

# Genetics of Nonalcoholic Fatty Liver Disease: From Pathogenesis to Therapeutics

Silvia Sookoian, MD, PhD, FAASLD<sup>1,2</sup> Carlos J. Pirola, PhD, FAHA<sup>1,3</sup>

<sup>1</sup>Institute of Medical Research A Lanari, School of Medicine, University of Buenos Aires, Ciudad Autonoma de Buenos Aires (C1427ARN), Argentina

<sup>2</sup>Department of Clinical and Molecular Hepatology, National Scientific and Technical Research Council (CONICET), University of Buenos Aires, Institute of Medical Research (IDIM), Ciudad Autónoma de Buenos Aires (C1427ARN), Argentina

<sup>3</sup>Department of Molecular Genetics and Biology of Complex Diseases, National Scientific and Technical Research Council (CONICET), University of Buenos Aires, Institute of Medical Research (IDIM), Ciudad Autonoma de Buenos Aires (C1427ARN), Argentina

Address for correspondence Silvia Sookoian, MD, PhD, Instituto de Investigaciones Médicas (UBA-CONICET), Combatientes de Malvinas 3150, CABA- C1427ARN, Argentina (e-mail: ssookoian@intramed.net). Carlos J. Pirola, PhD, Instituto de Investigaciones Médicas (UBA-CONICET), Combatientes de Malvinas 3150, CABA- C1427ARN, Argentina (e-mail: pirola.carlos@conicet.gov.ar).

Semin Liver Dis 2019;39:124–140.

## Abstract

### Keywords

- ▶ genetics
- ▶ PNPLA3
- ▶ TM6SF2
- ▶ STAT3
- ▶ GCKR
- ▶ HSD17B13
- ▶ GWAS
- ▶ NASH
- ▶ systems biology
- ▶ druggability

Here, the authors review the remarkable genetic discoveries that have illuminated the biology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH). The authors integrate genes associated with NAFLD and NASH into regulatory pathways to elucidate the disease pathogenesis. They review the evidence for molecular mediators of chronic liver damage, which suggests that convergent pathophenotypes, including inflammation and fibrosis, share common genetic modifiers. They further demonstrate that genes involved in the genetic susceptibility of NAFLD and NASH participate in cross-phenotype associations with diseases of the metabolic syndrome, including type 2 diabetes, obesity, and cardiovascular disease. However, immune-related loci associated with NAFLD and NASH exhibit some level of pleiotropy influencing disparate phenotypes, such as premature birth or sepsis. They finally focus on the translation of current genetic knowledge of NAFLD and NASH toward precision medicine. They provide evidence of genetic findings that can be leveraged to identify therapeutic targets.

Nonalcoholic fatty liver disease (NAFLD) is a condition manifested by an abnormal accumulation of fat in the liver, which can present signs of hepatocyte injury and chronic damage, such as those that characterize nonalcoholic steatohepatitis (NASH).<sup>1</sup> The disease can progress into severe clinical forms, including NASH-fibrosis, cirrhosis, and even hepatocellular carcinoma (HCC).<sup>1</sup>

Epidemiological observations derived from population-based studies,<sup>2,3</sup> familial aggregation studies,<sup>4,5</sup> and twin studies<sup>6,7</sup> have long provided evidence that NAFLD is at some level a heritable trait. NAFLD has been observed in a cluster of families, whereby Schwimmer et al found that fatty liver is

significantly more common in siblings (59%) and parents (78%) of children with NAFLD.<sup>4</sup> The heritability estimates of NAFLD range from 20 to 70%, depending on the study design and diagnostic approaches used in determining the liver phenotype.<sup>2–6</sup>

Variants of over 100 loci have been explored in candidate-gene association studies (see ▶ **Table 1**). Findings yielded by these studies have generated plausible evidence indicating that several loci are involved in the genetic susceptibility of NAFLD, including nuclear receptors, transcription factors that regulate lipid- and carbohydrate-related biosynthetic processes, inflammatory response, and fibrogenesis.<sup>8–10</sup>

published online  
March 25, 2019

Copyright © 2019 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA.  
Tel: +1(212) 584-4662.

DOI <https://doi.org/10.1055/s-0039-1679920>.  
ISSN 0272-8087.

**Table 1** Training gene list based on published evidence of the genetic component of NAFLD and NASH

Gene symbol (gene description)
RPL13AP7 (ribosomal protein L13a pseudogene 7)
ABCB11 (ATP binding cassette subfamily B member 11)
ACSL4 (acyl-CoA synthetase long chain family member 4)
ACTR5 (ARP5 actin related protein 5 homolog)
ADIPOQ (adiponectin, C1Q and collagen domain containing)
ADIPOR1 (adiponectin receptor 1)
ADIPOR2 (adiponectin receptor 2)
ADRB2 (adrenoceptor $\beta$ 2)
ADRB3 (adrenoceptor $\beta$ 3)
AGTR1 (angiotensin II receptor type 1)
APOC3 (apolipoprotein C3)
APOE (apolipoprotein E)
ARHGEF40 (Rho guanine nucleotide exchange factor 40)
C1orf94 (chromosome 1 open reading frame 94)
CACNA2D1 (calcium voltage-gated channel auxiliary subunit $\alpha$ 2delta 1)
CD14 (CD14 molecule)
CDH2 (cadherin 2)
CFTR (cystic fibrosis transmembrane conductance regulator)
CLOCK (clock circadian regulator)
CNTN5 (contactin 5)
COL13A1 (collagen type XIII $\alpha$ 1 chain)
CRACR2A (calcium release activated channel regulator 2A)
CYP2E1 (cytochrome P450 family 2 subfamily E member 1)
DCLK1 (doublecortin like kinase 1)
DGAT1 (diacylglycerol O-acyltransferase 1)
DGAT2 (diacylglycerol O-acyltransferase 2)
DYSF (dysferlin)
EHBP1L1 (EH domain binding protein 1 like 1)
ENPP1 (ectonucleotide pyrophosphatase/phosphodiesterase 1)
ETS1 (ETS proto-oncogene 1, transcription factor)
FABP2 (fatty acid binding protein 2)
FARP1 (FERM, ARH/RhoGEF and pleckstrin domain protein 1)
FDFT1 (farnesyl-diphosphate farnesyltransferase 1)
GATAD2A (GATA zinc finger domain containing 2A)
GC (GC, vitamin D binding protein)
GCKR (glucokinase regulator)
GCLC (glutamate-cysteine ligase catalytic subunit)
HFE (homeostatic iron regulator)
HS3ST1 (heparan sulfate-glucosamine 3-sulfotransferase 1)
HSD17B13 (hydroxysteroid 17- $\beta$ dehydrogenase 13)

(Continued)

**Table 1** (Continued)

Gene symbol (gene description)
IL18RAP (interleukin 18 receptor accessory protein)
IL1B (interleukin 1 $\beta$ )
IL6 (interleukin 6)
IRS1 (insulin receptor substrate 1)
KHDRBS3 (KH RNA binding domain containing, signal transduction associated 3)
KLF6 (Kruppel-like factor 6)
LCP1 (lymphocyte cytosolic protein 1)
LEPR (leptin receptor)
LINC00322 (long intergenic nonprotein coding RNA 322)
LIPC (lipase C, hepatic type)
PRG1 (p53-responsive gene 1)
LTBP3 (latent transforming growth factor $\beta$ binding protein 3)
LYPLAL1 (lysophospholipase like 1)
MACROD2 (MACRO domain containing 2)
MBOAT7 (membrane bound O-acyltransferase domain containing 7)
MC4R (melanocortin 4 receptor)
MIF (macrophage migration inhibitory factor)
MTCYBP22 (mitochondrially encoded cytochrome b pseudogene 22)
MTHFR (methylenetetrahydrofolate reductase)
MTTP (microsomal triglyceride transfer protein)
MUM1 (melanoma associated antigen (mutated) 1)
NCAN (neurocan)
NFIC (nuclear factor I C)
NGF (nerve growth factor)
NR1I2 (nuclear receptor subfamily 1 group I member 2)
OTX2P1 (orthodenticle homeobox 2 pseudogene 1)
PALLD (palladin, cytoskeletal associated protein)
PARVB (parvin $\beta$ )
PBX2P1 (PBX homeobox 2 pseudogene 1)
PDGFA (platelet derived growth factor subunit A)
PEMT (phosphatidylethanolamine N-methyltransferase)
PNPLA3 (patatin like phospholipase domain containing 3)
PPARA (peroxisome proliferator-activated receptor $\alpha$ )
PPARG (peroxisome proliferator-activated receptor gamma)
PPARGC1A (PPARG coactivator 1 $\alpha$ , PGC-1a)
PPP1R3B (protein phosphatase 1 regulatory subunit 3B)
PTGS2 (prostaglandin-endoperoxide synthase 2)
PTPRU (protein tyrosine phosphatase, receptor type U)
PZP (PZP, $\alpha$ -2-macroglobulin like)
RAB37 (RAB37, member RAS oncogene family)

(Continued)

**Table 1** (Continued)

Gene symbol (gene description)
SAMM50 (SAMM50 sorting and assembly machinery component)
SDK1 (sidekick cell adhesion molecule 1)
SEL1L3 (SEL1L family member 3)
SERPINA1 (serpin family A member 1)
SLC38A8 (solute carrier family 38 member 8)
SLC46A3 (solute carrier family 46 member 3)
SLC9A9 (solute carrier family 9 member A9)
SOD2 (superoxide dismutase 2)
SPINK1 (serine peptidase inhibitor, Kazal type 1)
ST8SIA1 (ST8 $\alpha$ -N-acetyl-neuraminide $\alpha$ -2,8-sialyltransferase 1)
STAT3 (signal transducer and activator of transcription 3)
TCF7L2 (transcription factor 7 like 2)
TEX36 (testis expressed 36)
TLR4 (toll like receptor 4)
TM6SF2 (transmembrane 6 superfamily member 2)
TMEM56 (transmembrane protein 56)
TNF (tumor necrosis factor)
TNFSF10 (TNF superfamily member 10)
TRAPPC9 (trafficking protein particle complex 9)
UCP1 (uncoupling protein 1)
UGT1A1 (UDP glucuronosyltransferase family 1 member A1)
YIPF1 (Yip1 domain family member 1)
ZNF512 (zinc finger protein 512)
ZP4 (zona pellucida glycoprotein 4)

Abbreviations: NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis.

Nevertheless, authors of a large majority of candidate-gene studies on NAFLD and NASH have failed to convincingly demonstrate a robust causal relationship between the associated variant and the disease. This could be explained by the limited number of functional mechanistic studies designed to test the hypotheses driving the investigations, or simply by the lack of statistical power.

Conversely, discoveries of variants of three genes (*PNPLA3*-rs738409, *TM6SF2*-rs58542916, and glucokinase regulator gene [*GCKR*]-rs780094 or *GCKR*-rs1260326) that regulate metabolic traits have been driven by genome-wide approaches,<sup>2,3,11,12</sup> including genome-wide association (GWAS) and exome-wide association (EWAS) studies. The association between these gene variants with the risk of NAFLD in cohorts of diverse ethnical backgrounds around the world was demonstrated by several authors.<sup>13-15</sup> In addition, variants in these genes have been associated with the risk of NASH and histological features of the disease severity, including liver fibrosis.<sup>13,16-18</sup>

Variants in additional loci, including a missense (p.Gly17-Glu, rs641738 C/T) variant in exon 1 of transmembrane channel-like 4 (*TMC4*)/intergenic-downstream of membrane-bound O-acyltransferase domain-containing 7 (*MBOAT7*), have been associated with a modest risk of NAFLD and NASH in Italian population.<sup>19</sup> However, this association could not be replicated in populations of other ethnicities.<sup>20-22</sup>

More recently, a study that involved the analysis of exome-sequence data coupled to electronic health records of 46,455 patients taking part in a large collaborative study revealed a loss-of-function variation in hydroxysteroid 17- $\beta$  dehydrogenase 13 (*HSD17B13*) gene that confers protection against chronic liver injury and mitigates progressive NASH among European Americans.<sup>23</sup>

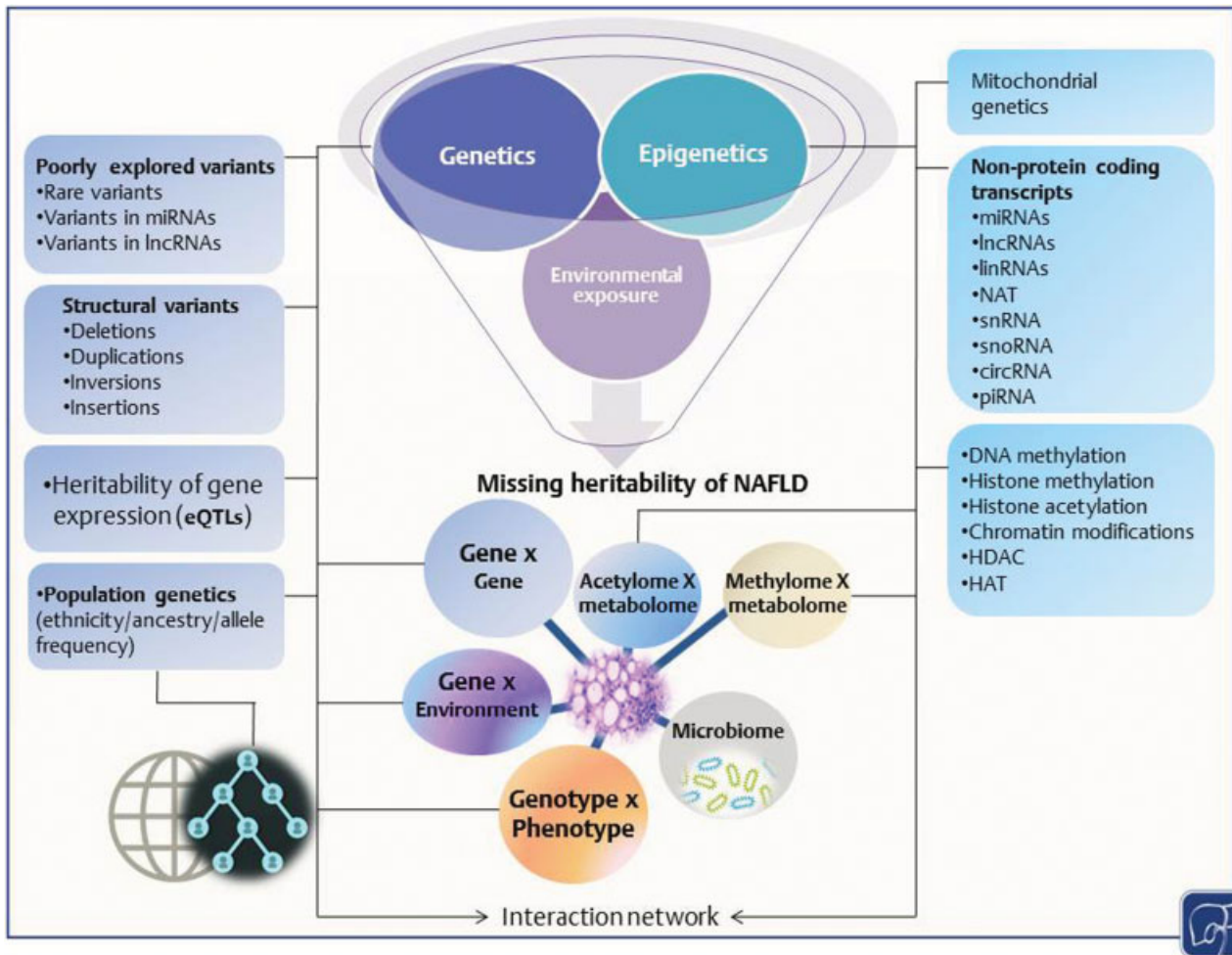
From the histopathologic point of view, NAFLD refers to potentially progressive lesions ranging from isolated steatosis (NAFL) to NASH<sup>1</sup> as explained above. It is then reasonable to hypothesize that NAFL, NASH, and NASH-fibrosis share genetic modifiers. In fact, variants in locus influencing the risk of NAFLD, including *PNPLA3*, *TM6SF2*, and *GCKR*, contribute to the risk of NASH as well. For example, the rs738409 presents a significant effect not only on liver fat accumulation (GG homozygous carriers show 73% higher lipid fat content when compared with CC homozygous) but also on the susceptibility of a more aggressive disease (GG homozygous carriers have 3.24-fold greater risk of higher inflammatory scores and 3.2-fold greater risk of developing fibrosis when compared with CC homozygous).<sup>13</sup>

While single nucleotide polymorphisms (SNPs) currently known as involved in the genetic risk of NAFLD cannot distinguish between isolated steatosis and NASH, some of these variants can inform the chances of presenting a more advanced disease. For example, NASH is 3.5-fold more frequently observed in GG homozygous than in CC homozygous carriers.<sup>13</sup>

The questions arise as to why much of the genetic variants do not allow us to differentiate NAFL from progressive NASH. Many explanations for this question can be suggested, which vary from the assumption that there is no NASH without NAFL, and consequently variants influencing the risk of NAFL directly or indirectly affect the predisposition to NASH to a more pragmatic point of view that questions the designs of genetic studies, including imprecise phenotyping and the use of controls of uncertain comparability. Certainly, the evidence suggests that some factors either genetic or environmental should affect the progression of NAFL to NASH if they are stages of a single disease. These factors are likely related with the inflammatory response and fibrogenesis process.

## The Missing Heritability of NAFLD and NASH

According to the available evidence, the effect of variants uncovered from GWAS or EWAS<sup>2,3,11,12,23</sup> explains a small portion of the disease variance. In fact, variants in the loci mentioned above explain up to approximately 10% of the variance NAFLD-related phenotypes. Hence, it is clear that knowledge of the phenotypic variance of NAFLD and NASH,



**Fig. 1** Missing heritability of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH). There is considerable disparity in the magnitude of heritability estimates of NAFLD and NASH and the proportion of variance explained by single nucleotide polymorphisms (SNPs) uncovered from genome-wide association study (GWAS), exome-wide association study (EWAS), and candidate-gene association studies. A significant proportion of the disease burden could be explained by the missing heritability, which cover not only genetic and epigenetic modifiers but also the interaction with environmental exposure as well as with a highly interconnected and dynamic network of factors, including the microbiome.<sup>115,116</sup> The genetic component of NAFLD and NASH may be potentially explained by undiscovered rare variants, structural variation, including copy number variation, variants in micro-ribonucleic acids (miRNAs) and long noncoding RNAs (lncRNAs), and expression of quantitative trait loci (eQTLs). The allelic architecture of the human genome that substantially varies according to the different ethnic groups plays an important role. Variation across populations might explain differences in the prevalence and severity of the disease across different ethnic groups. Epigenetic factors include not only deoxyribonucleic acid (DNA) and histone methylation but also chromatin remodeling and nonprotein coding RNAs. Epigenetic inheritance also involves modifications of the histone code, including histone acetyltransferases (*HAT*) and deacetylases (*HDAC*). Abbreviations: circRNA, circular RNA; linRNAs, long intergenic RNA; NAT, natural antisense transcript; piRNA, PIWI-interacting RNA; snRNA, small nuclear RNA; snoRNA, small nucleolar RNA.

which stems from the interaction between the genetic component and environmental sources, is still lacking, resulting in what is known as missing heritability.

The missing heritability of NAFLD and NASH, like many other common diseases, includes a complex spectrum of factors that remain poorly explored; some of them are illustrated in ►Fig. 1. Mapping the genetic component of NAFLD and NASH should include not only the search for rare variants, which probably would have substantial effect/s on the phenotype but the exploration of structural variation, for example, copy number variants. Given the role of mitochondria on the physiology of the disease, it is also worthwhile to characterize the genetic diversity of the mitochondrial deoxyribonucleic acid (mtDNA).

Novel evidence on the variability of the mtDNA genome suggests a significant role of mitochondrial genetics in the pathogenesis of NAFLD and the natural history of the disease.<sup>24</sup> A comprehensive exploration of the complete liver mtDNA-mutation spectrum in patients with NAFLD and in different stages of the disease showed that NAFLD is associated with increased liver mtDNA mutational burden, including point mutations in genes of the oxidative phosphorylation.<sup>24</sup> In addition, patients with advanced fibrosis had an overall enrichment of 1.4-fold mutation rate compared with those in whom fibrosis was mild or absent.<sup>24</sup> The accumulation of liver mtDNA-polymorphic sites in subunits of the OXPHOS was paralleled by the emergence of an OXPHOS-deficient phenotype.<sup>24</sup> Specifically, profiling of liver OXPHOS-gene and protein expression

provided evidence of liver mtDNA mutational rate impacts on mitochondrial function.<sup>24</sup> These observations are supported by several studies that highlighted the importance of mitochondrial homeostasis in the pathogenesis of NAFLD and the disease progression into NASH and NASH-fibrosis.<sup>25–27</sup>

It is expected that variants in micro-ribonucleic acids (miRNAs), which may affect miRNA function, would account for a sizeable proportion of the disease risk and/or the association of NAFLD with comorbidities. For instance, rs41318021, a miR-122-related sequence genetic variation in the 3' untranslated region of the L-arginine transporter gene (*SLC7A1*) was associated with arterial hypertension in patients with NAFLD.<sup>28</sup> Across the disease spectrum, variants in long noncoding RNAs (lncRNAs) have been shown to cover a portion of genetic component as well.<sup>29</sup> A survey of genetic variation associated with lncRNA-genomic regions uncovered the rs2829145 A/G located in a lncRNA (*lnc-JAM2-6*), which was associated with NAFLD and the disease severity. Moreover, prediction of regulatory elements in *lnc-JAM2-6* indicated potential sequence-specific binding motifs of oncogenes MAF bZIP transcription factor K (*MAFK*) and JunD proto-oncogene, AP-1 transcription factor subunit (*JUND*), as well as transcription factors involved in inflammatory response.<sup>29</sup> Results from a pilot GWAS on NAFLD showed that intergenic or intron variants with predicted functionality in lncRNAs might be associated with steatosis, lobular inflammation, and liver fibrosis.<sup>30</sup> SNPs in *LYPLAL1* (rs12137855), *PPP1R3B* (rs4240624), and *TRIB1* (rs2954021), of which the predicted functionality in the corresponding loci is a long lincRNA, were associated with liver fat content.<sup>3</sup>

The concept of heritability is derived from a mathematical calculation involving three phenotype variance sources: genetics (G), environment (E), and individual (noise). Hence, a single number (G) represents the fraction of variation between individuals in a population that is due to their genetic background. Unfortunately, knowledge of the G × E interaction in the biology of NAFLD and NASH remains largely unexplored (► Fig. 1). There is, however, remarkable evidence showing an interaction effect of variants in genes predisposing an individual to NAFLD, specifically variants in *PNPLA3*, *TM6SF2*, and *GCKR* and adiposity.<sup>31</sup> The greatest effects were observed for the interaction between rs738409-G risk allele and obesity, which was found to affect the entire spectrum of NAFLD, from steatosis, to steatohepatitis, to end-stage liver disease, as well as liver enzyme levels.<sup>31</sup> For example, in persons homozygous for the G-allele compared with homozygous CC, the risk of progression to cirrhosis varies from 2.4-fold among lean (body mass index [BMI] < 25 kg/m<sup>2</sup>) to 5.8-fold among obese (BMI > 35 kg/m<sup>2</sup>) subjects.<sup>31</sup> While interaction effects between obesity and variants of *TM6SF2* (rs58542926) and *GCKR* (rs1260326) were also observed, the magnitude of the effect and the impact on the disease spectrum are not only much lesser, but are rather confined to the amount of liver fat deposition.<sup>31</sup> These results suggest that an excessive caloric intake is more determinant in individuals at genetic risk, as expected.

It remains uncertain whether the gene–environment interactions mentioned above are limited to European ancestry

populations. Hence, a note of caution must be added as the allelic architecture of the human genome substantially varies according to the different ethnic groups. A detailed explanation of population genetics of variants in locus of interest can be found in the International Genome Sample Resource (<http://www.internationalgenome.org/>), which includes data generated by the 1000 Genomes Project (Africa, American, East Asian, European, and South Asian population). An important consideration is that the large majority of genetic studies of NAFLD and NASH remain mostly limited to Caucasian population. Further studies in non-Europeans, which are clearly underrepresented in the big studies currently available, might yield intriguing results, in particular regarding extreme genotypes or rapidly progressive forms of the disease.<sup>32,33</sup>

Although the concept is beyond this review, it is worthy to note that a noticeable portion of the missing heritability of NAFLD may be explained by epigenetic factors,<sup>25,34–38</sup> which may change gene expression by modifying accessibility of transcription machinery to chromatin. Among the most prominent epigenetic factors are DNA methylation and histone covalent modifications. We found that DNA methylation not only of the nuclear but mitochondrial genome loci is associated with NAFLD pathophenotypes.<sup>25,37,39</sup>

Most importantly, knowledge is lacking regarding a broad range of not sufficiently studied interactions, which are not limited to G × E, gene–gene (G × G), and genotype–phenotype (G × P) but other interactions that could explain the variance of the disease (► Fig. 1). It is plausible to presume that the NAFLD–NASH heritability gap might be explained by the intricate relationship among genetic variance of the nuclear and mitochondrial genome, the phenotype, and the yet unexplored interactions with epigenetic and environmental factors, including the microbiome. Future explorations into this interaction network could help unravel the missing heritability of NAFLD.

## Genetic Knowledge of NAFLD: Integrated Pathways of Disease Pathogenesis

Genes associated with a given disease often provide clues on its pathogenesis and mechanisms of tissue-associated damage. For example, variants in *PNPLA3*<sup>40–43</sup> and *TM6SF2*<sup>12,44,45</sup> have been functionally profiled to confirm a putative relationship with and a responsible effect on the variability of liver fat content. When their findings are interpreted jointly, the studies highlighted above yield insights into the significance of liver fat composition and lipid droplet biology and dynamics, as well as patterns of liver fat mobilization, in the pathogenesis of NAFLD.

Despite the wealth of knowledge from early hypothesis-driven genetic studies and genome-wide investigations, the precise mechanisms that explain the variability of the NAFLD phenotype are not fully understood. We also lack an understanding of the precise processes that govern the disease progression, as well as the molecular mechanisms associated with the degree of disease severity.

To offer a framework for overcoming these limitations, we used a tool that expands annotation details of genes/proteins

to perform an integrative analysis. Specifically, we integrated the genes/loci discovered either via candidate-gene association studies or genome-wide investigations into the Protein ANalysis THrough Evolutionary Relationships (PANTHER) database (<http://pantherdb.org>). PANTHER contains comprehensive information on the evolution and function of protein-coding genes from *Homo sapiens* to a wide range of completely sequenced genome. The training set of genes is shown in ►Table 1. This list includes a search in the GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) using the “Nonalcoholic fatty liver disease” search string, as well as genes that have been associated with the genetic risk of NAFLD and NASH.<sup>8,10,46–48</sup> We used the Gene Ontology (GO) data set to infer and integrate information pertinent to biological process of all genes listed in ►Table 1. The top ranked GO biological processes were adipokinetic hormone receptor activity (GO:0097003), adiponectin binding (GO:0055100), retinol O-fatty-acyltransferase activity (GO:0050252), and  $\beta$ -adrenergic receptor activity (GO:0004939) that presented a > 100-fold change enrichment (see ►Table 2 for the complete list that includes *p*-values and fold changes). Overrepresentation and enrichment tests based on Reactome pathways highlighted signaling to signal transducer and activator of transcription 3 (*STAT3*) (R-HSA-198745) (> 100-fold change), acyl chain remodeling of diacylglycerol (DAG) and triacylglycerol (TAG) (R-HSA-1482883), adenosine monophosphate-activated protein kinase-mediated *chREBP* transcriptional activation and caspase activation (R-HSA-163680 and R-HSA-140534, respectively), and chylomicron-mediated lipid transport (R-HSA-174800) as significantly enriched (see ►Table 2 for the complete list). As a result, we may infer that the pathogenesis of NAFLD and NASH is heavily mediated not only by processes associated with TAG and DAG remodeling, but also with hepatocyte response to interleukins/cytokines/adipokines, cell-death immune-mediated pathways, and acute-phase protein genes.

## Shared Molecular Regulatory Pathways of Chronic Liver Damage

Chronic liver diseases, particularly NAFLD and alcoholic liver disease (ALD), share the pathogenic pathways and mechanisms.<sup>49–52</sup> There are also consistent similarities in the pathogenesis of complex cholestatic disorders, including primary biliary cholangitis (formerly known as primary biliary cirrhosis) and primary sclerosing cholangitis.<sup>53–55</sup> Furthermore, chronic liver damage is associated with conserved pathogenic mechanisms, in particular hepatocyte cell death pathways, inflammatory processes that involve immune response, and fibrogenesis.<sup>52,56</sup>

It is therefore biologically plausible to presume that genetic predisposition of convergent pathophenotypes, specifically liver inflammation and fibrosis, is similar, as discussed later.

An interesting example of that is the *PNPLA3*-rs738409 (I148M) variant that was initially discovered in a GWAS of NAFLD. Subsequently, evidence of its involvement in the susceptibility of cirrhosis and end-stage liver disease of

patients with ALD emerged, including the development of alcohol-related cirrhosis<sup>57–59</sup> and HCC.<sup>60,61</sup> In addition, patients homozygous for the risk-G allele of the rs738409 variant seem to be more susceptible to developing severe alcoholic hepatitis, while also having a greater risk of poor survival rates.<sup>62,63</sup> Summarized evidence also suggests an involvement of the rs738409 variant in the risk and severity of chronic hepatitis C.<sup>64</sup>

These remarkable observations suggest that the rs738409 (directly by a *cis* or *trans* eQTL effect of *PNPLA3* gene and/or by coding protein isoforms with diverse functions) might have a causal role in inflammation, fibrosis, and hepatocarcinogenesis. *In vitro* studies showed that *PNPLA3* is required for hepatic stellate cell (HSC) activation and that the rs738409 G variant potentiates the profibrogenic features of HSCs.<sup>65</sup> Although findings yielded by previous studies on *PNPLA3* protein regulation indicate that the adiponutrin protein-family exhibits phospholipase but not retinyl esterase activity,<sup>66,67</sup> some evidence suggests that the rs738409 variant may be involved in retinol release.<sup>68</sup> On the other hand, recent *in vitro* studies showed that overexpression of *PNPLA3*-Met148 variant is associated with 1.75-fold increase in lactic acid, suggesting a shift of cellular response toward anaerobic metabolism and mitochondrial dysfunction.<sup>69</sup> This particular metabolic profile has also been observed in patients with NASH.<sup>70</sup> Furthermore, *PNPLA3* silencing has been associated with global metabolic perturbations that resemble a catabolic response associated with protein breakdown.<sup>69</sup> These metabolic changes may support the involvement of *PNPLA3* in broader metabolic functions in the liver.

More recently, the splice variant rs72613567 in the *HSD17B13* gene was found to protect patients with chronic liver disease, including NAFLD and ALD, from severe and progressive damage, regardless of the etiology.<sup>23</sup> These findings were replicated in two recent studies.<sup>71,72</sup>

There are other liver-related traits, such as serum aminotransferase levels, of which the genetic component of variability is highly influenced by the aforementioned variants, irrespective of the underlying cause of liver disease.<sup>23,73</sup>

While shared biology and genetics might explain the pathogenesis of chronic liver damage, the magnitude of the loci that are potentially involved in shared mechanisms is unknown. Based on the available evidence, *PNPLA3* and probably *TM6SF2* could explain commonality in pathogenic pathways of metabolic liver disease. However, some interesting observations suggest that other genes might potentially influence the shared mechanisms of liver damage. For example, variants in nuclear receptor subfamily 1, group I, member 2 (nuclear *pregnane X receptor*) have been associated with NAFLD predisposition<sup>74</sup> and with drug clearance and drug-induced liver injury.<sup>75</sup>

Variants/mutations in homeostatic iron regulator (HFE), a membrane protein that is similar to major histocompatibility complex class I-type proteins and that is involved in iron storage disorder, have been associated not only with hereditary hemochromatosis<sup>76</sup> but also with an increased risk of HCC in patients with alcoholic cirrhosis.<sup>77</sup> HFE variants have been also involved in the susceptibility of NAFLD,<sup>78,79</sup>

**Table 2** Integrated pathways of disease pathogenesis: Gene Ontology (GO) molecular function and Reactome prediction of NAFLD-predisposing genes

Annotation data set	Fold enrichment	Raw <i>p</i> -value	FDR
<b>GO molecular function (GO annotation number)</b>			
Adipokinetic hormone receptor activity (GO:0097003)	> 100	1.51E-04	3.91E-02
Adiponectin binding (GO:0055100)	> 100	2.51E-04	5.31E-02
Retinol O-fatty-acyltransferase activity (GO:0050252)	> 100	2.51E-04	5.08E-02
Beta-adrenergic receptor activity (GO:0004939)	> 100	2.51E-04	4.87E-02
Long-chain fatty acid binding (GO:0036041)	45.38	6.72E-05	2.61E-02
Acylglycerol O-acyltransferase activity (GO:0016411)	27.12	2.26E-05	1.75E-02
Fatty acid binding (GO:0005504)	22.47	4.44E-05	2.30E-02
RNA polymerase II repressing transcription factor binding (GO:0001103)	21.26	5.43E-05	2.53E-02
O-acyltransferase activity (GO:0008374)	15.73	1.61E-04	3.95E-02
Nuclear receptor activity (GO:0004879)	15.73	1.61E-04	3.75E-02
Transcription factor activity, direct ligand regulated sequence-specific DNA binding (GO:0098531)	15.73	1.61E-04	3.57E-02
Monocarboxylic acid binding (GO:0033293)	14.9	3.01E-05	1.75E-02
Carboxylic acid binding (GO:0031406)	7.25	6.50E-05	2.75E-02
Organic acid binding (GO:0043177)	6.85	9.14E-05	2.66E-02
Cytokine receptor binding (GO:0005126)	5.6	1.12E-04	3.06E-02
Lipid binding (GO:0008289)	3.92	1.52E-05	1.76E-02
Identical protein binding (GO:0042802)	2.99	2.03E-07	9.42E-04
Protein dimerization activity (GO:0046983)	2.91	2.70E-05	1.79E-02
Signaling receptor binding (GO:0005102)	2.7	1.10E-05	1.70E-02
Molecular function regulator (GO:0098772)	2.51	2.22E-05	2.07E-02
Enzyme binding (GO:0019899)	2.45	6.17E-06	1.44E-02
Protein binding (GO:0005515)	1.33	8.31E-05	2.58E-02
Molecular function (GO:0003674)	1.15	6.75E-05	2.42E-02
<b>Reactome pathway (identifier number)</b>			
Signaling to STAT3 (R-HSA-198745)	> 100	2.51E-04	4.54E-02
Acyl chain remodeling of DAG and TAG (R-HSA-1482883)	84.28	1.47E-05	7.33E-03
AMPK inhibits chREBP transcriptional activation activity (R-HSA-163680)	73.75	2.02E-05	6.69E-03
Ligand-dependent caspase activation (R-HSA-140534)	42.14	8.13E-05	1.80E-02
Chylomicron-mediated lipid transport (R-HSA-174800)	25.65	3.01E-04	4.28E-02
Caspase activation via extrinsic apoptotic signaling pathway (R-HSA-5357769)	24.58	3.37E-04	4.19E-02
Transcriptional regulation of white adipocyte differentiation (R-HSA-381340)	12.61	6.40E-05	1.59E-02
Lipid digestion, mobilization, and transport (R-HSA-73923)	11.8	1.65E-05	6.56E-03
Glycerophospholipid biosynthesis (R-HSA-1483206)	9.02	2.86E-04	4.74E-02
PPARA activates gene expression (R-HSA-1989781)	8.94	2.98E-04	4.56E-02
Regulation of lipid metabolism by Peroxisome proliferator-activated receptor $\alpha$ (PPAR $\alpha$ ) (R-HSA-400206)	8.78	3.23E-04	4.28E-02
Metabolism of vitamins and cofactors (R-HSA-196854)	8.08	1.24E-04	2.47E-02
Fatty acid, triacylglycerol, and ketone body metabolism (R-HSA-535734)	6.99	2.46E-05	6.99E-03
Metabolism of lipids and lipoproteins (R-HSA-556833)	5.25	4.26E-09	4.24E-06
Metabolism (R-HSA-1430728)	3.2	2.10E-09	4.17E-06

Abbreviations: DAG, diacylglycerol; FDR, false discovery rate; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis; TAG, triacylglycerol.

Note: Enrichment analysis was performed by the PANTHER software available at <http://pantherdb.org/>; <sup>120</sup> analysis type: PANTHER Overrepresentation Test (Released December 5, 2017). Annotation version: PANTHER version 13.1 and Reactome version 58.

Statistical analysis: Fisher's exact test with false discovery rate (FDR) multiple test correction. Analyzed list: training set was the list of genes associated with NAFLD or NASH in candidate gene association studies or genome-wide approaches (see ► **Table 1**).

Reference list: Homo sapiens (all genes in the database).

although findings yielded by a systematic review of available data does not support this association.<sup>47</sup>

A missense (p.Glu366Lys, also known as PI<sup>+</sup>Z) variant in *SERPINA1* (serpin family A member 1) gene that is known as a predisposing factor for developing  $\alpha$ -1-antitrypsin deficiency<sup>80</sup> has been recently associated with the risk of cirrhosis in NAFLD and alcohol misuse.<sup>81</sup>

### NAFLD Genes and Pleiotropy: Cross-Associations between NAFLD-Predisposing Genes and Phenotypes of the Metabolic Syndrome

Genome-wide association of complex diseases have demonstrated that a large number of SNPs are implicated in the susceptibility of multiple—not necessarily related—traits. The effect of one gene on different phenotypes is known as pleiotropy.<sup>82,83</sup> While the concept of pleiotropy has been largely confined to the field of evolutionary biology, it become evident during the past 10 years, when genome-wide approaches revealed cross-phenotype associations among a broad category of complex traits.<sup>84</sup> In fact, it is estimated that approximately 4.6% of SNPs discovered by GWAS show pleiotropic effects,<sup>84</sup> and 44% of genes reported in the GWAS Catalog are associated with more than one phenotype.<sup>85</sup>

Gene-based connectivity network based on gene/protein cooccurrence suggests genetic commonality between NAFLD and features of the metabolic syndrome (MetS), specifically obesity, type 2 diabetes, and arterial hypertension.<sup>86</sup> For example, a rare nonsense (rs149847328, p.Arg227Ter) mutation in *GCKR* was associated with a rapidly progressive clinical form of NASH, which might be the first rare genetic form of the disease.<sup>33</sup> Interestingly, *GCKR* is considered a susceptibility gene for a form of maturity-onset diabetes of the young.<sup>87</sup>

There are, however, paradoxical examples of alleles that impart risk of developing NAFLD but are protective against phenotypes that are closely related with the disease, including cardiovascular disease (CVD). For example, carriers of the minor T allele (EK + KK) of the *TM6SF2* E167K (rs58542926 C/T) variant are protected from CVD, including myocardial infarction,<sup>88</sup> and show low levels of total plasma cholesterol, low-density lipoprotein cholesterol, and triglycerides.<sup>12,14,18,44,89</sup> At the same time, the minor-T allele of rs58542926 is a risk factor for NAFLD and NASH.<sup>12,14,17,18,89</sup>

Together, these observations highlight the concept of shared genetic basis of diverse phenotypes. This assumption not only fits into biologically meaningful associations, for example, immune-mediated and/or metabolic diseases, but also traits/diseases that a priori present a certain level of dissimilarity in their pathogenic mechanisms.

We explored the extent of pleiotropy of loci known to be associated with the genetic risk of NAFLD and NASH. This exploration was performed by literature-enrichment analysis offered by the Genset2Diseases (GS2D) Web server (<http://cbdm.uni-mainz.de/geneset2diseases>)—a tool that computes associations of genes with diseases using biomedical literature annotations.<sup>90</sup> GS2D algorithm prioritizes all human genes according to their relation to a biomedical topic

using all available scientific abstracts and orthology information.<sup>90</sup>

As expected, we found that a high proportion of NAFLD-related genes (listed in ▶Table 1) are also involved in the pathogenesis of phenotypes of the MetS (▶Table 3). Examples of shared NAFLD and MetS-related loci include *PPARGC1A*, a master regulator of carbohydrates and fat metabolism and mitochondrial function that has been associated with NAFLD, insulin resistance, and liver mitochondrial copy number,<sup>25</sup> as well as with cardiac development<sup>91</sup> and cardiac disease.<sup>92</sup> Another example is clock circadian regulator (*CLOCK*) that has been linked to MetS in rodents<sup>93</sup> and in humans, specifically obesity<sup>94,95</sup> and NAFLD.<sup>96</sup>

It could be argued that this analysis is inflated by highly correlated traits and outcomes, such as NAFLD, type 2 diabetes, insulin resistance, atherosclerosis, dyslipidemia, etc. Nevertheless, the analysis offered some surprising findings as well. For example, 7 out of 104 input genes were significantly associated in the literature with ulcerative colitis or premature birth (▶Table 3), and 5 out of 104 were associated with abortion, sepsis, and nasal polyps (▶Table 3). These results, however, must be interpreted with caution, as further work on the confirmation of causality and curation of data are needed. Still, it is expected that—if confirmed—these results may open a window for therapeutic explorations, whereby drugs can be designed to focus on pleiotropic loci or pleiotropic molecular targets that cover multiple traits, even though those traits are not obviously associated.

### Genetics of NAFLD and Precision Medicine

With the advances of the genetic knowledge of NAFLD and NASH, it becomes possible to use this information for clinical applications. Genetic data could be leveraged to identify individuals at risk of NAFLD, or to estimate the risk of severe histological outcomes, including NASH and NASH-fibrosis (▶Fig. 2). Genetic markers are already being used as tools for personalized clinical practice, including treatment decisions (▶Fig. 2). Specifically, *PNPLA3*-rs738409 was incorporated into combined screening algorithms that included clinical and biochemical data. Nevertheless, the utility of the variant in NAFLD risk estimation remains inferior to classical predictive or imaging approaches. For example, Kotronen et al proposed the *NAFLD liver fat score*, which showed an area under the receiver operating characteristic curve (AUROC) of 0.872 (95% confidence interval [CI]: 0.84–0.91) in predicting liver fat content.<sup>97</sup> The addition of rs738409 to the score composed by the presence of type 2 diabetes, along with the level of serum fasting insulin and aminotransferases, improved the prediction accuracy by only < 1%.<sup>97</sup> A more sophisticated multipanel score—the *NAFLD multicomponent score*—which integrates omics-derived and clinical variables, the rs738409, and proteomic data, showed an AUROC of 0.932 for the NAFLD population risk identification.<sup>98</sup> Despite this high predictive value, this biomarker panel would be neither practical nor cost-effective for large-scale population screening programs.



**Table 3** Extent of pleiotropy in NAFLD-predisposing genes

Disease	Genes count	Fold change	p-Value	FDR	Genes <sup>#</sup> (numbers of times that relevant citations regarding each gene appear in biomedical literature)
Metabolic syndrome	21	10.15	0.000 e <sup>+00</sup>	0.000e <sup>+00</sup>	ADRB2 <sup>9</sup> , ADRB3 <sup>22</sup> , AGTR1 <sup>10</sup> , APOC3 <sup>20</sup> , APOE <sup>35</sup> , FABP2 <sup>13</sup> , GCKR <sup>10</sup> , IRS1 <sup>8</sup> , LEPR <sup>20</sup> , LIPC <sup>8</sup> , MC4R <sup>9</sup> , MTPP <sup>7</sup> , ENPP1 <sup>11</sup> , PPARG <sup>51</sup> , UCP1 <sup>5</sup> , ADIPOQ <sup>163</sup> , CLOCK <sup>7</sup> , PPARGC1A <sup>8</sup> , ADIPOR1 <sup>7</sup> , ADIPOR2 <sup>5</sup> , PNPLA3 <sup>10</sup>
Insulin resistance	22	6.08	5.038e <sup>-12</sup>	9.236e <sup>-11</sup>	ADRB3 <sup>35</sup> , APOC3 <sup>15</sup> , FABP2 <sup>28</sup> , GCKR <sup>8</sup> , IL6 <sup>53</sup> , IRS1 <sup>99</sup> , LEPR <sup>30</sup> , LIPC <sup>10</sup> , MC4R <sup>8</sup> , MTPP <sup>8</sup> , ENPP1 <sup>43</sup> , PPARG <sup>10</sup> , PPARG <sup>141</sup> , TCF7L2 <sup>48</sup> , TNF <sup>101</sup> , UCP1 <sup>9</sup> , DGAT1 <sup>5</sup> , ADIPOQ <sup>200</sup> , PPARGC1A <sup>28</sup> , ADIPOR1 <sup>31</sup> , ADIPOR2 <sup>28</sup> , PNPLA3 <sup>21</sup>
Hypertriglyceridemia	7	14.31	5.279e <sup>-07</sup>	5.807e <sup>-06</sup>	APOC3 <sup>38</sup> , APOE <sup>36</sup> , FABP2 <sup>6</sup> , GCKR <sup>8</sup> , LIPC <sup>8</sup> , PPARG <sup>8</sup> , ADIPOQ <sup>6</sup>
Morbid obesity	8	10.85	6.807e <sup>-07</sup>	6.240e <sup>-06</sup>	ADRB3 <sup>8</sup> , LEPR <sup>13</sup> , MC4R <sup>19</sup> , PPARG <sup>18</sup> , UCP1 <sup>7</sup> , ADIPOQ <sup>24</sup> , PPARGC1A <sup>9</sup> , PNPLA3 <sup>6</sup>
Dyslipidemias	8	10.32	9.975e <sup>-07</sup>	7.837e <sup>-06</sup>	ADRB2 <sup>5</sup> , APOC3 <sup>17</sup> , APOE <sup>39</sup> , GCKR <sup>6</sup> , LIPC <sup>9</sup> , PPARG <sup>13</sup> , PPARG <sup>11</sup> , ADIPOQ <sup>11</sup>
Alcoholic liver diseases	5	26.58	1.018e <sup>-06</sup>	6.999e <sup>-06</sup>	CD14 <sup>6</sup> , CYP2E1 <sup>10</sup> , HFE <sup>9</sup> , TNF <sup>10</sup> , PNPLA3 <sup>8</sup>
Obesity	21	3.08	2.624e <sup>-06</sup>	1.604e <sup>-05</sup>	ADRB2 <sup>81</sup> , ADRB3 <sup>99</sup> , FABP2 <sup>27</sup> , GCKR <sup>11</sup> , IRS1 <sup>31</sup> , LEPR <sup>152</sup> , LIPC <sup>21</sup> , MC4R <sup>199</sup> , ENPP1 <sup>38</sup> , PPARG <sup>26</sup> , PPARG <sup>174</sup> , TCF7L2 <sup>43</sup> , UCP1 <sup>46</sup> , DGAT1 <sup>6</sup> , ADIPOQ <sup>200</sup> , CLOCK <sup>16</sup> , PPARGC1A <sup>28</sup> , ADIPOR1 <sup>21</sup> , ADIPOR2 <sup>13</sup> , PNPLA3 <sup>44</sup> , LYPLAL1 <sup>8</sup>
Chronic periodontitis	6	13.99	4.096e <sup>-06</sup>	2.253e <sup>-05</sup>	CD14 <sup>7</sup> , IL1B <sup>33</sup> , IL6 <sup>21</sup> , PTGS2 <sup>10</sup> , TLR4 <sup>7</sup> , TNF <sup>13</sup>
Diabetes mellitus, type 2	25	2.52	9.088e <sup>-06</sup>	4.544e <sup>-05</sup>	ADRB3 <sup>37</sup> , AGTR1 <sup>30</sup> , APOC3 <sup>34</sup> , APOE <sup>111</sup> , FABP2 <sup>39</sup> , GC <sup>9</sup> , GCKR <sup>52</sup> , IL6 <sup>98</sup> , IRS1 <sup>85</sup> , LEPR <sup>32</sup> , LIPC <sup>25</sup> , MC4R <sup>32</sup> , MTPP <sup>8</sup> , ENPP1 <sup>60</sup> , PPARG <sup>28</sup> , PPARG <sup>200</sup> , SOD2 <sup>25</sup> , TCF7L2 <sup>200</sup> , TNF <sup>123</sup> , UCP1 <sup>17</sup> , ADIPOQ <sup>200</sup> , PPARGC1A <sup>75</sup> , ADIPOR1 <sup>26</sup> , ADIPOR2 <sup>22</sup> , LYPLAL1 <sup>5</sup>
Polycystic ovary syndrome	9	5.75	2.564e <sup>-05</sup>	1.175e <sup>-04</sup>	IL6 <sup>16</sup> , IRS1 <sup>25</sup> , PPARG <sup>29</sup> , TCF7L2 <sup>16</sup> , TNF <sup>18</sup> , ADIPOQ <sup>52</sup> , PPARGC1A <sup>5</sup> , ADIPOR1 <sup>6</sup> , ADIPOR2 <sup>5</sup>
Overweight	7	7.82	3.058e <sup>-05</sup>	1.294e <sup>-04</sup>	ADRB2 <sup>7</sup> , ADRB3 <sup>7</sup> , LEPR <sup>10</sup> , MC4R <sup>10</sup> , ADIPOQ <sup>48</sup> , CLOCK <sup>6</sup> , PNPLA3 <sup>11</sup>
Hyperlipidemias	6	9.38	4.170e <sup>-05</sup>	1.638e <sup>-04</sup>	ADRB3 <sup>5</sup> , APOC3 <sup>8</sup> , APOE <sup>67</sup> , FABP2 <sup>6</sup> , LIPC <sup>10</sup> , PPARG <sup>11</sup>
Atherosclerosis	14	3.39	5.669e <sup>-05</sup>	2.079e <sup>-04</sup>	AGTR1 <sup>10</sup> , APOC3 <sup>12</sup> , APOE <sup>61</sup> , CD14 <sup>11</sup> , GCKR <sup>5</sup> , IL6 <sup>38</sup> , LIPC <sup>7</sup> , MIF <sup>9</sup> , PPARG <sup>12</sup> , PPARG <sup>29</sup> , TLR4 <sup>33</sup> , TNF <sup>45</sup> , ADIPOQ <sup>57</sup> , PPARGC1A <sup>7</sup>
Diabetes, gestational	6	8.05	9.822e <sup>-05</sup>	3.376e <sup>-04</sup>	IRS1 <sup>7</sup> , PPARG <sup>16</sup> , TCF7L2 <sup>17</sup> , TNF <sup>15</sup> , ADIPOQ <sup>37</sup> , PPARGC1A <sup>5</sup>
Weight loss	6	7.67	1.291e <sup>-04</sup>	4.177e <sup>-04</sup>	ADRB3 <sup>11</sup> , FABP2 <sup>5</sup> , LEPR <sup>11</sup> , MC4R <sup>16</sup> , ADIPOQ <sup>28</sup> , CLOCK <sup>6</sup>
Coronary artery disease	14	2.81	4.014e <sup>-04</sup>	1.226e <sup>-03</sup>	AGTR1 <sup>23</sup> , APOC3 <sup>34</sup> , APOE <sup>85</sup> , CD14 <sup>17</sup> , FABP2 <sup>5</sup> , GCKR <sup>9</sup> , IL6 <sup>63</sup> , LIPC <sup>26</sup> , MTHFR <sup>89</sup> , PPARG <sup>14</sup> , PPARG <sup>33</sup> , ADIPOQ <sup>82</sup> , PPARGC1A <sup>9</sup> , ADIPOR1 <sup>7</sup>
Periodontitis	5	7.64	4.909e <sup>-04</sup>	1.421e <sup>-03</sup>	CD14 <sup>15</sup> , IL1B <sup>68</sup> , IL6 <sup>28</sup> , TLR4 <sup>24</sup> , TNF <sup>33</sup>
Diabetic nephropathies	8	4.29	5.476e <sup>-04</sup>	1.506e <sup>-03</sup>	AGTR1 <sup>20</sup> , APOE <sup>27</sup> , MTHFR <sup>29</sup> , ENPP1 <sup>9</sup> , PPARG <sup>27</sup> , SOD2 <sup>8</sup> , TCF7L2 <sup>7</sup> , ADIPOQ <sup>31</sup>
Glucose intolerance	5	6.71	8.854e <sup>-04</sup>	2.319e <sup>-03</sup>	IRS1 <sup>7</sup> , LEPR <sup>6</sup> , PPARG <sup>12</sup> , TCF7L2 <sup>14</sup> , ADIPOQ <sup>31</sup>
Premature birth	7	4.45	9.875e <sup>-04</sup>	2.469e <sup>-03</sup>	ADRB2 <sup>10</sup> , CD14 <sup>7</sup> , IL1B <sup>19</sup> , IL6 <sup>33</sup> , MTHFR <sup>17</sup> , TLR4 <sup>13</sup> , TNF <sup>29</sup>
Nasal polyps	5	5.54	2.085e <sup>-03</sup>	4.986e <sup>-03</sup>	CFTR <sup>7</sup> , IL1B <sup>7</sup> , IL6 <sup>9</sup> , PTGS2 <sup>9</sup> , TNF <sup>10</sup>
Abortion, habitual	5	5.07	3.053e <sup>-03</sup>	6.996e <sup>-03</sup>	APOE <sup>9</sup> , IL1B <sup>12</sup> , IL6 <sup>16</sup> , MTHFR <sup>68</sup> , TNF <sup>24</sup>
<i>Helicobacter</i> infections	6	3.69	5.725e <sup>-03</sup>	1.260e <sup>-02</sup>	CD14 <sup>10</sup> , IL1B <sup>117</sup> , MIF <sup>8</sup> , PTGS2 <sup>33</sup> , TLR4 <sup>37</sup> , TNF <sup>57</sup>
Hepatitis B, chronic	5	4.18	6.921e <sup>-03</sup>	1.464e <sup>-02</sup>	HFE <sup>7</sup> , IL6 <sup>20</sup> , MIF <sup>7</sup> , TNF <sup>44</sup> , PNPLA3 <sup>8</sup>
Sepsis	5	3.36	1.673e <sup>-02</sup>	3.408e <sup>-02</sup>	CD14 <sup>41</sup> , IL6 <sup>51</sup> , MIF <sup>12</sup> , TLR4 <sup>46</sup> , TNF <sup>55</sup>
Diabetes mellitus	7	2.63	1.719e <sup>-02</sup>	3.377e <sup>-02</sup>	PPARG <sup>10</sup> , PPARG <sup>35</sup> , TCF7L2 <sup>20</sup> , UCP1 <sup>6</sup> , PPARGC1A <sup>10</sup> , ADIPOR1 <sup>7</sup> , ADIPOR2 <sup>6</sup>
Colitis, ulcerative	7	2.53	2.051e <sup>-02</sup>	3.890e <sup>-02</sup>	CD14 <sup>12</sup> , IL1B <sup>20</sup> , MIF <sup>8</sup> , STAT3 <sup>18</sup> , TLR4 <sup>28</sup> , TNF <sup>55</sup> , NR1I2 <sup>5</sup>
Pulmonary disease, chronic obstructive	7	2.47	2.336e <sup>-02</sup>	4.283e <sup>-02</sup>	ADRB2 <sup>23</sup> , CFTR <sup>13</sup> , GC <sup>16</sup> , GCLC <sup>5</sup> , IL6 <sup>32</sup> , SERPINA1 <sup>45</sup> , TNF <sup>55</sup>

Abbreviations: EWAS, exome-wide association study; FDR, false discovery rate; GWAS, genome-wide association study; NAFLD, nonalcoholic fatty liver disease; NASH, nonalcoholic steatohepatitis.

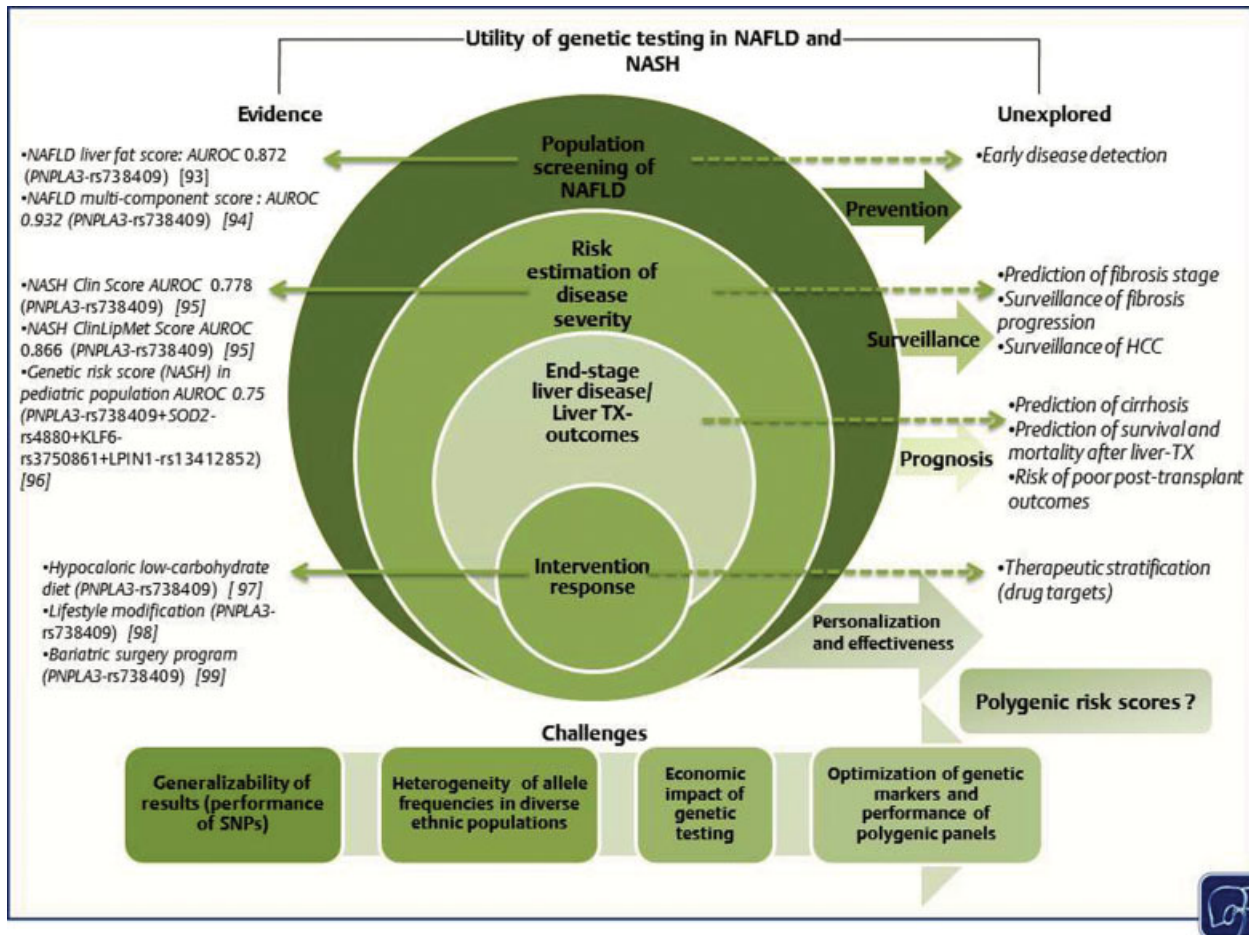
Note: The exploration was performed by literature-enrichment analysis offered by the *Geneset2Diseases* (GS2D) Web server (<http://cbdm.uni-mainz.de/geneset2diseases>), a tool that computes associations of genes with diseases using biomedical literature annotations.<sup>90</sup> The training set consisted of a list of genes extracted from published gene associations with NAFLD and NASH in candidate-gene association studies and genome-wide approaches (GWAS and EWAS); the full list is shown in ► **Table 1**.

Disease: Disease term from the MeSH vocabulary (based on biomedical references represented by MEDLINE records).

Genes count: The search was restricted using the following filters: For a gene set, minimum number of genes significantly associated with a disease = 5 and minimum number of disease-related citations for a gene = 5.

Fold change: (number of input genes significantly associated with the disease in the literature / number of input genes) / (total number of genes significantly associated with the disease in the literature / total number of genes).

p-Value: Computed by Fisher's exact test; FDR computed by Benjamini-Hochberg method. Gene #: list of genes (gene symbols) of input genes significantly associated with the disease and, in superscript, numbers of relevant citations in the literature.



**Fig. 2** Genetics of nonalcoholic fatty liver disease (NAFLD) and precision medicine. This figure shows examples of the use of genetic markers in the clinical setting, as well as potential yet unexplored applications.

Risk estimation of the disease severity and progression, including NASH and NASH-fibrosis, offers greater opportunities of clinical translation. In fact, the use of genetic testing might open a window for the development of gene-based strategies for the diagnosis of NASH, thus moving the diagnosis of the disease severity from an invasive (liver biopsy) toward a noninvasive approach. Unfortunately, there is still no evidence of superiority in terms of efficacy and accuracy of rs738409—or other variants—in predicting liver histology as compared with the liver biopsy. For example, a combination of laboratory test (aspartate transaminase and fasting insulin), circulating metabolites, and rs738409 genotypes into the *NASH Clinical Score* and the *NASH ClinLipMet score* showed NASH prediction value; the AUROC for NASH was 0.778 (95% CI: 0.709–0.846) and 0.866 (95% CI: 0.820–0.913), respectively.<sup>99</sup> Similar explorations have been conducted in pediatric settings, in which a polygenic risk score that included combinations of variants in four loci (*PNPLA3*-rs738409, *SOD2*-rs4880, *KLF6*-3750861, and *LPIN1*-13412852) showed an AUROC for NASH of 0.75 (95% CI: 0.67–0.82).<sup>100</sup> It has to be emphasized that the more variants (a worse scenario if minor allele frequencies are low) are included in a polygenic score, the lower the frequency of individuals at risk will be found.

It should be noted that genetic assessments provide static information for the explored phenotype or disease trait. However, genetic markers could be used to dynamically predict the response to any therapeutic intervention, as shown in ► **Fig. 2**. For example, findings yielded by pilot studies indicate that information of the homozygosity status of rs738409 risk-G allele was useful in predicting the absolute change in liver fat content of patients enrolled in a program of hypocaloric low-carbohydrate diet<sup>101</sup> or reduced caloric intake.<sup>102</sup> The variant in *PNPLA3* seems to be also useful in predicting changes in body weight of morbidly NAFLD patients enrolled in a bariatric surgery program.<sup>103</sup>

Potential avenues for future research that would significantly affect prevention, surveillance, and prognosis assessment of NAFLD and NASH are summarized in ► **Fig. 2**. Poorly explored but promising uses of genetic markers include, for example, surveillance of HCC that could occur in cirrhotic and noncirrhotic patients with NASH, or assessment of liver transplantation prognostic outcomes (► **Fig. 2**). The potential interplay between the recipient and donor genotype of variants of interest suggests an interesting yet poorly explored research avenue in the field of precision medicine. Findings yielded by a small number of studies suggest that the *PNPLA3*-rs738409 G allele in either the donor or the

recipient could be a risk factor for NAFLD recurrence or appearance after liver transplantation.<sup>104,105</sup>

The potential value of using genetic markers for treatment decisions pertaining to patients enrolled in NASH clinical trials remains largely unexploited. Nonetheless, it is expected that this specific clinical application will be explored in the near future as the use of novel drugs for the treatment of NASH becomes available in the market. Potential shortcomings and limitations of genetic markers in clinical decision making are shown in ►Fig. 2.

Finally, while it is known that NAFLD is a polygenic and complex disease, the use of polygenic risk scores in the NASH diagnosis and prognosis and its interaction with environmental exposure remain largely unknown. Yet, the use of polygenic risk scores in personalized NAFLD care should be tested and optimized to perform well in diverse ethnic groups because the frequency of the risk alleles varies significantly among populations.<sup>2,11–14,23,106</sup> Remarkable examples of allele frequency disparity among populations are *PNPLA3*-rs738309, of which the frequency of the G-risk allele varies from 12% in African population to 48% in South American (Mexican, Colombian, Peruvian, and Puerto Rican) population (as shown in <http://www.ensembl.org>), and *HSD17B13*-rs72613567, of which the frequency of the A-protective insertion allele varies from 5% in African population to 34% among East Asian population (figures of population genetics were extracted from the 1000 Genomes Project, <http://www.internationalgenome.org/>).

## Nonalcoholic Steatohepatitis Treatment Inferred from Genetic Discoveries

There are currently no approved pharmacologic therapies for NASH. However, many novel drugs are being tested for safety and efficacy.<sup>107</sup> Some of these drugs have been designed based on the available knowledge of NAFLD pathogenesis and the underlying mechanisms of the disease progression, including metabolic pathways, inflammatory cascades, and/or fibrogenesis.<sup>107</sup>

Patients with NAFLD and NASH currently receive lifestyle recommendations, and are eventually medicated with known and relatively safe drugs, for example,  $\alpha$  tocopherol (vitamin E), ursodeoxycholic acid (UDCA), metformin, losartan, or the insulin sensitizer pioglitazone,<sup>107,108</sup> which are already available on the market. These drugs are usually prescribed not necessarily for the treatment of NASH but for the treatment of associated comorbidities, for example, type 2 diabetes and arterial hypertension. Hence, their use in the treatment of NASH is purely empirical and/or pragmatic, guided by the assumption of a putative effect on the disease. Despite this limitation, some of the commonly prescribed drugs, including vitamin E and pioglitazone, have been shown to lead to a partial improvement in liver outcomes, such as liver enzymes.<sup>108</sup>

To answer the question of whether medications that patients receive in ordinary clinical practice are in line with disease mechanisms inferred from genetic discoveries, we performed text-mined chemical–gene–disease interac-

tions by the Comparative Toxicogenomics Database (CTD; <http://ctdbase.org>). We specifically modeled the interaction network among genes associated with NAFLD and NASH that was reported in previous studies (►Table 1), genes associated with fibrosis and inflammation (mined from the curated gene–disease associations that are established by both the CTD data set and Online Mendelian Inheritance in Man), and drugs that have been used or are currently in use for the treatment of NASH ( $\alpha$  tocopherol–vitamin E, UDCA, metformin, pioglitazone, losartan, and liraglutide).

It is evident from the above that translating the information generated from NAFLD genetic studies into new treatment drugs and/or clinical biomarkers was a major challenge. In fact, data generated from either candidate–gene association studies or genome-wide surveys have not been exploited for drug discovery, even though some genes involved are shared by NAFLD and general processes, such as inflammation and fibrosis (►Fig. 3). Nevertheless, we obtained some remarkable results. For example, some of these drugs—including vitamin E, pioglitazone, and even losartan—are predicted to target genes associated with the genetic risk of NAFLD or NASH (►Fig. 3); conversely, liraglutide seems not to match any genes discovered in genetic studies (►Fig. 3).

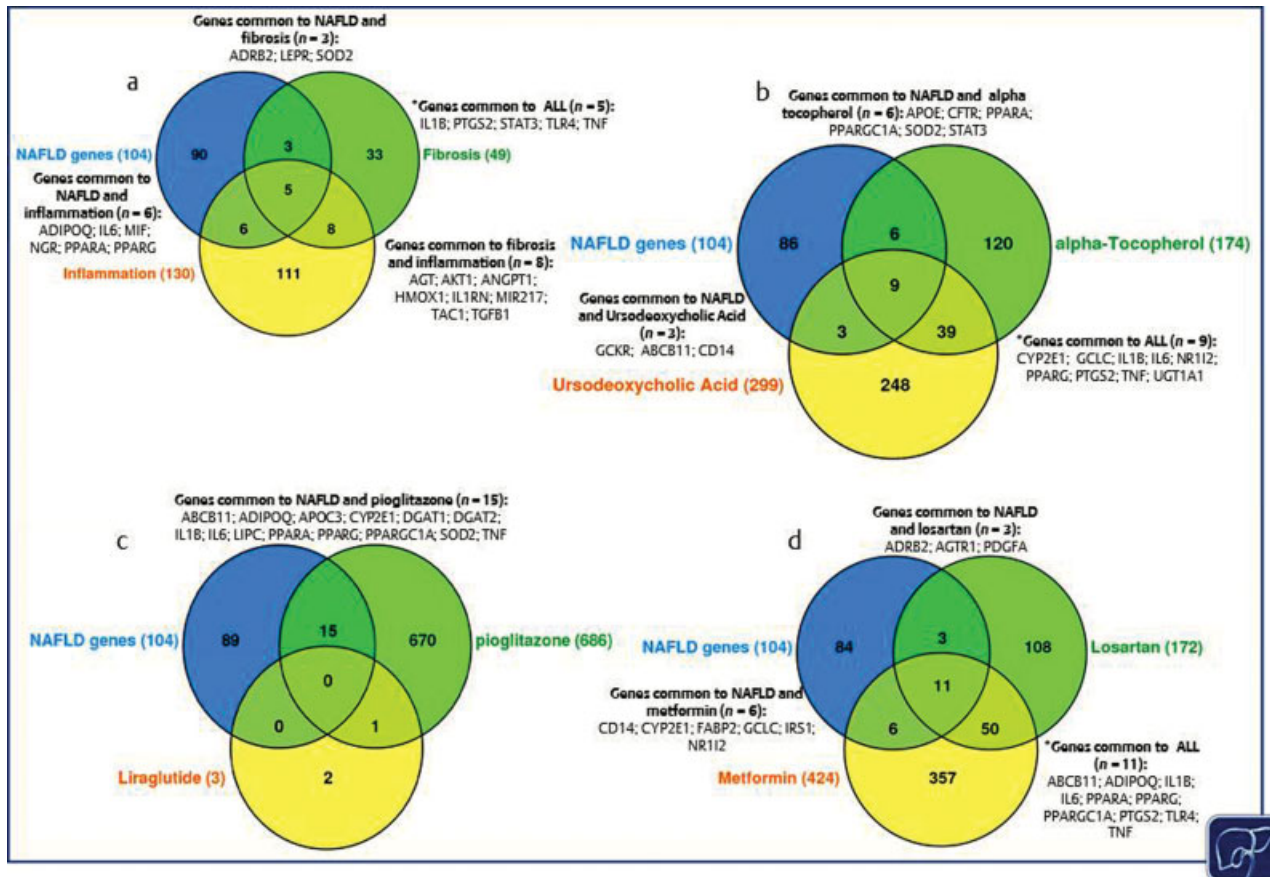
Particularly interesting are the following targets: peroxisome proliferator-activated receptor alpha (*PPAR $\alpha$* ) and peroxisome proliferator-activated receptor gamma (*PPAR $\gamma$* ) and its coactivator *PPARG* coactivator 1 alpha (*PGC1 $\alpha$* ), *STAT3*, adrenoceptor beta 2 (*ADRB3*), and tumor necrosis factor (*TNF*).

Elafibranor (code name GFT505), a dual *PPAR $\alpha$*  and *PPAR $\delta$*  ligand that is currently in phase III, has been proven to consistently ameliorate histological outcomes associated with the disease severity.<sup>110</sup> This pharmacological agent, which has been specifically designed to target *PPARs*, represents a remarkable example of a drug with potentially pleiotropic and systemic effects.<sup>111</sup>

## NAFLD and NASH Genes and the Druggable Proteome

Variants associated with the greatest effects on NAFLD and NASH are indeed missense SNPs (*PNPLA3*-I148M and *TM6SF2*-E167K) that not only explain modest changes in gene/protein expression levels but hardly represent “druggable” targets.<sup>106</sup> These two loci present either pleiotropic metabolic effects,<sup>69</sup> or are associated with dual and opposite effects on critical phenotypes, particularly *TM6SF2*-E167K variant, as already mentioned.<sup>14</sup> Hence, the potential use of these proteins as pharmacological targets by modulating their protein and/or enzymatic activity is rather limited.<sup>106</sup>

As a proof-of-concept, we performed an *in silico* “druggability” prediction of known NAFLD GWAS-discovered genes—including *PNPLA3*, *TM6SF2*, *GCKR*, and *HSD17B13*—based on protein structural druggability, ligand-based druggability, and network-based druggability implemented by the canSAR resource (<http://cansar.icr.ac.uk/>). This resource contains information of the whole human proteome, as well as



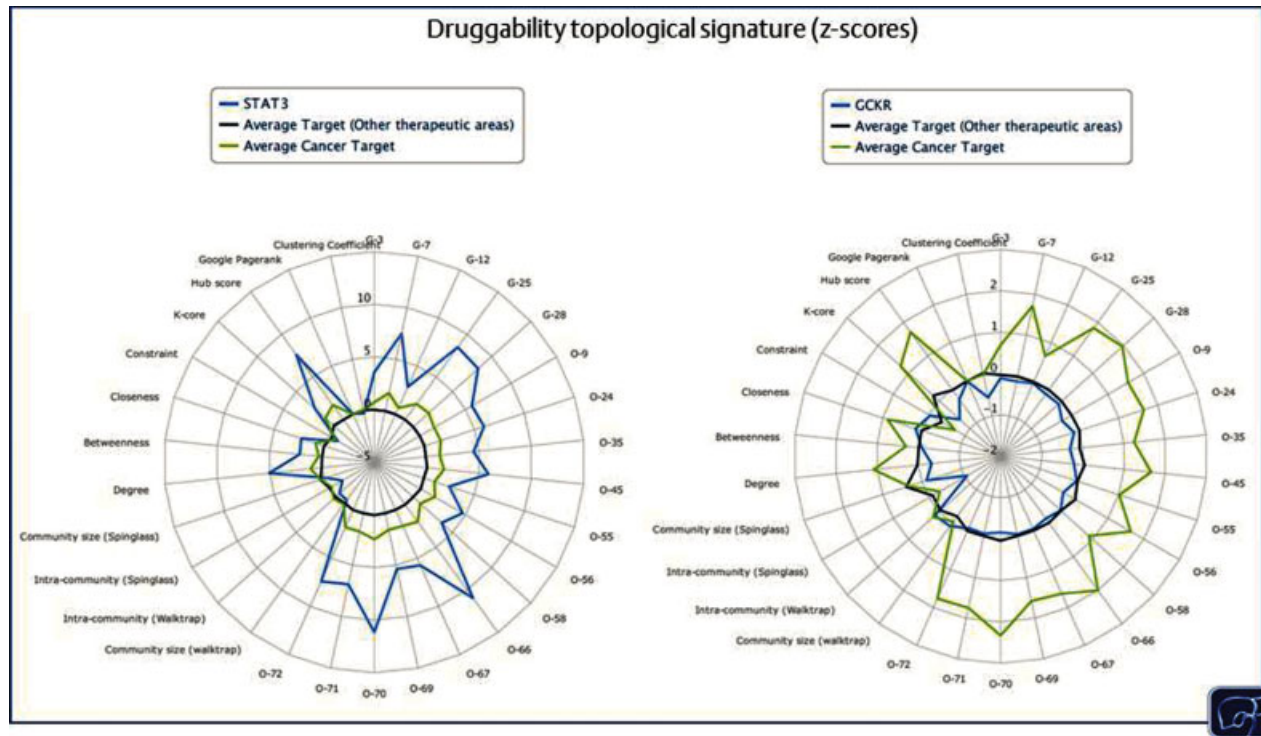
**Fig. 3** Nonalcoholic fatty liver disease (NAFLD) genes and drug interaction network. This figure shows the shared genes associated with NAFLD and nonalcoholic steatohepatitis (NASH) (see ▶Table 1 for the training set of genes,  $n = 104$ ), genes associated with fibrosis ( $n = 49$ ), and inflammation ( $n = 130$ ) (the list of genes under these terms is automatically established from the Comparative Toxicogenomics Database [CTD] data set) (A), and drugs that have been used or are currently in use for the treatment of NASH:  $\alpha$ -tocopherol/vitamin E, ursodeoxycholic acid (UDCA) (B), pioglitazone, liraglutide (C), and metformin and losartan (D). Numbers after terms in Venn graphs, including genes associated with the selected drugs, indicate the number of genes stored in the CTD database for a given term: UDCA,  $n = 299$ ; vitamin E,  $n = 174$ ; pioglitazone,  $n = 686$ ; liraglutide,  $n = 3$ ; losartan,  $n = 172$ ; and metformin,  $n = 424$ . CTD integrates information on chemicals, including chemical structures, curated interacting genes and proteins, curated and inferred disease relationships, and enriched pathways and functional annotations, which were extracted from the U.S. National Library of Medicine, the Online Mendelian Inheritance in Man (OMIM) database, and the gene database at the National Center for Biotechnology Information (NCBI). The interaction network was modeled by the CTD (<http://ctdbase.org>).

2,136 model organisms and 8,631 protein families.<sup>112</sup> Predictions are based on the premise that a protein is “druggable” if its activity can be modulated by its binding to a drug-like small compound.<sup>112</sup> The results yielded by this analysis revealed that neither PNPLA3 nor TM6SF2 have any “druggable” protein structure or are associated with any bioactive compound, or are potentially druggable by any predicted ligand-based approach. Assessment of the same parameters for GCKR and HSD17B13 shows a contrasting scenario, as both proteins are potentially druggable targets based on their molecular target three-dimensional structure (Protein Data Bank) and ligandability prediction, which were performed for all identified pockets within each protein structure. Based on the homology of closest druggable structure(s), which examines the structure of the protein and identifies any cavities on the protein surface where a drug-like compound could bind, we found that HSD17B13 and GCKR have a structural druggability of 66.67 and 100%,

respectively. Nevertheless, druggability prediction using different approaches, including tumor-tissue and cell line expression, and mutational analysis indicated that overall druggability percentile of GCKR is 44.08%, including druggability for cancer (46.36%) and other therapeutics (17.32%).

Specific focused analysis by the canSAR resource on candidate genes previously associated with NAFLD and the disease severity, for instance *STAT3*,<sup>113</sup> revealed that the protein coded by this gene presents a ligand-based druggability score of 97%. This specific score indicates the likely druggability of the protein based on the chemical properties of different compounds tested against the protein itself and/or its homologs. *STAT3* protein presents an overall druggability percentile of 99.21%, and druggability for cancer therapeutics of 99.39%. Furthermore, structural druggability of *STAT3* is 100%.

Network-based druggability assessment for *STAT3* and GCKR proteins, which examines the structure or the protein–



**Fig. 4** Signal transducer and activator of transcription 3 (*STAT3*) and glucokinase regulator gene (*GCKR*) radar network-predicted druggability plots. Radar plots showing representative network property profiles of *STAT3* and *GCKR* as potential drug targets (blue plot). The predicted network druggability is compared with the randomized network model of an average cancer target (green plot) or an average target for a noncancer drug. Prediction was performed by the canSAR resource available at <https://cansar.icr.ac.uk>. The network descriptors are divided into three categories: Substructures, Topological, and Community-based. The substructures were obtained from Przulj.<sup>119</sup> The graphlets are labeled as G-n and the orbits as O-. Topological descriptors: Betweenness centrality: A measure for quantifying the influence of one protein on the communication between other proteins in a network. Closeness centrality: Measures how many steps are required for a protein to reach every other protein—a lower number of steps indicates faster communication. Burt's constraint: Burt's Structural Hole and Ego Networks. Constraint is higher when a protein's neighbors are also connected, making the protein more redundant. k-core: A k-core is a fully connected subgraph in which each protein has a degree of at least k. Kleinberg hub score: A measure of how authoritative each protein is based on the principal eigenvector of the network's adjacency matrix. Google PageRank: A measure of the relative importance of the protein within the network; as the protein-protein network is an undirected graph, PageRank is positively correlated with degree distribution. Clustering coefficient: The probability that the neighbors are also connected to each other, calculated by the ratio of triangles connected to the protein. Community-based descriptors: Community Size (Walktrap): Based on hierarchical clustering and attempts to find densely connected subgraphs via random walks across the network. Community size (Spinglass): Based on partitioning clustering, where the number of communities to detect is predefined. Intra-community: Ratio of inter- to intracommunity communication. A higher number indicates that the protein's neighbors are in the same community. Spinglass inner: Number of interactions within the community. Spinglass outer: Number of interactions between the community and the rest of the network.

protein interaction around the target, suggests that *STAT3* but not *GCKR* is a good drug target, as disrupting its activity would affect different and relevant cellular processes (► **Fig. 4**). In fact, *STAT3* performs better than average targets of other therapeutic areas, even cancer (► **Fig. 4**).

A recently published experimental study in mice in which the researchers used a novel small *STAT3* inhibitor molecule (C188-9) has demonstrated its beneficial effects on liver-related outcomes.<sup>114</sup> C188-9 not only reduced tumor development but also improved liver steatosis, inflammation, and pathological lesions of NASH in mice with hepatocyte-specific deletion of *Pten* gene.<sup>114</sup> Further experimental and clinical evidence indicates that *STAT3* is not only involved in the regulatory circuit of liver fibrogenesis (► **Fig. 3**),<sup>115</sup> but is also involved in NASH by exacerbating insulin resistance.<sup>116</sup>

In conclusion, as genetic studies of NAFLD and NASH continue to expand, they are likely to provide insights into the mechanisms of disease pathogenesis and progression.

Knowledge on variants associated with the susceptibility of NASH offers an interesting opportunity not only for individualized risk prediction and prognosis, but also for the individual assessment of therapeutic response. Hence, future medicine in the field of NASH would benefit from patient-optimized strategies, which rather than being implemented on a wide scale may be tailored to the genetic makeup of each patient.

## Main Concepts and Learning Points

### 1. Concepts

- NAFLD is a polygenic complex disease
- NAFLD gene-regulatory networks
- Shared pathogenic mechanisms of chronic liver damage
- NAFLD genes: pleiotropy or just biologically meaningful associations?
- NAFLD and the druggable proteome

## 2. Learning Points

- The genetic component of NAFLD and NASH is largely explained by variants in genes that regulate glucose and fat homeostasis.
- Integrated pathways of disease pathogenesis suggest > 100-fold change enrichment in adiponectin and STAT3 activated-signaling pathways, retinol O-fatty-acyltransferase, and  $\beta$ -adrenergic activity.
- Convergent pathophenotypes, including liver inflammation and fibrosis, share molecular regulatory pathways and disease-predisposing genes.
- NAFLD-associated genes overlap with loci that were originally thought to play a role in the metabolic syndrome-associated traits.
- Data generated from candidate-gene association studies and genome-wide surveys can be leveraged to identify therapeutic targets.

### Note

The authors apologize to the colleagues whose works could not be cited owing to manuscript length limitations.

### Funding

Agencia Nacional de Promoción Científica y Tecnológica, Fondo para la Investigación Científica y Tecnológica (*Fon-CyT*) (PICT 2014–0432 and PICT 2015–0551 to S.S. and PICT 2014–1816 and PICT 2016–0135 to C.J.P.).

### Conflicts of Interest

None.

## References

- 1 Brunt EM, Wong VW, Nobili V, et al. Nonalcoholic fatty liver disease. *Nat Rev Dis Primers* 2015;1:15080
- 2 Palmer ND, Musani SK, Yerges-Armstrong LM, et al. Characterization of European ancestry nonalcoholic fatty liver disease-associated variants in individuals of African and Hispanic descent. *Hepatology* 2013;58(03):966–975
- 3 Speliotes EK, Yerges-Armstrong LM, Wu J, et al; NASH CRN; GIANT Consortium; MAGIC Investigators; GOLD Consortium. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. *PLoS Genet* 2011;7(03):e1001324
- 4 Schwimmer JB, Celedon MA, Lavine JE, et al. Heritability of nonalcoholic fatty liver disease. *Gastroenterology* 2009;136(05):1585–1592
- 5 Wagenknecht LE, Scherzinger AL, Stamm ER, et al. Correlates and heritability of nonalcoholic fatty liver disease in a minority cohort. *Obesity (Silver Spring)* 2009;17(06):1240–1246
- 6 Loomba R, Schork N, Chen CH, et al; Genetics of NAFLD in Twins Consortium. Heritability of hepatic fibrosis and steatosis based on a prospective twin study. *Gastroenterology* 2015;149(07):1784–1793
- 7 Tarnoki AD, Tarnoki DL, Bata P, et al. Heritability of non-alcoholic fatty liver disease and association with abnormal vascular parameters: a twin study. *Liver Int* 2012;32(08):1287–1293
- 8 Sookoian S, Pirola CJ. The genetic epidemiology of nonalcoholic fatty liver disease: toward a personalized medicine. *Clin Liver Dis* 2012;16(03):467–485
- 9 Hernaez R. Genetic factors associated with the presence and progression of nonalcoholic fatty liver disease: a narrative review. *Gastroenterol Hepatol* 2012;35(01):32–41
- 10 Sookoian S, Pirola CJ. Genetic predisposition in nonalcoholic fatty liver disease. *Clin Mol Hepatol* 2017;23(01):1–12
- 11 Romeo S, Kozlitina J, Xing C, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* 2008;40(12):1461–1465
- 12 Kozlitina J, Smagris E, Stender S, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. *Nat Genet* 2014;46(04):352–356
- 13 Sookoian S, Pirola CJ. Meta-analysis of the influence of I148M variant of patatin-like phospholipase domain containing 3 gene (PNPLA3) on the susceptibility and histological severity of non-alcoholic fatty liver disease. *Hepatology* 2011;53(06):1883–1894
- 14 Pirola CJ, Sookoian S. The dual and opposite role of the TM6SF2-rs58542926 variant in protecting against cardiovascular disease and conferring risk for nonalcoholic fatty liver: a meta-analysis. *Hepatology* 2015;62(06):1742–1756
- 15 Zain SM, Mohamed Z, Mohamed R. Common variant in the glucokinase regulatory gene rs780094 and risk of nonalcoholic fatty liver disease: a meta-analysis. *J Gastroenterol Hepatol* 2015;30(01):21–27
- 16 Sookoian S, Castaño GO, Burgueño AL, Gianotti TF, Rosselli MS, Pirola CJ. A nonsynonymous gene variant in the adiponutrin gene is associated with nonalcoholic fatty liver disease severity. *J Lipid Res* 2009;50(10):2111–2116
- 17 Liu YL, Reeves HL, Burt AD, et al. TM6SF2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. *Nat Commun* 2014;5:4309
- 18 Dongiovanni P, Petta S, Maglio C, et al. Transmembrane 6 superfamily member 2 gene variant disentangles nonalcoholic steatohepatitis from cardiovascular disease. *Hepatology* 2015;61(02):506–514
- 19 Mancina RM, Dongiovanni P, Petta S, et al. The MBOAT7-TMC4 variant rs641738 increases risk of nonalcoholic fatty liver disease in individuals of European descent. *Gastroenterology* 2016;150(05):1219–1230.e6
- 20 Kawaguchi T, Shima T, Mizuno M, et al. Risk estimation model for nonalcoholic fatty liver disease in the Japanese using multiple genetic markers. *PLoS One* 2018;13(01):e0185490
- 21 Koo BK, Joo SK, Kim D, et al. Additive effects of PNPLA3 and TM6SF2 on the histological severity of non-alcoholic fatty liver disease. *J Gastroenterol Hepatol* 2018;33(06):1277–1285
- 22 Sookoian S, Flichman D, Garaycochea ME, et al. Lack of evidence supporting a role of TMC4-rs641738 missense variant-MBOAT7-intergenic downstream variant-in the susceptibility to nonalcoholic fatty liver disease. *Sci Rep* 2018;8(01):5097
- 23 Abul-Husn NS, Cheng X, Li AH, et al. A protein-truncating HSD17B13 variant and protection from chronic liver disease. *N Engl J Med* 2018;378(12):1096–1106
- 24 Sookoian S, Flichman D, Scian R, et al. Mitochondrial genome architecture in non-alcoholic fatty liver disease. *J Pathol* 2016;240(04):437–449
- 25 Sookoian S, Rosselli MS, Gemma C, et al. Epigenetic regulation of insulin resistance in nonalcoholic fatty liver disease: impact of liver methylation of the peroxisome proliferator-activated receptor  $\gamma$  coactivator 1 $\alpha$  promoter. *Hepatology* 2010;52(06):1992–2000
- 26 Sunny NE, Parks EJ, Browning JD, Burgess SC. Excessive hepatic mitochondrial TCA cycle and gluconeogenesis in humans with nonalcoholic fatty liver disease. *Cell Metab* 2011;14(06):804–810
- 27 Koliaki C, Szendroedi J, Kaul K, et al. Adaptation of hepatic mitochondrial function in humans with non-alcoholic fatty liver is lost in steatohepatitis. *Cell Metab* 2015;21(05):739–746
- 28 Pirola CJ, Fernández Gianotti T, Castaño GO, et al. Circulating microRNA signature in non-alcoholic fatty liver disease: from

- serum non-coding RNAs to liver histology and disease pathogenesis. *Gut* 2015;64(05):800–812
- 29 Sookoian S, Rohr C, Salatino A, et al. Genetic variation in long noncoding RNAs and the risk of nonalcoholic fatty liver disease. *Oncotarget* 2017;8(14):22917–22926
  - 30 Chalasani N, Guo X, Loomba R, et al; Nonalcoholic Steatohepatitis Clinical Research Network. Genome-wide association study identifies variants associated with histologic features of non-alcoholic fatty liver disease. *Gastroenterology* 2010;139(05):1567–1576, 1576.e1–1576.e6
  - 31 Stender S, Kozlitina J, Nordestgaard BG, Tybjærg-Hansen A, Hobbs HH, Cohen JC. Adiposity amplifies the genetic risk of fatty liver disease conferred by multiple loci. *Nat Genet* 2017;49(06):842–847
  - 32 Boonvisit S, Yoshida K, Nakayama K, Watanabe K, Miyashita H, Iwamoto S. Identification of deleterious rare variants in MTP, PNPLA3, and TM6SF2 in Japanese males and association studies with NAFLD. *Lipids Health Dis* 2017;16(01):183
  - 33 Pirola CJ, Flichman D, Dopazo H, et al. A rare nonsense mutation in the glucokinase regulator gene is associated with a rapidly progressive clinical form of nonalcoholic steatohepatitis. *Hepatol Commun* 2018;2(09):1030–1036
  - 34 Ahrens M, Ammerpohl O, von Schönfels W, et al. DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. *Cell Metab* 2013;18(02):296–302
  - 35 Kitamoto T, Kitamoto A, Ogawa Y, et al. Targeted-bisulfite sequence analysis of the methylation of CpG islands in genes encoding PNPLA3, SAMM50, and PARVB of patients with non-alcoholic fatty liver disease. *J Hepatol* 2015;63(02):494–502
  - 36 Murphy SK, Yang H, Moylan CA, et al. Relationship between methylome and transcriptome in patients with nonalcoholic fatty liver disease. *Gastroenterology* 2013;145(05):1076–1087
  - 37 Pirola CJ, Scian R, Gianotti TF, et al. Epigenetic modifications in the biology of nonalcoholic fatty liver disease: the role of DNA hydroxymethylation and TET proteins. *Medicine (Baltimore)* 2015;94(36):e1480
  - 38 Zeybel M, Hardy T, Robinson SM, et al. Differential DNA methylation of genes involved in fibrosis progression in non-alcoholic fatty liver disease and alcoholic liver disease. *Clin Epigenetics* 2015;7:25
  - 39 Pirola CJ, Gianotti TF, Burgueño AL, et al. Epigenetic modification of liver mitochondrial DNA is associated with histological severity of nonalcoholic fatty liver disease. *Gut* 2013;62(09):1356–1363
  - 40 Mitsche MA, Hobbs HH, Cohen JC. Patatin-like phospholipase domain-containing protein 3 promotes transfer of essential fatty acids from triglycerides to phospholipids in hepatic lipid droplets. *J Biol Chem* 2018;293(18):6958–6968
  - 41 BasuRay S, Smagris E, Cohen JC, Hobbs HH. The PNPLA3 variant associated with fatty liver disease (I148M) accumulates on lipid droplets by evading ubiquitylation. *Hepatology* 2017;66(04):1111–1124
  - 42 Smagris E, BasuRay S, Li J, et al. Pnpla3<sup>I148M</sup> knockin mice accumulate PNPLA3 on lipid droplets and develop hepatic steatosis. *Hepatology* 2015;61(01):108–118
  - 43 Li JZ, Huang Y, Karaman R, et al. Chronic overexpression of PNPLA3<sup>I148M</sup> in mouse liver causes hepatic steatosis. *J Clin Invest* 2012;122(11):4130–4144
  - 44 Mahdessian H, Taxiarchis A, Popov S, et al. TM6SF2 is a regulator of liver fat metabolism influencing triglyceride secretion and hepatic lipid droplet content. *Proc Natl Acad Sci U S A* 2014;111(24):8913–8918
  - 45 Smagris E, Gilyard S, BasuRay S, Cohen JC, Hobbs HH. Inactivation of Tm6sf2, a gene defective in fatty liver disease, impairs lipidation but not secretion of very low density lipoproteins. *J Biol Chem* 2016;291(20):10659–10676
  - 46 Anstee QM, Daly AK, Day CP. Genetics of alcoholic and nonalcoholic fatty liver disease. *Semin Liver Dis* 2011;31(02):128–146
  - 47 Hernaez R, Yeung E, Clark JM, Kowdley KV, Brancati FL, Kao WH. Hemochromatosis gene and nonalcoholic fatty liver disease: a systematic review and meta-analysis. *J Hepatol* 2011;55(05):1079–1085
  - 48 Kahali B, Halligan B, Speliotes EK. Insights from genome-wide association analyses of nonalcoholic fatty liver disease. *Semin Liver Dis* 2015;35(04):375–391
  - 49 Sookoian S, Pirola CJ. Systems biology elucidates common pathogenic mechanisms between nonalcoholic and alcoholic fatty liver disease. *PLoS One* 2013;8(03):e58895
  - 50 Gao B, Tsukamoto H. Inflammation in alcoholic and nonalcoholic fatty liver disease: friend or foe? *Gastroenterology* 2016;150(08):1704–1709
  - 51 Goossens N, Hoshida Y. Is hepatocellular cancer the same disease in alcoholic and nonalcoholic fatty liver diseases? *Gastroenterology* 2016;150(08):1710–1717
  - 52 Greuter T, Malhi H, Gores GJ, Shah VH. Therapeutic opportunities for alcoholic steatohepatitis and nonalcoholic steatohepatitis: exploiting similarities and differences in pathogenesis. *JCI Insight* 2017;2(17):2
  - 53 Sookoian S, Pirola CJ. Genetic determinants of acquired cholestasis: a systems biology approach. *Front Biosci (Landmark Ed)* 2012;17:206–220
  - 54 Hirschfeld GM, Chapman RW, Karlsen TH, Lammert F, Lazaridis KN, Mason AL. The genetics of complex cholestatic disorders. *Gastroenterology* 2013;144(07):1357–1374
  - 55 Reichert MC, Hall RA, Krawczyk M, Lammert F. Genetic determinants of cholangiopathies: molecular and systems genetics. *Biochim Biophys Acta Mol Basis Dis* 2018;1864(4 Pt B):1484–1490
  - 56 Sookoian S, Flichman D, Garaycochea ME, San Martino J, Castaño GO, Pirola CJ. Metastasis-associated lung adenocarcinoma transcript 1 as a common molecular driver in the pathogenesis of nonalcoholic steatohepatitis and chronic immune-mediated liver damage. *Hepatol Commun* 2018;2(06):654–665
  - 57 Stickel F, Buch S, Lau K, et al. Genetic variation in the PNPLA3 gene is associated with alcoholic liver injury in Caucasians. *Hepatology* 2011;53(01):86–95
  - 58 Chamorro AJ, Torres JL, Mirón-Canelo JA, González-Sarmiento R, Laso FJ, Marcos M. Systematic review with meta-analysis: the I148M variant of patatin-like phospholipase domain-containing 3 gene (PNPLA3) is significantly associated with alcoholic liver cirrhosis. *Aliment Pharmacol Ther* 2014;40(06):571–581
  - 59 Buch S, Stickel F, Trépo E, et al. A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. *Nat Genet* 2015;47(12):1443–1448
  - 60 Trepo E, Guyot E, Ganne-Carré N, et al. PNPLA3(rs738409C>G) is a common risk variant associated with hepatocellular carcinoma in alcoholic cirrhosis. *Hepatology* 2012;55(04):1307–1308
  - 61 Stickel F, Buch S, Nischalke HD, et al. Genetic variants in PNPLA3 and TM6SF2 predispose to the development of hepatocellular carcinoma in individuals with alcohol-related cirrhosis. *Am J Gastroenterol* 2018;113(07):1099
  - 62 Salameh H, Raff E, Erwin A, et al. PNPLA3 gene polymorphism is associated with predisposition to and severity of alcoholic liver disease. *Am J Gastroenterol* 2015;110(06):846–856
  - 63 Atkinson SR, Way MJ, McQuillan A, Morgan MY, Thursz MR. Homozygosity for rs738409:G in PNPLA3 is associated with increased mortality following an episode of severe alcoholic hepatitis. *J Hepatol* 2017;67(01):120–127
  - 64 Salameh H, Hanayneh MA, Masadeh M, et al. PNPLA3 as a genetic determinant of risk for and severity of non-alcoholic fatty liver disease spectrum. *J Clin Transl Hepatol* 2016;4(03):175–191
  - 65 Bruschi FV, Claudel T, Tardelli M, et al. The PNPLA3 I148M variant modulates the fibrogenic phenotype of human hepatic stellate cells. *Hepatology* 2017;65(06):1875–1890

- 66 Jenkins CM, Mancuso DJ, Yan W, Sims HF, Gibson B, Gross RW. Identification, cloning, expression, and purification of three novel human calcium-independent phospholipase A2 family members possessing triacylglycerol lipase and acylglycerol transacylase activities. *J Biol Chem* 2004;279(47):48968–48975
- 67 Lake AC, Sun Y, Li JL, et al. Expression, regulation, and triglyceride hydrolase activity of Adiponutrin family members. *J Lipid Res* 2005;46(11):2477–2487
- 68 Pingitore P, Dongiovanni P, Motta BM, et al. PNPLA3 overexpression results in reduction of proteins predisposing to fibrosis. *Hum Mol Genet* 2016;25(23):5212–5222
- 69 Min HK, Sookoian S, Pirola CJ, Cheng J, Mirshahi F, Sanyal AJ. Metabolic profiling reveals that PNPLA3 induces widespread effects on metabolism beyond triacylglycerol remodeling in Huh-7 hepatoma cells. *Am J Physiol Gastrointest Liver Physiol* 2014;307(01):G66–G76
- 70 Sookoian S, Castaño GO, Pirola CJ. PNPLA3 I148M variant is associated with metabolic stress-response phenotype in patients with nonalcoholic fatty liver disease. *Hepatology* 2015;61(05):1769
- 71 Pirola CJ, Garaycoechea M, Flichman D, et al. Splice variant rs72613567 prevents worst histologic outcomes in patients with nonalcoholic fatty liver disease. *J Lipid Res* 2019;60(01):176–185
- 72 Ma Y, Belyaeva OV, Brown PM, et al. HSD17B13 is a hepatic retinol dehydrogenase associated with histological features of non-alcoholic fatty liver disease. *Hepatology* 2018. doi: 10.1002/hep.30350. [Epub ahead of print]
- 73 Chambers JC, Zhang W, Sehmi J, et al; Alcohol Genome-wide Association (AlcGen) Consortium; Diabetes Genetics Replication and Meta-analyses (DIAGRAM+) Study; Genetic Investigation of Anthropometric Traits (GIANT) Consortium; Global Lipids Genetics Consortium; Genetics of Liver Disease (GOLD) Consortium; International Consortium for Blood Pressure (ICBP-GWAS); Meta-analyses of Glucose and Insulin-Related Traits Consortium (MAGIC). Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet* 2011;43(11):1131–1138
- 74 Sookoian S, Castaño GO, Burgueño AL, Gianotti TF, Rosselli MS, Pirola CJ. The nuclear receptor PXR gene variants are associated with liver injury in nonalcoholic fatty liver disease. *Pharmacogenet Genomics* 2010;20(01):1–8
- 75 Wang YM, Chai SC, Brewer CT, Chen T. Pregnane X receptor and drug-induced liver injury. *Expert Opin Drug Metab Toxicol* 2014;10(11):1521–1532
- 76 Feder JN, Gnirke A, Thomas W, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet* 1996;13(04):399–408
- 77 Nahon P, Sutton A, Rufat P, et al. Liver iron, HFE gene mutations, and hepatocellular carcinoma occurrence in patients with cirrhosis. *Gastroenterology* 2008;134(01):102–110
- 78 George DK, Goldwurm S, MacDonald GA, et al. Increased hepatic iron concentration in nonalcoholic steatohepatitis is associated with increased fibrosis. *Gastroenterology* 1998;114(02):311–318
- 79 Nelson JE, Bhattacharya R, Lindor KD, et al. HFE C282Y mutations are associated with advanced hepatic fibrosis in Caucasians with nonalcoholic steatohepatitis. *Hepatology* 2007;46(03):723–729
- 80 Chappell S, Hadzic N, Stockley R, Guetta-Baranes T, Morgan K, Kalsheker N. A polymorphism of the alpha1-antitrypsin gene represents a risk factor for liver disease. *Hepatology* 2008;47(01):127–132
- 81 Strnad P, Buch S, Hamesch K, et al. Heterozygous carriage of the alpha1-antitrypsin Pi\*Z variant increases the risk to develop liver cirrhosis. *Gut* 2018;gutjnl-2018-316228
- 82 Stearns FW. One hundred years of pleiotropy: a retrospective. *Genetics* 2010;186(03):767–773
- 83 Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet* 2016;48(07):709–717
- 84 Sivakumaran S, Agakov F, Theodoratou E, et al. Abundant pleiotropy in human complex diseases and traits. *Am J Hum Genet* 2011;89(05):607–618
- 85 Chesmore K, Bartlett J, Williams SM. The ubiquity of pleiotropy in human disease. *Hum Genet* 2018;137(01):39–44
- 86 Sookoian S, Pirola CJ. Nonalcoholic fatty liver disease and metabolic syndrome: Shared genetic basis of pathogenesis. *Hepatology* 2016;64(05):1417–1420
- 87 Iwata M, Maeda S, Kamura Y, et al. Genetic risk score constructed using 14 susceptibility alleles for type 2 diabetes is associated with the early onset of diabetes and may predict the future requirement of insulin injections among Japanese individuals. *Diabetes Care* 2012;35(08):1763–1770
- 88 Holmen OL, Zhang H, Fan Y, et al. Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nat Genet* 2014;46(04):345–351
- 89 Sookoian S, Castaño GO, Scian R, et al. Genetic variation in transmembrane 6 superfamily member 2 and the risk of non-alcoholic fatty liver disease and histological disease severity. *Hepatology* 2015;61(02):515–525
- 90 Andrade-Navarro MA, Fontaine JF. Gene Set to Diseases (GS2D): disease enrichment analysis on human gene sets with literature data. *Genom Comput Biol* 2016;2(01):e33
- 91 Rowe GC, Jiang A, Arany Z. PGC-1 coactivators in cardiac development and disease. *Circ Res* 2010;107(07):825–838
- 92 Lai CQ, Tucker KL, Parnell LD, et al. PPARGC1A variation associated with DNA damage, diabetes, and cardiovascular diseases: the Boston Puerto Rican Health Study. *Diabetes* 2008;57(04):809–816
- 93 Turek FW, Joshi C, Kohsaka A, et al. Obesity and metabolic syndrome in circadian Clock mutant mice. *Science* 2005;308(5724):1043–1045
- 94 Scott EM, Carter AM, Grant PJ. Association between polymorphisms in the Clock gene, obesity and the metabolic syndrome in man. *Int J Obes* 2008;32(04):658–662
- 95 Sookoian S, Gemma C, Gianotti TF, Burgueño A, Castaño G, Pirola CJ. Genetic variants of Clock transcription factor are associated with individual susceptibility to obesity. *Am J Clin Nutr* 2008;87(06):1606–1615
- 96 Sookoian S, Castaño G, Gemma C, Gianotti TF, Pirola CJ. Common genetic variations in CLOCK transcription factor are associated with nonalcoholic fatty liver disease. *World J Gastroenterol* 2007;13(31):4242–4248
- 97 Kotronen A, Peltonen M, Hakkarainen A, et al. Prediction of non-alcoholic fatty liver disease and liver fat using metabolic and genetic factors. *Gastroenterology* 2009;137(03):865–872
- 98 Wood GC, Chu X, Argyropoulos G, et al. A multi-component classifier for nonalcoholic fatty liver disease (NAFLD) based on genomic, proteomic, and phenomic data domains. *Sci Rep* 2017;7:43238
- 99 Zhou Y, Orešič M, Leivonen M, et al. Noninvasive detection of nonalcoholic steatohepatitis using clinical markers and circulating levels of lipids and metabolites. *Clin Gastroenterol Hepatol* 2016;14(10):1463–1472.e6
- 100 Nobili V, Donati B, Panera N, et al. A 4-polymorphism risk score predicts steatohepatitis in children with nonalcoholic fatty liver disease. *J Pediatr Gastroenterol Nutr* 2014;58(05):632–636
- 101 Sevastianova K, Kotronen A, Gastaldelli A, et al. Genetic variation in PNPLA3 (adiponutrin) confers sensitivity to weight loss-induced decrease in liver fat in humans. *Am J Clin Nutr* 2011;94(01):104–111
- 102 Shen J, Wong GL, Chan HL, et al. PNPLA3 gene polymorphism and response to lifestyle modification in patients with nonalcoholic fatty liver disease. *J Gastroenterol Hepatol* 2015;30(01):139–146



- 103 Krawczyk M, Jiménez-Agüero R, Alustiza JM, et al. PNPLA3 p. I148M variant is associated with greater reduction of liver fat content after bariatric surgery. *Surg Obes Relat Dis* 2016;12(10):1838–1846
- 104 Finkenstedt A, Auer C, Glodny B, et al. Patatin-like phospholipase domain-containing protein 3 rs738409-G in recipients of liver transplants is a risk factor for graft steatosis. *Clin Gastroenterol Hepatol* 2013;11(12):1667–1672
- 105 Kim H, Lee KW, Lee K, et al. Effect of PNPLA3 I148M polymorphism on histologically proven non-alcoholic fatty liver disease in liver transplant recipients. *Hepatol Res* 2018;48(03):E162–E171
- 106 Sookoian S, Arrese M, Pirola CJ. Genetics meets therapy? Exome-wide association study reveals a loss-of-function variant in HSD17B13 (17-beta-hydroxysteroid dehydrogenase 13) that protects patients from liver damage and NAFLD-progression. *Hepatology* 2019;69(02):907–910
- 107 Konerman MA, Jones JC, Harrison SA. Pharmacotherapy for NASH: current and emerging. *J Hepatol* 2018;68(02):362–375
- 108 Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. *Hepatology* 2018;67(01):328–357
- 109 Sanyal AJ, Chalasani N, Kowdley KV, et al; NASH CRN. Pioglitazone, vitamin E, or placebo for nonalcoholic steatohepatitis. *N Engl J Med* 2010;362(18):1675–1685
- 110 Ratziu V, Harrison SA, Francque S, et al; GOLDEN-505 Investigator Study Group. Elafibranor, an agonist of the peroxisome proliferator-activated receptor- $\alpha$  and - $\delta$ , induces resolution of nonalcoholic steatohepatitis without fibrosis worsening. *Gastroenterology* 2016;150(05):1147–1159.e5
- 111 Sookoian S, Pirola CJ. Elafibranor for the treatment of NAFLD: one pill, two molecular targets and multiple effects in a complex phenotype. *Ann Hepatol* 2016;15(04):604–609
- 112 Bulusu KC, Tym JE, Coker EA, Schierz AC, Al-Lazikani B. canSAR: updated cancer research and drug discovery knowledgebase. *Nucleic Acids Res* 2014;42(Database issue):D1040–D1047
- 113 Sookoian S, Castaño G, Gianotti TF, Gemma C, Rosselli MS, Pirola CJ. Genetic variants in STAT3 are associated with nonalcoholic fatty liver disease. *Cytokine* 2008;44(01):201–206
- 114 Jung KH, Yoo W, Stevenson HL, et al. Multifunctional effects of a small-molecule STAT3 inhibitor on NASH and hepatocellular carcinoma in mice. *Clin Cancer Res* 2017;23(18):5537–5546
- 115 Xiang DM, Sun W, Ning BF, et al. The HLF/IL-6/STAT3 feedforward circuit drives hepatic stellate cell activation to promote liver fibrosis. *Gut* 2018;67(09):1704–1715
- 116 Min HK, Mirshahi F, Verdianelli A, et al. Activation of the GP130-STAT3 axis and its potential implications in nonalcoholic fatty liver disease. *Am J Physiol Gastrointest Liver Physiol* 2015;308(09):G794–G803
- 117 Caussy C, Hsu C, Lo MT, et al; Genetics of NAFLD in Twins Consortium. Link between gut-microbiome derived metabolite and shared gene-effects with hepatic steatosis and fibrosis in NAFLD. *Hepatology* 2018. Doi: 10.1002/hep.29892
- 118 Puri P, Sanyal AJ. The intestinal microbiome in nonalcoholic fatty liver disease. *Clin Liver Dis* 2018;22(01):121–132
- 119 Przulj N. Biological network comparison using graphlet degree distribution. *Bioinformatics* 2007;23(02):e177–e183
- 120 Mi H, Huang X, Muruganujan A, et al. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* 2017;45(D1):D183–D189