP. Finally, HCC was considered as outcomes. In the unadjusted model pPRS-Discordant showed the highest AUC (AUC=0.6812, $p=9.98*10^{-16}$) and resulted the best predictor together with PRS-5 (AUC=0.6631, $p=4.23*10^{-13}$) and PRS-Chen (AUC=0.6626, $p=5.00*10^{-13}$) at the pairwise comparison (PRS-Chen vs pPRS-Discordant: $p_{DE\ LONG}=0.144$; PRS-Chen vs pPRS-5: $p_{DE\ LONG}=0.972$; PRS-5 vs pPRS-Discordant: $p_{DE\ LONG}=0.131$) (Figure 8A). When corrected for the major confounders, the model revealed that pPRS-Discordant was the best predictor of HCC among the all the PRSs considered in the analysis (AUC=0.8666, $p=2.01*10^{-58}$). This score showed a better performance than PRS-5 (AUC=0.8666, $p_{DE\ LONG}=0.087$) and PRS-Chen (AUC=0.8621, $p_{DE\ LONG}=0.036$), which conversely showed at pairwise comparison similar predictivity ($p_{DE\ LONG}=0.583$) (Figure 8B). PRS-FNI showed an higher AUC than PRS-FIB-4 both in the unadjusted (AUC=0.6258 vs AUC=0.5680 respectively, $p_{DE\ LONG}=6.02*10^{-2}$, Figure 8A) and adjusted models, even if in the latter the

two showed a comparable predictivity, even if in the latter the two showed a comparable predictivity, probably due to the fact that the effect of confounders was likely masking the effect of the single SNPs included in the scores (AUC=0.8613 vs AUC=0.8589 respectively, $p_{DE\ LONG}$=0.451, Figure 8B).

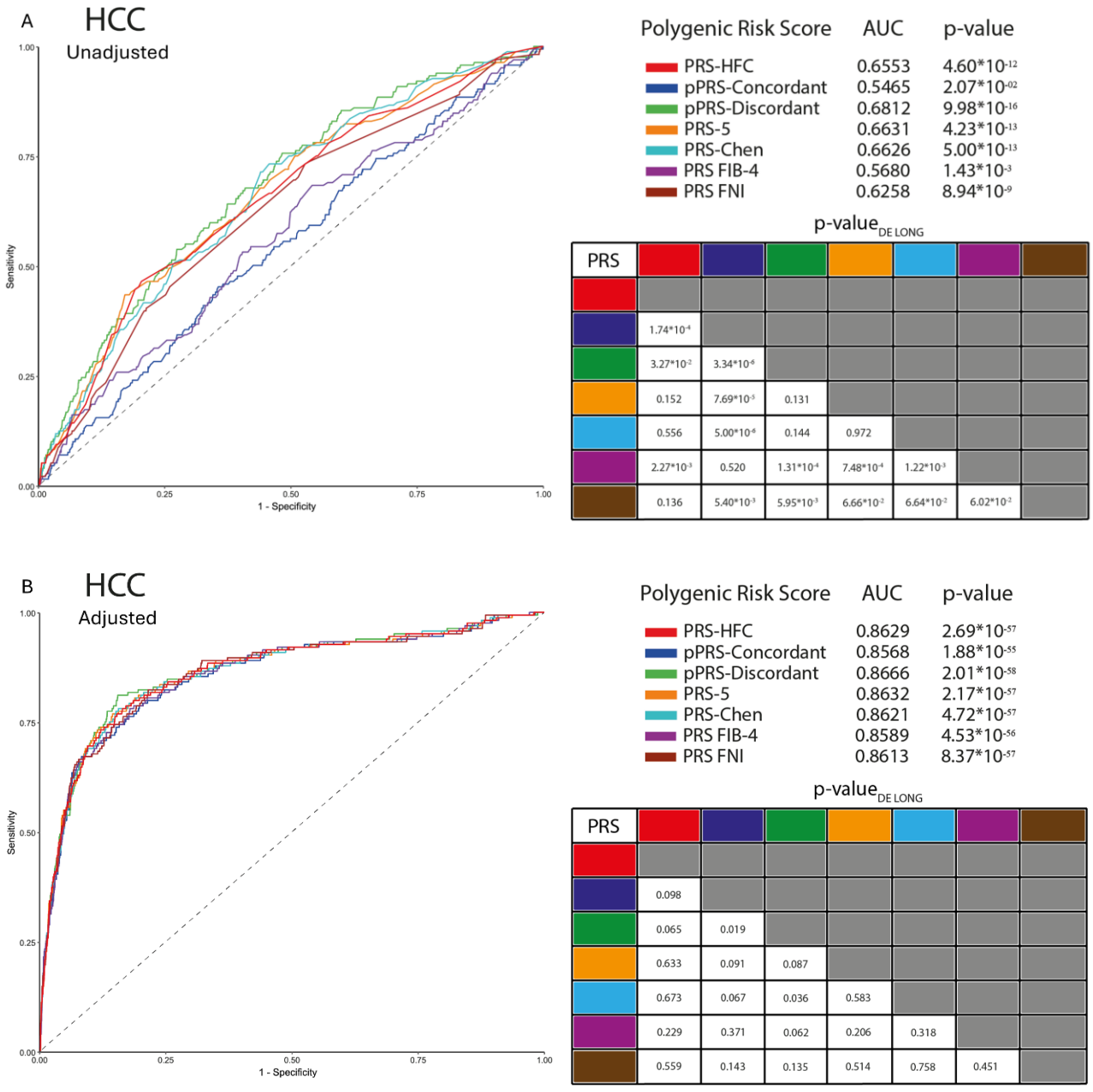

Figure 8. Comparison of the performance of the two newly developed PRS for FNI and FIB-4 with PRS available in literature considering both unadjusted (A) and adjusted (B) models for HCC (N=164 vs 6861).

Taken together, these results indicate that even though pPRS-discordant is a better marker of prediction that pPRS-concordant for each of the tested phenotype, with a particular mention to HCC, which maintain the higher AUC value with the highest significance.

All in all, the two novel PRS-FIB-4 and PRS-FNI developed from the GWAS summary statistics of the Milan BioBank showed in both models a performance which is at least comparable or better to the pPRS-Concordant for liver fibrosis (unadjusted: PRS-FIB-4 vs pPRS-Concordant: $p_{DE\ LONG}$=0.280; adjusted: PRS-FIB-4 vs pPRS-Concordant: $p_{DE\ LONG}$=0.206; unadjusted: PRS-FNI vs pPRS-Concordant: $p_{DE\ LONG}$=3.03*10^{-7}; adjusted: PRS-FNI vs pPRS-Concordant: $p_{DE\ LONG}$=1.47*10^{-3}) (Figure 6), for liver cirrhosis (unadjusted: PRS-FIB-4 vs pPRS-Concordant: $p_{DE\ LONG}$=0.813; adjusted: PRS-FIB-4 vs pPRS-Concordant: $p_{DE\ LONG}$=0.479; unadjusted: PRS-FNI vs pPRS-Concordant: $p_{DE\ LONG}$=1.39*10^{-9}; adjusted: PRS-FNI vs pPRS-Concordant: $p_{DE\ LONG}$=3.58*10^{-3}) (Figure 7) and for HCC (unadjusted: PRS-FIB-4 vs pPRS-Concordant: $p_{DE\ LONG}$=0.520; adjusted: PRS-FIB-4 vs pPRS-Concordant: $p_{DE\ LONG}$=0.371; unadjusted: PRS-FNI vs pPRS-Concordant: $p_{DE\ LONG}$=5.40*10^{-3}; adjusted: PRS-FNI vs pPRS-Concordant: $p_{DE\ LONG}$=0.143) (Figure 8).

All in all, although all the tested PRSs can reliably predict all the phenotypes of advanced liver disease, there is still the need of a PRS which is specific for the etiology of liver insults in order to discriminate specific sub- and endophenotypes within the context of liver disease.

Phenome wide association study of *MRC1* rs2461143 T>C variant

To investigate possible mechanism of association between the *MRC1* rs2461143 T>C variant and FIB-4, a PheWAS was carried out taking into consideration 211 phenotypes collected and available in the Milan Biobank, adjusting for age, biological sex, T2D status and *PNPLA3* rs738408 C>T variant.

Trait	Variant	β [95% C.I.]	p	padj
AST	MRC1 rs2461143 T>C	0.11 [0.08;0.14]	2.39E-12	5.05E-10
HSI	MRC1 rs2461143 T>C	-0.05 [-0.06;-0.03]	2.82E-09	2.97E-07
FIB-4	MRC1 rs2461143 T>C	0.07 [0.04;0.09]	5.29E-09	3.72E-07
FAST score	MRC1 rs2461143 T>C	0.10 [0.06;0.14]	4.33E-06	2.28E-04
FNI score	MRC1 rs2461143 T>C	0.09 [0.04;0.13]	8.79E-05	0.004
Glutamic acid	MRC1 rs2461143 T>C	0.10 [0.03;0.17]	0.005	0.165
TSH	MRC1 rs2461143 T>C	0.08 [0.01;0.15]	0.018	0.484

Table 6. PheWAS of MRC1 rs2661143 T>C variant with the phenotypes available in the Milan Biobank.

Among the significant findings, a positive association with marker of liver inflammation was found such as AST levels (β [95% C.I.]= 0.11 [0.08;0.14], FDR= $5.05*10^{-10}$), FAST score (β [95% C.I.]= 0.10 [0.06;0.14], FDR= $2.28*10^{-4}$) and FNI (β [95% C.I.]= 0.09 [0.04;0.13], FDR=$4*10^{-3}$) (Table 6).

Conversely, an inverse association with HSI and the variant (β [95% C.I.]=-0.05 [-0.06;-0.03], FDR=$2.97*10^{-7}$) was reported, suggesting a mechanism of action and inflammation leading to hepatic fibrosis which is at least primary decoupled from lipid content of the liver (Table 6). Finally, increasing levels of glutamic acid (β [95% C.I.]=0.10 [0.03;0.17], p=$5*10^{-3}$) and TSH (β [95% C.I.]=0.08 [0.01;0.15], p=$1.80*10^{-2}$) were nominally associated with the *MRC1* rs2461143 T>C variant (Table 6).

Impact of *MRC1* genotype on gene expression

Given the overall results obtained from GWASs of noninvasive scores (FIB-4 and FNI) of liver fibrosis in the Milan Biobank, the GTEx portal was interrogated to assess if any of the genetic variant associated with these traits could represent an expression quantitative trait locus (eQTL), i.e. whenever any of these genetic variations was associated with changes in the respective gene expression levels, in the liver. The interrogation of GTEx portal revealed that none of these variant represented an eQTL in liver tissue (Table 7).

SNP	Gene	p	NES
rs10145931	IFI27	0.610	0.052
rs56255430	TM6SF2	0.120	-0.100
rs2461143	MRC1	0.950	0.003
rs738408	PNPLA3	0.800	-0.015

*Table 7. Evaluation in GTEx public repository of possible eQTL in liver tissue taking into consideration significative (p<5*10-8) variants associated with noninvasive scores of liver fibrosis.*

Then, given that t both *MRC1* and *IFI27* are well known to be expressed primarily in immune cells and inflammatory cells, and since these variants, not in LD with any missense variants altering the protein sequence, resulted significantly associated (p≤$5*10^{-8}$) with FIB-4 and FNI from GWAS, whole blood entry was explored to assess if these variants, namely *MRC1*

rs2461143 T>C and *IFI27* rs10145931 T>C, might represent potential eQTL in this tissue.

The *MRC1* rs2461143 T>C risk variant, but not *IFI27* rs10145931 T>C (Figure 9A, p=0.307), represented an eQTL, being associated with a decrease of *MRC1* expression in whole blood tissue (Figure 9B, $p=2.37*10^{-5}$).

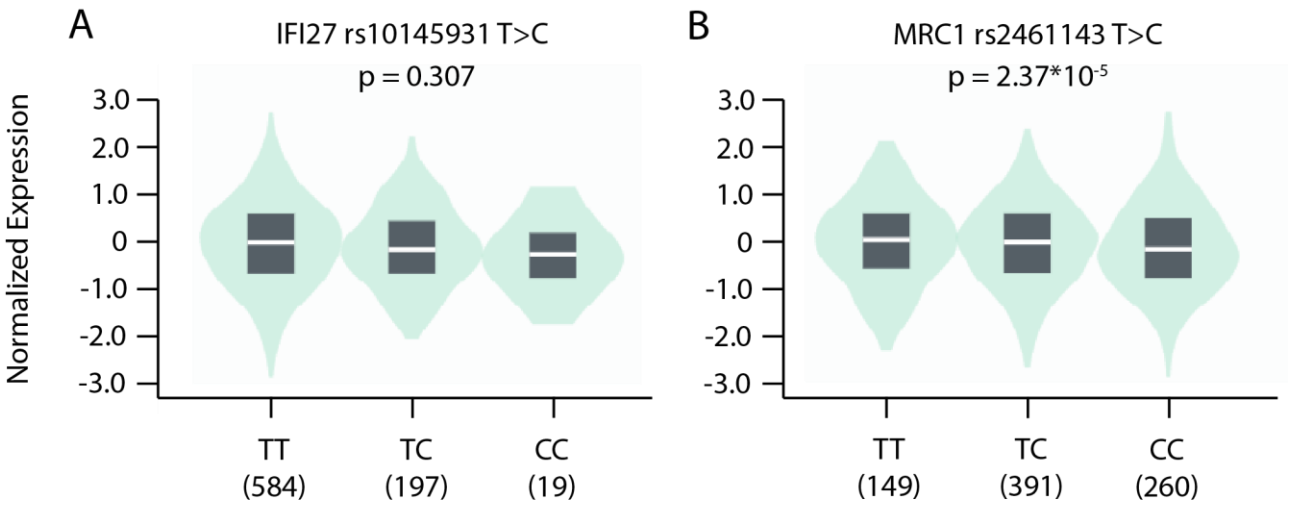

Figure 9. MRC1 rs2461143 T>C genetic variation represents an eQTL in whole blood tissue in GTEx portal.

MRC1 is primarily expressed in immune cells, and its protein product, which consist in a pattern recognition receptor, is found on the surface of LSECs, macrophages and dendritic cells. It plays a crucial role in the innate immune system, taking part in processes like the regulation of inflammation, tissue homeostasis and repair, and liver fibrosis in principle via fibrotic remodeling by cytokines secretion, which, in turn, promotes the activation of HSCs triggering collagen deposition[71,72].

MRC1 rs2461143 T>C variant associated with increased ballooning in the Milan BioBank

To evaluate the impact of the variant on the morphological features of liver tissue, the association of the variant with histological features in the biopsy cohort of the Milan BioBank was tested (N=437), considering as main outcomes the steatosis, ballooning and necroinflammation grade, fibrosis stage and the NAS.

Hepatocellular ballooning resulted associated with the rs2461143 T>C variant (β [95% C.I.]=0.08 [0.01;0.16], p=0.03) suggesting its involvement in hepatocytes degeneration anticipating the inflammatory processes (Table 8).

Histological Trait	Variant	β [95% C.I.]	p
Steatosis, grade	MRC1 rs2461143 T>C	-0.05 [-0.18;0.07]	0.407
Ballooning, grade	MRC1 rs2461143 T>C	0.08 [0.01;0.16]	0.030
Necroinflammation, grade	MRC1 rs2461143 T>C	0.01 [-0.07;0.09]	0.803
Fibrosis, stage	MRC1 rs2461143 T>C	0.03 [-0.04;0.12]	0.394
NAFLD Activity Score (NAS)	MRC1 rs2461143 T>C	0.04 [-0.17;0.26]	0.698
NASH, yes	MRC1 rs2461143 T>C	-0.13 [-.046;0.19]	0.420

Table 8. Association study in biopsy cohort to evaluate the association of MRC1 rs2461143 with histological features of the liver (N=437) in the Milan Biobank.

MRC1 is downregulated in patients with advanced liver disease

Then, given the association of the *MRC1* rs2461143 T>C variant and reduced expression of *MRC1 in immune cells*, the regulation of the gene in response to liver damage features was investigated in a transcriptomic cohort composed of 125 people who underwent bariatric surgery. In this cohort, *MRC1* downregulation was associated with the presence of ballooning (Figure 10B, p=0.007, N=15 vs 110 controls), necroinflammation (Figure 10C, p=0.018, N=63 vs 62 controls) and MASH (Figure 10D, p=0.016, N=14 vs 111 controls), but not with hepatic steatosis (Figure 10A, p=0.148, N=104 vs 21 controls).

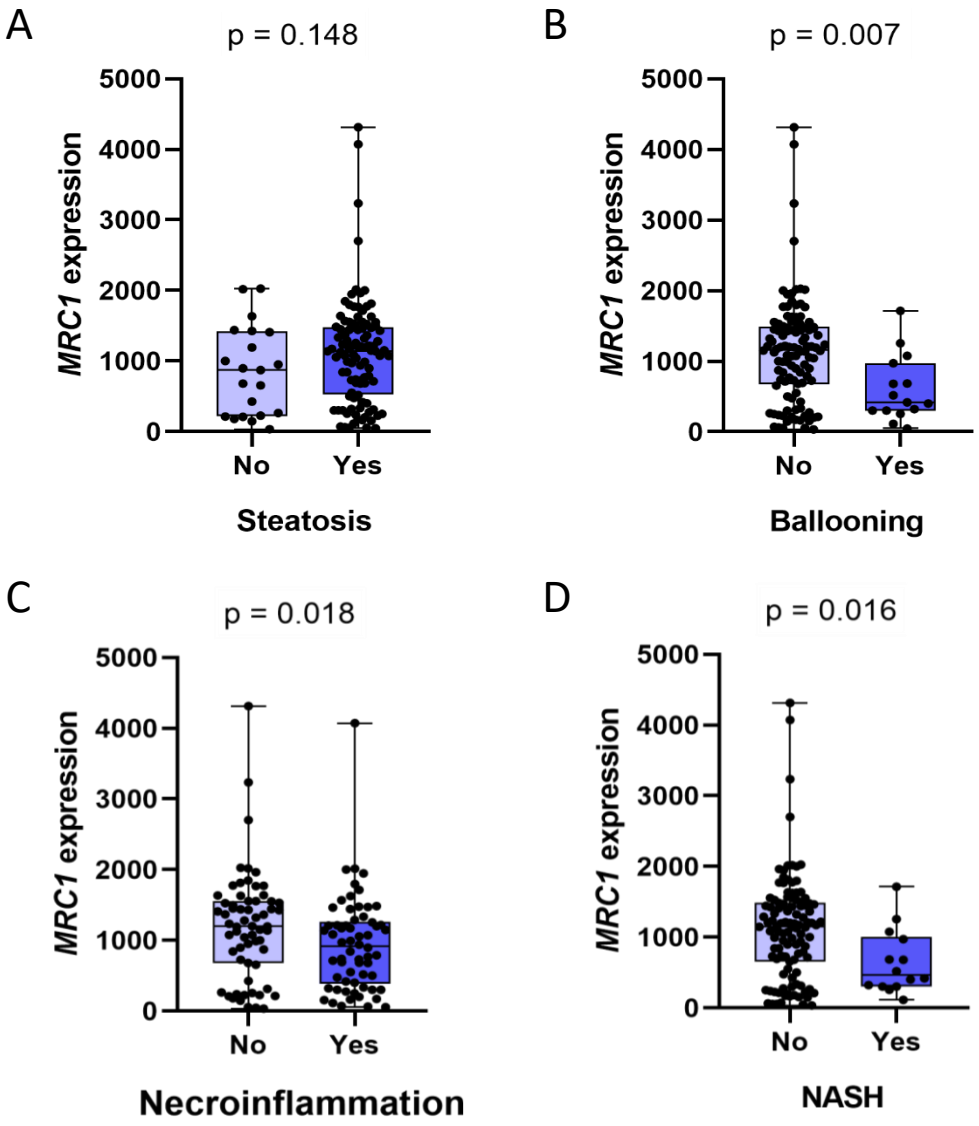

Figure 10. The downregulation of MRC1 gene is associated with hepatocellular ballooning (p=0.007),necroinflammation of the liver (p=0.018) and NASH status (p=0.016) in a transcriptomic cohort composed by 125 bariatric patients.

Furthermore, in the whole cohort, in which the sequencing data of 125 bariatric patients[68] were joined together with additional 261 MASH patients[67] were considered (N=341), increasing fibrosis stage was also associated with *MRC1* downregulation (p=0.044) (Figure 11).

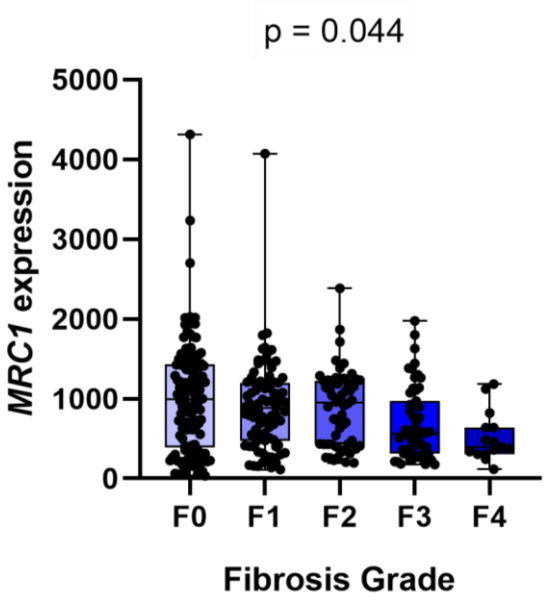

Figure 11. The downregulation of MRC1 gene is also associated with liver fibrosis grade (p=0.044) in an extended transcriptomic cohort composed by 341 MASH patients.

MRC1 downregulation is associated with signature of inflammation at the transcriptomic level

To pinpoint the molecular signatures underling the downregulation of *MRC1*, a transcriptomic analysis was performed between patients with higher *MRC1* and low *MRC1* expression (N=341, 228 vs 113 subjects respectively). At transcriptomic level, this differential expression analysis highlighted the alteration of genes involved in immune modulation and inflammatory pathways in patients with reduced *MRC1* expression: a total of 162 genes were differentially expressed, 126 downregulated and 36 upregulated (Figure 12).

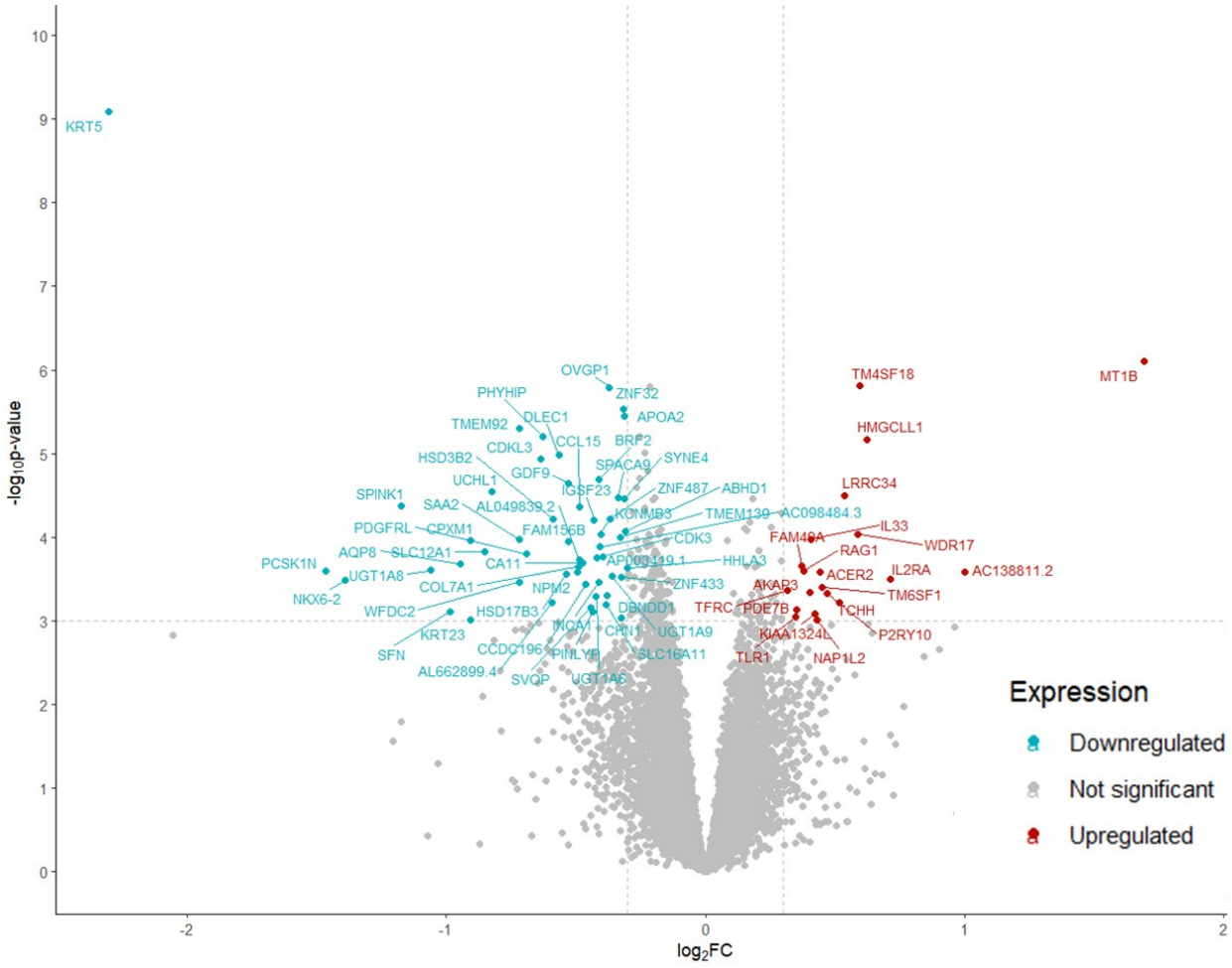

Figure 12. Differential expression analysis in the extended MASH cohort (N=341) revealed the alteration of genes involved in immune modulation and inflammatory pathways in patients with reduced MRC1 expression. A total of 162 genes were differentially expressed, 126 downregulated and 36 upregulated.

The top 10 genes significantly upregulated were *MT1B* (log_2 fold-change=1.69, p=0.005), encoding the a metal-binding protein involved in oxidative stress defense and inflammation; *AC138811.2* (log_2 fold-change=1.00, p=0.038) is a long non-coding RNA of unknown function but located near immune-related loci; *IL2RA* (log_2 fold-change=0.71, p=0.040) encoding the alpha subunit of the interleukin-2 receptor, essential for lymphocyte activation and regulatory T cell function; *HMGCLL1* (log_2 fold-change=0.62, p=0.009) participating in leucine catabolism and likely to influence lipid and energy metabolism during inflammation; *TM4SF18* (log_2 fold-change=0.59, p=0.005) potentially mediating immune cell adhesion;

Gene	log_2 fold-change	padj
MT1B	1.69	0.005
AC138811.2	1.00	0.038
IL2RA	0.71	0.040
HMGCLL1	0.62	0.009
TM4SF18	0.59	0.005
WDR17	0.59	0.029
LRRC34	0.54	0.020
P2RY10	0.47	0.046
TM6SF1	0.45	0.043
ACER2	0.44	0.038
KRT5	-2.30	1.212E-05
PCSK1N	-1.47	0.038
NKX6-2	-1.39	0.040
SPINK1	-1.17	0.023
UGT1A8	-1.06	0.038
AQP8	-0.94	0.037
CPXM1	-0.91	0.029
SLC12A1	-0.85	0.033
UCHL1	-0.83	0.020
TMEM92	-0.72	0.009

Table 9. Top 10 genes significantly upregulated and downregulated in the extender MASH cohort (N=341).

WDR17 (log_2 fold-change=0.59, p=0.029) linked to cell cycle regulation; *LRRC34* which modulates nucleic acid interactions; *P2RY10* (log_2 fold-change=0.47, p=0.046), a G-protein–coupled receptor responsive to lysophospholipids, mediating immune cell trafficking; *TM6SF1* (log_2 fold-change=0.45, p=0.043) encoding a transmembrane protein implicated in lipid metabolism and endoplasmic reticulum function; and *ACER2* (log_2 fold-change=0.44, p=0.038) which regulates sphingolipid metabolism and inflammatory signaling

(Table 9).

Conversely, patients with low *MRC1* expression displayed reduced expression of several protective or structural genes: *KRT5* (log_2 fold-change=-2.30, p=1.21*10^{-5}) encoding keratin 5, which is important for epithelial structural integrity; *PCSK1N* (log_2 fold-change=-1.47, p=0.038) that modulates cytokine processing; *NKX6-2* (log_2 fold-change=-1.39, p=0.040), a transcription factor involved in cell differentiation; *SPINK1* (log_2 fold-change=-1.17, p=0.023) exerting a protective role against protease-mediated tissue injury; *UGT1A8* (log_2 fold-change=-1.06, p=0.038), which is involved in detoxification; *AQP8* (log_2 fold-change=-0.94, p=0.037) encoding aquaporin-8, a water channel protein expressed in hepatocytes and biliary epithelium; *CPXM1* (log_2 fold-change=-0.91, p=0.029) which is involved in extracellular matrix remodeling; *SLC12A1* (log_2 fold-change=-0.85, p=0.033) encoding a sodium-potassium-chloride cotransporter important in ion homeostasis; *UCHL1* (log_2 fold-change=-0.83, p=0.020) is a ubiquitin carboxy-terminal hydrolase involved in protein turnover and cellular stress responses and *TMEM92* (log_2 fold-change=-0.72, p=0.009) encodes a poorly characterized transmembrane protein (Table 9).

Overall, this transcriptional profile in individuals with low *MRC1* expression suggests a shift toward chronic immuno-inflammatory processes, characterized by upregulation of immune activation and signaling genes alongside downregulation of epithelial integrity and tissue-protective factors, potentially contributing to hepatic fibrogenesis.

Then, considering this set, a GSEA in patients with low *MRC1* expression, revealed the upregulation of inflammatory pathways such as the signaling by interleukins (Normalized Enrichment Score=1.84, p=0.01) and cytokine signaling in immune system (Normalized Enrichment Score=1.68, p=0.036) (Figure 13).

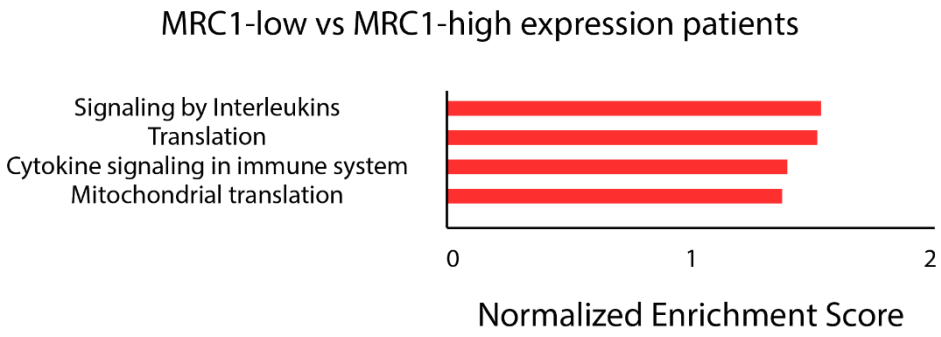

Figure 13. Gene set enrichment analysis (GSEA) in patients with low MRC1 expression, revealed the upregulation of inflammatory pathways in the extended MASH cohort (N=341)

To zoom into the cell populations which express the *MRC1* gene then scRNA-Seq dataset including 3 healthy individuals and 3 MASH patients was analyzed. In this cohort, only LSECs and immune cells expressed the gene and only these two populations were considered to carry out further analyses. In keeping with bulk transcriptomic data, both immune cells and LSECs cells showed a decrease in *MRC1* expression in people with MASH patients compared to those with normal liver ($p=2.2*10^{-16}$ and $p=6.6*10^{-7}$ respectively) (Figure 14).

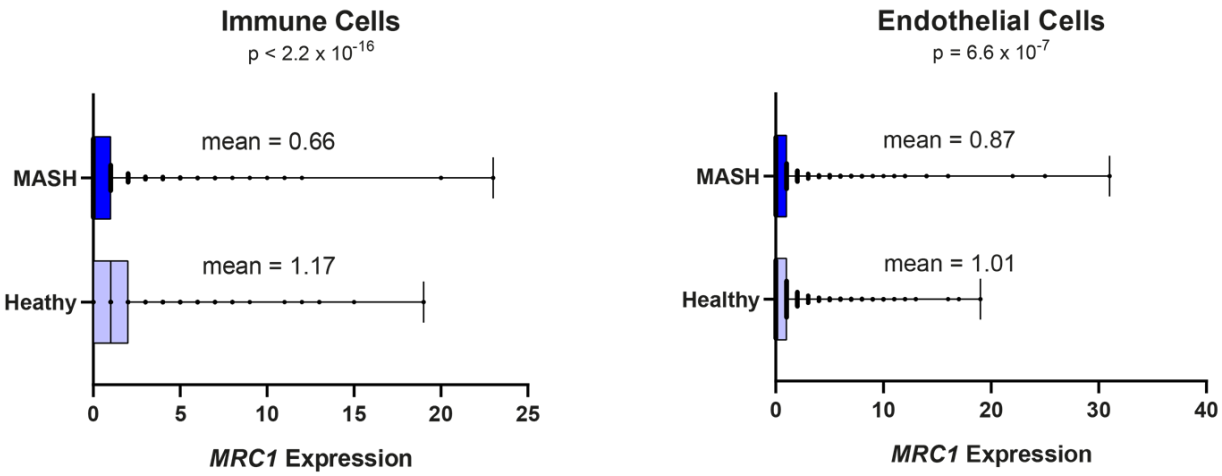

Figure 14. MRC1 expression comparison between healthy (N=3) and MASH (N=3) individuals in immune and endothelial cells (LSECs) with single cell resolution.

Finally, an overrepresentation analysis (ORA) revealed also markers of inflammation in both the cell populations: in LSECs the low expression of *MRC1* correlated with an overrepresentation of genes involved in epithelial to mesenchymal transition (EMT, OR=3.24, FDR=4.21*10^{-7}), TNF-α signaling via NFK-β (OR=2.13, FDR=7.04*10^{-4}) and inflammatory response (OR=1.87, FDR=0.038) (Figure 15).

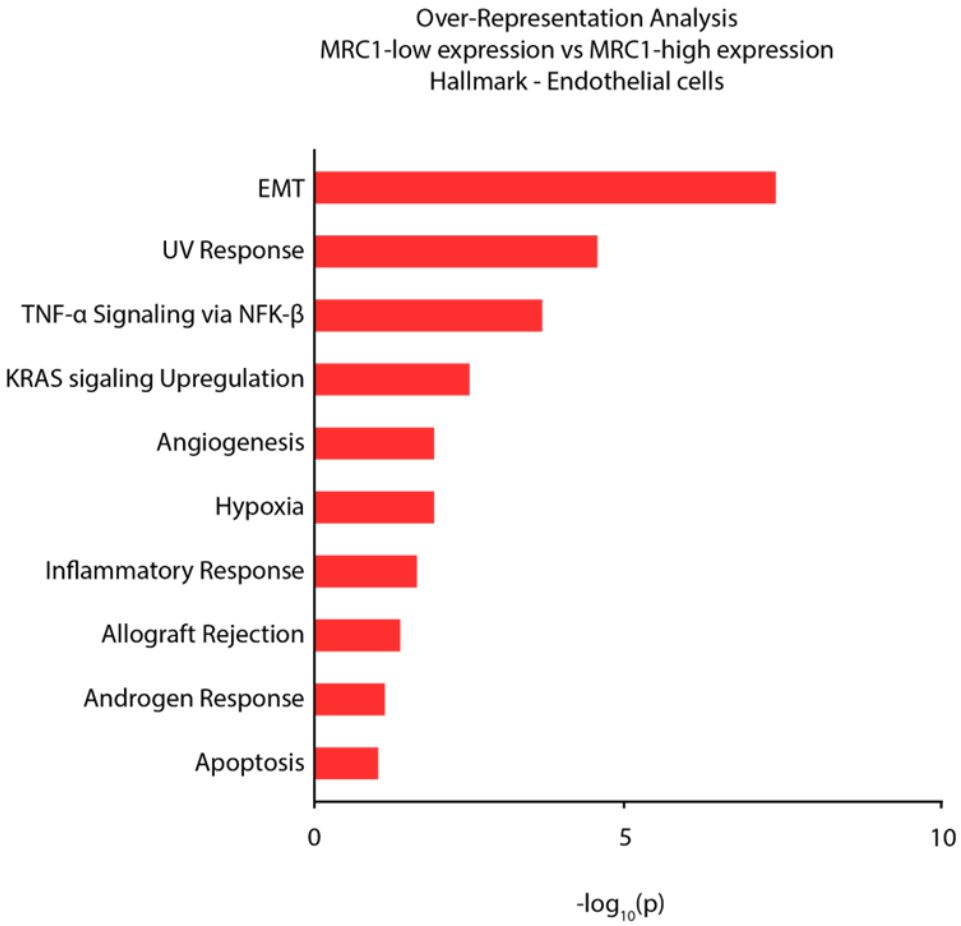

Figure 15. Overrepresentation analysis (ORA) revealed enhanced pathways of inflammation in endothelial cells (LSECs).

Similarly, in immune cells the low expression of *MRC1* correlated with an overrepresentation of genes involved in EMT (OR=2.53, FDR=0.002), inflammatory response (OR=2.32, FDR=0.006), TNF-α signaling via NFK-β (OR=2.01, FDR=0.022), IL-6/JAK/STAT3 signaling (OR=2.47, FDR=0.036) and complement (OR=1.81, FDR=0.049) (Figure 16).

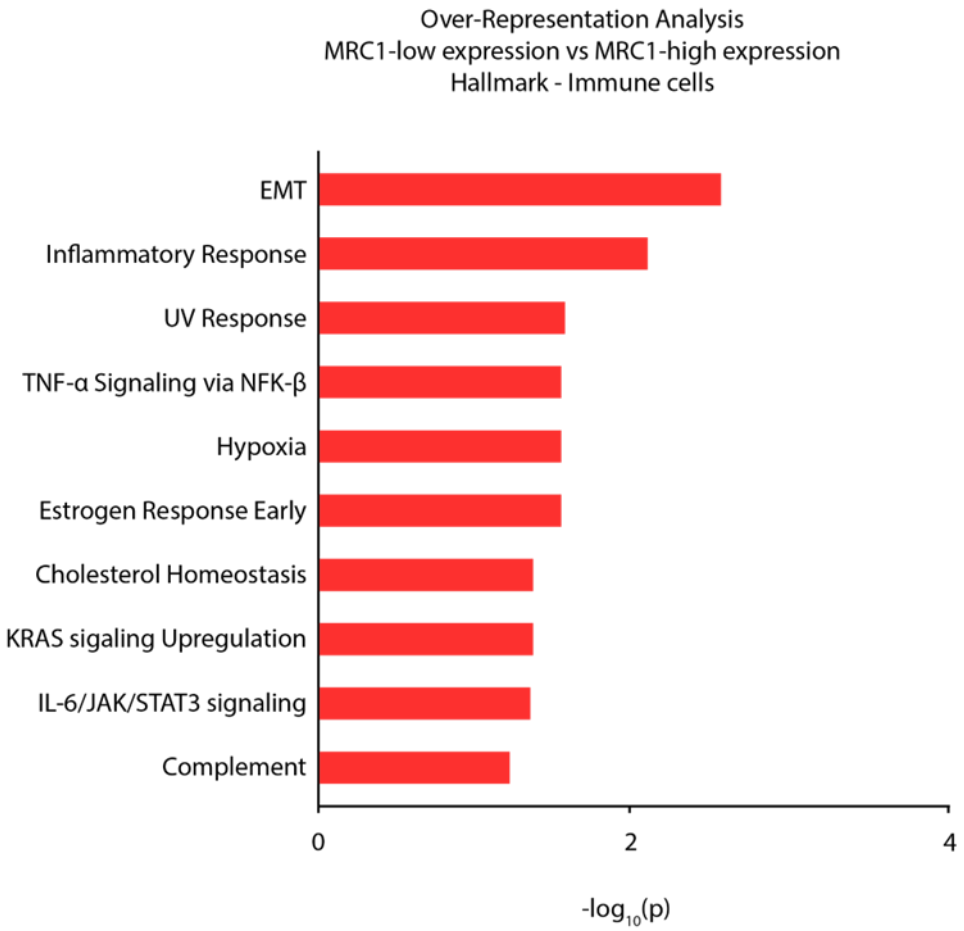

Figure 16. Overrepresentation analysis (ORA) revealed enhanced pathways of inflammation in immune cells.

All in all, these data suggest that *MRC1* downregulation may be causally linked to increased propensity to transition from steatosis to fibrotic steatohepatitis.

Discussion

Here was introduced the Milan Biobank, the first Italian genetic biobank enriched in patient affected by MASLD. The main findings were a) the validation of known loci (*PNPLA3*, *TM6SF2*) b) the validation and multi-omic characterization of *MRC1* as a new MASLD fibrosis-associated locus, involving immune regulation and fibrotic remodeling; c) the evidence, via PRSs analysis, that systemic and liver-specific pathways of MASLD progression can be genetically untangled.

The current investigation sought to deconstruct the genetic foundation of liver fibrosis in MASLD via a cross-sectional GWAS in the Milan Biobank, a substantial multicenter Italian cohort. Our investigation validated strongly established loci (*PNPLA3*, *TM6SF2*) and revealed novel associations at *MRC1*, *IFI27* and *DIAPH2*. These results underscore the complementary utility of national biobank-based studies and broaden the knowledge of biological mechanisms involved in fibrotic development in MASLD. The significant association of the loci with highly reported signatures of liver damage represents a positive control for the reliability of the analysis in the Milan Biobank; in fact, genetic variations at *PNPLA3* are known to confer a markedly increased risk of increasingly severe histological features of SLD, without a strong effect on metabolic syndrome component traits[75], and mutation in *TM6SF2*, involved in the loading of TGs on the nascent APOB, are known to predispose to liver fibrosis by altering the lipid clearance from hepatocytes and by hindering the regulation of lipid metabolism and VLDL[33–35]. The GWAS approach allowed to pinpoint a novel locus of association for circulating liver enzymes; in fact, although *LRRIQ3* is generally poorly expressed in liver with limited data connecting this gene to hepatocyte biology or HCC and even if literature involves, rather than *IL1RAPL1* specifically, other IL-1 family members and accessory genes in explaining the importance of the IL-1 signaling axis in liver inflammation and HCC, GWAS on ALT highlighted the role of *VMA21* to the onset of

liver damage, since it is well known that this gene contributes to hepatic steatosis, defective lipophagy, ER stress and altered cholesterol metabolism[76].

The results obtained by the analysis of other circulating liver enzymes underpins the consistency of the approach used to setup the overall cohort in the Milan Biobank; in liver, AST was associated, beyond *PNPLA3* and *MRC1*, with *NFKB1* which is ubiquitously expressed in the liver, including hepatocytes, KCs, and HSCs and plays a pivotal role as central mediator of inflammation, immunity, and survival signaling, especially in response to TNF-α, IL-1β, LPS, and oxidative stress[77]. Interestingly, NF-κB function in the liver has a double-edged personality: acute activation is protective and promotes regeneration, but chronic activation drives inflammation, fibrosis, and cancer[77,78]. Therapeutic intervention therefore aims at selective inhibition in specific liver cell types to prevent chronic disease without sacrificing regeneration or immunity.

RORA regulates genes involved in lipid metabolism such as apolipoproteins *APOA1*, *APOA5*, *APOC3* and *PPARG* and it has specific and redundant functions as modulator of expression of genes encoding phase I and phase II proteins involved in the metabolism of lipids[79]. *EXOC3L4* as well is expressed in the liver, and it represents another genetic link to variation in liver enzymes previously reported in literature[54]. In addition, *LPL* gene has been reported to cause hyperlipoproteinemia and lipid metabolism disorders[80,81]. Furthermore, the direction of the association of variants associated with circulating CHO, HDL, LDL and, more importantly, TGs is consistent with the hypothesis of the existence of uncoupling mechanisms responsible for the higher liver TGs which is attributable to retention of liver fat in the liver rather than to an increase in uptake and synthesis of energy substrates[48].

Another results supporting this statement and the reliability of the overall analysis in the Milan Biobank in the significance of *PNPLA3* locus as top hit of non-invasive scores of liver fibrosis; in fact PNPLA3-I148M significantly promoted intracellular free cholesterol aggregation in LX-2 cells by decreasing cholesterol efflux protein (ABCG1) expression,

inducing mitochondrial dysfunction and structural damage through the accumulation of free cholesterol, thereby promoting the activation of LX-2 cells and the development of liver fibrosis[51,82]. The significance of this locus in the association analysis of circulating GGT levels may be attributed to the fact that the Milan Biobank is enriched in patients carrying alteration in the *PNPLA3* gene concomitantly with high levels of GGT, partially introducing a slight bias without anyway masking the effect of *GGT1* gene which represents the leading locus of this trait in the Italian population[82].

An additional contribution to this work is the PRSs for MASLD-associated traits. By examining the PRS, established and new loci of association were identified, contributing to the complexity indicators of fibrosis. Significantly, refined PRSs analyses indicate that a systemic understanding and liver-specific pathways may be unlinked, allowing for a better conceptualization of the disease's complexity. Patients with mostly systemic genetic risk factors are likely to advance through cardiometabolic comorbidities, while those with liver-specific risk factors would have a direct hepatic risk disease pathway. This kind of stratification is extremely useful in clinical practice, especially for providing appropriate therapies, as well as in clinical surveillance, in addition to therapeutic management and clinical decision-making. Besides this, the calculation of the two PRSs for FIB-4 and FNI showed promising results; in fact, the comparison of the performance of the validated PRSs already available in the literature with the newly developed PRSs for noninvasive scores of liver fibrosis against the main unfavorable outcomes of advanced MASLD in the Milan Biobank showed consistent and, even if slightly lower, reliable predictivity of liver fibrosis, cirrhosis and HCC. All in all, although all the tested PRSs can reliably predict all the phenotypes of advanced liver disease, there is still the need of a PRS which is specific for the etiology of liver insults in order to discriminate specific sub- and endophenotypes within the context of liver disease.

The main novelty among this study finding is the novel association highlighted between

MRC1 locus and FIB-4 in the Milan Biobank (rs2461143 T>C), validated and replicated also in the in an independent International biobank, the UKBB (rs56278466 T>G), whose signal comparison through LD analysis showed poor linkage between the two SNP, indicating the reliability of the association results obtained for this locus. Among the new associations, the *MRC1* locus stood out as of particular interest. *MRC1* encodes the mannose receptor (CD206), a pattern recognition receptor found on macrophages and LSECs, where it is involved in endocytosis of glycoproteins, clearance of collagen fragments, and modulation of immune activation[71,72].

Our results show that inherited variation at this locus affects fibrosis risk, implicating a genetic component to immune-mediated tissue remodeling. Furthermore, through the characterization of the gene exploiting publicly available and local transcriptomic data, was shown the correlation of the low expression of this gene and the phenotypes which are representative of advanced liver disease. Transcriptomic data corroborated this inference, as decreased *MRC1* expression was observed in patients with severe fibrosis and hepatocellular ballooning, and functional studies implicated *MRC1*-related pathways in the regulation of cytokine secretion and HSCs activation. This combined evidence from GWAS, replication, and multi-omics studies firmly implicates a central role of *MRC1* in fibrotic remodeling, making it both a mechanistic candidate and a putative biomarker for disease stratification. Furthermore, taken together, this data suggest a possible mechanism of action for *MRC1* during liver fibrosis, which is decoupled from steatosis, as indicated by the results obtained from the analysis of histological records in the Milan Biobank. All in all, *MRC1* could possibly take part within the liver in restraining the inflammatory process driven by immune cells and LSECs, as suggested by the marker of inflammation obtained from the transcriptomic analysis.

Limitations

This study shows several limitations. First, the study is cross-sectional, which means that the genetics of fibrosis markers can be studied only at a single time point and used to hypothesize about disease progression. This hypothesis can then be tested through prospective studies. Though the Milan Biobank cohort (7,025 participants) is large, it is modest compared to international resources, which limits the power to detect small-effect variants and makes replication in other cohorts difficult. Even with quality control and adjustments for ancestry, it is still possible that there is some confounding left over from population stratification. Compared to biopsy, scores like FIB-4 and FNI might incorrectly classify fibrosis, and associated variants cannot be fully understood without functional interpretation, which requires experimental validation. Insights can be gained through PRS, but they are still not accurate enough to be of clinical value. In spite of these issues, the study validates the idea of GWAS in national biobanks and finds new candidate loci for MASLD fibrosis progression through it, which is a good starting point for future functional and translational research.

Future perspectives

The final approach to complement this study will include the demonstration of associations between alterations in *MRC1* gene with liver related phenotypes in larger cohorts by meta-analysis and PheWAS. The final proof of *MRC1* involvement in liver fibrosis onset and development by enhancement of inflammatory process will be held by replication and validation of the results in wider international histological cohorts such as LITMUS.

Conclusions

This work conducted the first GWAS on non-invasive fibrosis markers in an Italian cohort, replicating and expanding MASLD's known risk loci. FIB-4 and FNI are traits which were not previously analyzed in published GWAS about MASLD, thereby providing through the genetic validation with the two main inherited determinants of MASH (*PNPLA3* and *TM6SF2*) an indirect support to the accuracy of this simple score for early screening of fibrotic MASH. The findings confirm the impact of *TM6SF2* and *PNPLA3* in liver disease, while bringing attention mainly to *MRC1* as newly proposed locus, by combining evidence from GWAS, replication in the UKBB, and multi-omics studies, pinpointing its biological meaningfulness and increasing the likelihood of its involvement in the fibrotic remodeling. This work also demonstrates the benefit of a national biobank brings in developing phenotype definitions and supports large-scale international collaborations. It also exemplifies the integration of PRSs in MASLD research, showing the diagnostic and management advantages that polygenic information offers for complex disease pathways and risk stratification, also given the strong evidence and the consistency of results obtained by GWASs on the principal trait and endophenotypes related to MASLD and liver damage in the Milan Biobank was offered. Moreover, by performing for the first time a GWAS on noninvasive scores of liver fibrosis and by analyzing the performance of validated PRSs with the newly developed ones for FIB-4 and FNI, the need for the development of specific pPRS in order to better discriminate etiological events related to an advanced phenotype in liver diseases was pinpointed and the clinical relevance of non-invasive marker used in clinical practice was demonstrated based on the consistency of the obtained results with the literature; this approach could pave the way to a new strategy can be helpful in the era of personalized risk scores and precision medicine.

Abbreviations

ABCG1 – ATP Binding Cassette Subfamily G Member 1

ACER2 – Alkaline Ceramidase 2

AK3 – Adenylate Kinase 3

ALDH1A2 – Aldehyde Dehydrogenase 1 Family Member A2

ALT – Alanine Aminotransferase

APOA1 – Apolipoprotein A1

APOA5 – Apolipoprotein A5

APOB / ApoB-100 – Apolipoprotein B-100

APOC1 – Apolipoprotein C1

APOC3 – Apolipoprotein C3

APOE – Apolipoprotein E

AQP8 – Aquaporin 8

AST – Aspartate Aminotransferase

AUROC – Area Under the Receiver Operating Characteristic curve

AUC – Area Under the Curve

BMI – Body Mass Index

CAP – Controlled Attenuation Parameter

CD206 (MRC1) – Macrophage Mannose Receptor C-type 1

CELSR2 – Cadherin EGF LAG Seven-Pass G-Type Receptor 2

CETP – Cholesteryl Ester Transfer Protein

CHO – Total Cholesterol

CPXM1 – Carboxypeptidase X, M14 Family Member 1

DAMPs – Damage-Associated Molecular Patterns

DIAPH2 – Diaphanous Related Formin 2

ECM – Extracellular Matrix

EMT – Epithelial-Mesenchymal Transition

EWAS – Exome-Wide Association Study

EXOC3L4 – Exocyst Complex Component 3 Like 4

FAST – FibroScan-AST score

FDR – False Discovery Rate

FIB-4 – Fibrosis-4 Index

FNI – Fibrotic NASH Index

GCKR – Glucokinase Regulator

GEO – Gene Expression Omnibus

GGT – Gamma-Glutamyl Transferase

GGT1 – Gamma-Glutamyltransferase 1

GSEA – Gene Set Enrichment Analysis

GTEx – Genotype-Tissue Expression project

GWAS – Genome-Wide Association Study

HbA1c – Glycated Hemoglobin A1c

HBS1L – HBS1 Like Translational GTPase

HCC – Hepatocellular Carcinoma

HCV – Hepatitis C Virus

HDL – High-Density Lipoprotein

IFI27 – Interferon Alpha Inducible Protein 27

HSI – Hepatic Steatosis Index

HIV – Human Immunodeficiency Virus

HMGCLL1 – 3-Hydroxy-3-Methylglutaryl-CoA Lyase Like 1

HNF1A – Hepatocyte Nuclear Factor 1-Alpha

HSCs – Hepatic Stellate Cells

HSD17B13 – Hydroxysteroid 17-Beta Dehydrogenase 13

IL-1β – Interleukin 1 beta

IL-18 – Interleukin 18

IL-6 – Interleukin 6

IL2RA – Interleukin 2 Receptor Subunit Alpha

IL1RAPL1 – Interleukin 1 Receptor Accessory Protein Like 1

IR – Insulin Resistance

JAK – Janus Kinase

KCs – Kupffer Cells

KRT5 – Keratin 5

LD – Linkage Disequilibrium

LDL – Low-Density Lipoprotein

LDLR – Low-Density Lipoprotein Receptor

LIPC – Hepatic Lipase C

LPL – Lipoprotein Lipase

LRRC34 – Leucine Rich Repeat Containing 34

LRRIQ3 – Leucine Rich Repeat And IQ Motif Containing 3

LPS – Lipopolysaccharide

LSM – Liver Stiffness Measurement

LSECs – Liver Sinusoidal Endothelial Cells

MAGEE2 – Melanoma Antigen Family E Member 2

MASLD – Metabolic dysfunction-Associated Steatotic Liver Disease

MASH – Metabolic dysfunction-Associated Steatohepatitis

MBOAT7 – Membrane Bound O-Acyltransferase Domain Containing 7

m-KCs – Monocyte-derived Kupffer Cells

MRC1 (CD206) – Macrophage Mannose Receptor C-type 1

MT1B – Metallothionein 1B

NAS – NAFLD Activity Score

NAFLD – Non-Alcoholic Fatty Liver Disease

NASH – Non-Alcoholic Steatohepatitis

NF-kβ / NFKB1 – Nuclear Factor kappa-light-chain-enhancer of activated B cells

NGS – Next-Generation Sequencing

NLRP3 – NLR Family Pyrin Domain Containing 3

NKX6-2 – NK6 Homeobox 2

ORA – Over Representation Analysis

OR – Odds Ratio

P2RY10 – Purinergic Receptor P2Y10

PAMPs – Pathogen-Associated Molecular Patterns

PCA – Principal Component Analysis

PCs – Principal Components

PCSK1N – Proprotein Convertase Subtilisin/Kexin Type 1 Inhibitor

PheWAS – Phenome-Wide Association Study

PLTs – Platelets

PNPLA3 – Patatin-like Phospholipase Domain Containing 3

PPARG – Peroxisome Proliferator-Activated Receptor Gamma

PRS – Polygenic Risk Score

pPRS – Partitioned Polygenic Risk Score

PSRC1 – Proline/Serine Rich Coiled-Coil 1

RNA-Seq – RNA Sequencing

RORA – RAR Related Orphan Receptor A

scRNA-Seq – Single-Cell RNA Sequencing

SLC12A1 – Solute Carrier Family 12 Member 1

SLD – Steatotic Liver Disease

SMARCA4 – SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin, Subfamily A, Member 4

snRNA-Seq – Single-Nucleus RNA Sequencing

SNPs – Single-Nucleotide Polymorphisms

SPINK1 – Serine Peptidase Inhibitor Kazal Type 1

STAT3 – Signal Transducer and Activator of Transcription 3

T2D – Type 2 Diabetes

TBL2 – Transducin Beta Like 2

TGF-β – Transforming Growth Factor Beta

TGs – Triglycerides

TM4SF18 – Transmembrane 4 L Six Family Member 18

TM6SF1 – Transmembrane 6 Superfamily Member 1

TM6SF2 – Transmembrane 6 Superfamily Member 2

TMEM92 – Transmembrane Protein 92

TNF-α – Tumor Necrosis Factor Alpha

TOPMed – Trans-Omics for Precision Medicine program

TRIB1 / TRIB1AL – Tribbles Pseudokinase 1 / TRIB1 Adjacent Locus

UGT1A8 – UDP Glucuronosyltransferase Family 1 Member A8

UKBB – UK Biobank

US – United States

VLDL – Very Low-Density Lipoprotein

VMA21 – Vacuolar ATPase Assembly Factor VMA21

VWA7 – Von Willebrand Factor A Domain Containing 7

WDR17 – WD Repeat Domain 17

ZPR1 – ZPR1 Zinc Finger Protein

Bibliography

1. Rinella, M. E. *et al.* A multisociety Delphi consensus statement on new fatty liver disease nomenclature. *J. Hepatol.* **79**, 1542–1556 (2023).
2. Solomon, A. *et al.* Interactions between Metabolic Syndrome, MASLD, and Arterial Stiffening: A Single-Center Cross-Sectional Study. *Healthcare* **11**, 2696 (2023).
3. Rinella, M. E. *et al.* AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. *Hepatol. Baltim. Md* **77**, 1797–1835 (2023).
4. Younossi, Z. M. *et al.* The economic and clinical burden of nonalcoholic fatty liver disease in the United States and Europe. *Hepatol. Baltim. Md* **64**, 1577–1586 (2016).
5. Ahmad, F. B. Mortality in the United States — Provisional Data, 2023. *MMWR Morb. Mortal. Wkly. Rep.* **73**, (2024).
6. Cohen, J. C., Horton, J. D. & Hobbs, H. H. Human Fatty Liver Disease: Old Questions and New Insights. *Science* **332**, 1519–1523 (2011).
7. Llovet, J. M. *et al.* Nonalcoholic steatohepatitis-related hepatocellular carcinoma: pathogenesis and treatment. *Nat. Rev. Gastroenterol. Hepatol.* **20**, 487–503 (2023).
8. Anstee, Q. M., Reeves, H. L., Kotsiliti, E., Govaere, O. & Heikenwalder, M. From NASH to HCC: current concepts and future challenges. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 411–428 (2019).
9. Starley, B. Q., Calcagno, C. J. & Harrison, S. A. Nonalcoholic fatty liver disease and hepatocellular carcinoma: a weighty connection. *Hepatol. Baltim. Md* **51**, 1820–1832 (2010).
10. Younossi, Z. M. *et al.* The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): a systematic review. *Hepatol. Baltim. Md* **77**, 1335–1347 (2023).
11. Wong, V. W.-S., Ekstedt, M., Wong, G. L.-H. & Hagström, H. Changing epidemiology, global trends and implications for outcomes of NAFLD. *J. Hepatol.* **79**, 842–852 (2023).
12. Singal, A. G., Lampertico, P. & Nahon, P. Epidemiology and surveillance for hepatocellular carcinoma: New trends. *J. Hepatol.* **72**, 250–261 (2020).
13. Seki, E. & Schwabe, R. F. Hepatic inflammation and fibrosis: functional links and key

pathways. *Hepatol. Baltim. Md* **61**, 1066–1079 (2015).

14. West, A. P. & Shadel, G. S. Mitochondrial DNA in innate immune responses and inflammatory pathology. *Nat. Rev. Immunol.* **17**, 363–375 (2017).
15. Watanabe, A. *et al.* Apoptotic hepatocyte DNA inhibits hepatic stellate cell chemotaxis via toll-like receptor 9. *Hepatol. Baltim. Md* **46**, 1509–1518 (2007).
16. Koyama, Y. & Brenner, D. A. Liver inflammation and fibrosis. *J. Clin. Invest.* **127**, 55–64 (2017).
17. Wree, A. *et al.* NLRP3 inflammasome activation results in hepatocyte pyroptosis, liver inflammation, and fibrosis in mice. *Hepatol. Baltim. Md* **59**, 898–910 (2014).
18. Wang, Q. *et al.* Hyperglycemia aggravates acute liver injury by promoting liver-resident macrophage NLRP3 inflammasome activation via the inhibition of AMPK/mTOR-mediated autophagy induction. *Immunol. Cell Biol.* **98**, 54–66 (2020).
19. Pierantonelli, I. *et al.* Lack of NLRP3-inflammasome leads to gut-liver axis derangement, gut dysbiosis and a worsened phenotype in a mouse model of NAFLD. *Sci. Rep.* **7**, 12200 (2017).
20. Van Herck, M. A. *et al.* The Differential Roles of T Cells in Non-alcoholic Fatty Liver Disease and Obesity. *Front. Immunol.* **10**, 82 (2019).
21. Tomita, K. *et al.* Tumour necrosis factor alpha signalling through activation of Kupffer cells plays an essential role in liver fibrosis of non-alcoholic steatohepatitis in mice. *Gut* **55**, 415–424 (2006).
22. Wang, H., Mehal, W., Nagy, L. E. & Rotman, Y. Immunological mechanisms and therapeutic targets of fatty liver diseases. *Cell. Mol. Immunol.* **18**, 73–91 (2021).
23. Luci, C., Bourinet, M., Leclère, P. S., Anty, R. & Gual, P. Chronic Inflammation in Non-Alcoholic Steatohepatitis: Molecular Mechanisms and Therapeutic Strategies. *Front. Endocrinol.* **11**, 597648 (2020).
24. Tran, S. *et al.* Impaired Kupffer Cell Self-Renewal Alters the Liver Response to Lipid Overload during Non-alcoholic Steatohepatitis. *Immunity* **53**, 627-640.e5 (2020).
25. Heyens, L. J. M., Busschots, D., Koek, G. H., Robaeys, G. & Francque, S. Liver Fibrosis in Non-alcoholic Fatty Liver Disease: From Liver Biopsy to Non-invasive Biomarkers in Diagnosis

and Treatment. *Front. Med.* **8**, 615978 (2021).

26. Elpek, G. Ö. Cellular and molecular mechanisms in the pathogenesis of liver fibrosis: An update. *World J. Gastroenterol. WJG* **20**, 7260–7276 (2014).
27. Zhou, W.-C., Zhang, Q.-B. & Qiao, L. Pathogenesis of liver cirrhosis. *World J. Gastroenterol. WJG* **20**, 7312–7324 (2014).
28. Speliotes, E. K. *et al.* Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet.* **7**, e1001324 (2011).
29. Sookoian, S., Pirola, C. J., Valenti, L. & Davidson, N. O. Genetic Pathways in Nonalcoholic Fatty Liver Disease: Insights From Systems Biology. *Hepatol. Baltim. Md* **72**, 330–346 (2020).
30. Chen, Y. *et al.* Genome-wide association meta-analysis identifies 17 loci associated with nonalcoholic fatty liver disease. *Nat. Genet.* **55**, 1640–1650 (2023).
31. Romeo, S. *et al.* Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat. Genet.* **40**, 1461–1465 (2008).
32. Grimaudo, S. *et al.* Association Between PNPLA3 rs738409 C>G Variant and Liver-Related Outcomes in Patients With Nonalcoholic Fatty Liver Disease. *Clin. Gastroenterol. Hepatol.* **18**, 935-944.e3 (2020).
33. Liu, Y.-L. *et al.* TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. *Nat. Commun.* **5**, 4309 (2014).
34. Li, B.-T. *et al.* Disruption of the ERLIN-TM6SF2-APOB complex destabilizes APOB and contributes to non-alcoholic fatty liver disease. *PLoS Genet.* **16**, e1008955 (2020).
35. Luo, F., Oldoni, F. & Das, A. TM6SF2: A Novel Genetic Player in Nonalcoholic Fatty Liver and Cardiovascular Disease. *Hepatol. Commun.* **6**, 448–460 (2021).
36. Alharthi, J. *et al.* A metabolic associated fatty liver disease risk variant in MBOAT7 regulates toll like receptor induced outcomes. *Nat. Commun.* **13**, 7430 (2022).
37. Wu, N. *et al.* Waist circumference mediates the association between rs1260326 in GCKR gene and the odds of lean NAFLD. *Sci. Rep.* **13**, 6488 (2023).
38. Su, W. *et al.* Role of HSD17B13 in the liver physiology and pathophysiology. *Mol. Cell.*

Endocrinol. **489**, 119–125 (2019).

39. Pirola, C. J. *et al.* Splice variant rs72613567 prevents worst histologic outcomes in patients with nonalcoholic fatty liver disease. *J. Lipid Res.* **60**, 176–185 (2019).
40. Valenti, L. & Casirati, E. Editorial: Unveiling new horizons for liver steatosis genetic variants beyond hepatocellular carcinoma diagnosis - exploring the potential of HSD17B13 inhibition. *Aliment. Pharmacol. Ther.* **58**, 727–728 (2023).
41. Ozturk, A., Olson, M. C., Samir, A. E. & Venkatesh, S. K. Liver fibrosis assessment: MR and US elastography. *Abdom. Radiol. N. Y.* **47**, 3037–3050 (2022).
42. Wang, J.-L. *et al.* Non-invasive diagnosis of non-alcoholic fatty liver disease: Current status and future perspective. *Heliyon* **10**, e27325 (2024).
43. Abarca Ruiz, J., Villagomez Estrada, M., Calderón, W. & Carrera, E. P-122 LIVER TRANSIENT ELASTOGRAPHY (FIBROSCAN). FIRST REPORT OF EXPERIENCE IN ECUADOR. *Ann. Hepatol.* **24**, 100482 (2021).
44. Chen, X. *et al.* Fibrosis-4 index predicts mortality in HIV/HCV co-infected patients receiving combination antiretroviral therapy in rural China. *Biosci. Trends* **13**, 32–39 (2019).
45. Tavaglione, F. *et al.* Development and Validation of a Score for Fibrotic Nonalcoholic Steatohepatitis. *Clin. Gastroenterol. Hepatol. Off. Clin. Pract. J. Am. Gastroenterol. Assoc.* **21**, 1523-1532.e1 (2023).
46. Newsome, P. N. *et al.* FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: a prospective derivation and global validation study. *Lancet Gastroenterol. Hepatol.* **5**, 362–373 (2020).
47. Priego-Parra, B. A. *et al.* Hepatic Steatosis Index (HSI): A Valuable Biomarker in Subjects with Metabolic Dysfunction-associated Fatty Liver Disease (MAFLD). *Ann. Hepatol.* **29**, (2024).
48. Jamialahmadi, O. *et al.* Partitioned polygenic risk scores identify distinct types of metabolic dysfunction-associated steatotic liver disease. *Nat. Med.* **30**, 3614–3623 (2024).
49. Dongiovanni, P. *et al.* Causal relationship of hepatic fat with liver damage and insulin resistance in nonalcoholic fatty liver. *J. Intern. Med.* **283**, 356–370 (2018).
50. Bianco, C. *et al.* Non-invasive stratification of hepatocellular carcinoma risk in non-alcoholic

fatty liver using polygenic risk scores. *J. Hepatol.* **74**, 775–782 (2021).

51. Pirazzi, C. *et al.* Patatin-like phospholipase domain-containing 3 (PNPLA3) I148M (rs738409) affects hepatic VLDL secretion in humans and in vitro. *J. Hepatol.* **57**, 1276–1282 (2012).
52. Prill, S. *et al.* The TM6SF2 E167K genetic variant induces lipid biosynthesis and reduces apolipoprotein B secretion in human hepatic 3D spheroids. *Sci. Rep.* **9**, 11585 (2019).
53. Hoffmann, T. J. *et al.* A large electronic-health-record-based genome-wide study of serum lipids. *Nat. Genet.* **50**, 401–413 (2018).
54. Chambers, J. C. *et al.* Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat. Genet.* **43**, 1131–1138 (2011).
55. Vujkovic, M. *et al.* A multiancestry genome-wide association study of unexplained chronic ALT elevation as a proxy for nonalcoholic fatty liver disease with histological and radiological validation. *Nat. Genet.* **54**, 761–771 (2022).
56. Ghouse, J. *et al.* Integrative common and rare variant analyses provide insights into the genetic architecture of liver cirrhosis. *Nat. Genet.* **56**, 827–837 (2024).
57. Kanai, M. *et al.* A second update on mapping the human genetic architecture of COVID-19. *Nature* **621**, E7–E26 (2023).
58. Bianco, C. *et al.* Predictors of controlled attenuation parameter in metabolic dysfunction. *United Eur. Gastroenterol. J.* **12**, 364–373 (2023).
59. Baselli, G. A. *et al.* Rare ATG7 genetic variants predispose patients to severe fatty liver disease. *J. Hepatol.* **77**, 596–606 (2022).
60. Sberna, A. L. *et al.* European Association for the Study of the Liver (EASL), European Association for the Study of Diabetes (EASD) and European Association for the Study of Obesity (EASO) clinical practice recommendations for the management of non-alcoholic fatty liver disease: evaluation of their application in people with Type 2 diabetes. *Diabet. Med. J. Br. Diabet. Assoc.* **35**, 368–375 (2018).
61. COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur. J. Hum. Genet. EJHG* **28**, 715–718 (2020).

62. Degenhardt, F. *et al.* Detailed stratified GWAS analysis for severe COVID-19 in four European populations. *Hum. Mol. Genet.* **31**, 3945–3966 (2022).
63. Severe Covid-19 GWAS Group *et al.* Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).
64. Price, A. L. *et al.* Long-Range LD Can Confound Genome Scans in Admixed Populations. *Am. J. Hum. Genet.* **83**, 132–135 (2008).
65. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
66. De Vincentis, A. *et al.* A Polygenic Risk Score to Refine Risk Stratification and Prediction for Severe Liver Disease by Clinical Fibrosis Scores. *Clin. Gastroenterol. Hepatol. Off. Clin. Pract. J. Am. Gastroenterol. Assoc.* **20**, 658–673 (2022).
67. Govaere, O. *et al.* Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. *Sci. Transl. Med.* **12**, eaba4448 (2020).
68. Baselli, G. A. *et al.* Liver transcriptomics highlights interleukin-32 as novel NAFLD-related cytokine and candidate biomarker. *Gut* **69**, 1855–1866 (2020).
69. Xiao, Y. *et al.* Hepatocytes demarcated by EphB2 contribute to the progression of nonalcoholic steatohepatitis. *Sci. Transl. Med.* **15**, eadc9653 (2023).
70. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
71. Linehan, S. A. *et al.* Enhanced expression of the mannose receptor by endothelial cells of the liver and spleen microvascular beds in the macrophage-deficient PU.1 null mouse. *Histochem. Cell Biol.* **123**, 365–376 (2005).
72. McGreal, E. P., Martinez-Pomares, L. & Gordon, S. Divergent roles for C-type lectins expressed by cells of the innate immune system. *Mol. Immunol.* **41**, 1109–1121 (2004).
73. Xiong, Z. *et al.* Integrated Bioinformatics and Validation Reveal IFI27 and Its Related Molecules as Potential Identifying Genes in Liver Cirrhosis. *Biomolecules* **14**, 13 (2023).
74. Kostrzewska-Poczekaj, M. *et al.* DIAPH2 alterations increase cellular motility and may

contribute to the metastatic potential of laryngeal squamous cell carcinoma. *Carcinogenesis* **40**, 1251–1259 (2019).

75. Speliotes, E. K. *et al.* PNPLA3 variants specifically confer increased risk for histologic nonalcoholic fatty liver disease but not metabolic disease. *Hepatol. Baltim. Md* **52**, 904–912 (2010).
76. Cannata Serio, M. *et al.* Mutations in the V-ATPase Assembly Factor VMA21 Cause a Congenital Disorder of Glycosylation With Autophagic Liver Disease. *Hepatol. Baltim. Md* **72**, 1968–1986 (2020).
77. Hayden, M. S. & Ghosh, S. Shared Principles in NF-κB Signaling. *Cell* **132**, 344–362 (2008).
78. Pikarsky, E. *et al.* NF-κB functions as a tumour promoter in inflammation-associated cancer. *Nature* **431**, 461–466 (2004).
79. Vu-Dac, N. *et al.* Transcriptional regulation of apolipoprotein A-I gene expression by the nuclear receptor RORalpha. *J. Biol. Chem.* **272**, 22401–22404 (1997).
80. Wang, S. *et al.* Identification and Characterization of Two Novel Compounds: Heterozygous Variants of Lipoprotein Lipase in Two Pedigrees With Type I Hyperlipoproteinemia. *Front. Endocrinol.* **13**, (2022).
81. Harshfield, E. L. *et al.* Genome-wide analysis of blood lipid metabolites in over 5000 South Asians reveals biological insights at cardiometabolic disease loci. *BMC Med.* **19**, 232 (2021).
82. Gou, Y. *et al.* PNPLA3-I148M Variant Promotes the Progression of Liver Fibrosis by Inducing Mitochondrial Dysfunction. *Int. J. Mol. Sci.* **24**, 9681 (2023).