CLINICAL and MOLECULAR HEPATOLOGY

Review

Precision medicine in nonalcoholic fatty liver disease: New therapeutic insights from genetics and systems biology

Silvia Sookoian[1,2] and Carlos J. Pirola[1,3]

[1]Institute of Medical Research A Lanari, School of Medicine, University of Buenos Aires, Autonomous City of Buenos Aires; Departments of [2]Clinical and Molecular Hepatology and [3]Molecular Genetics and Biology of Complex Diseases, Institute of Medical Research (IDIM), National Scientific and Technical Research Council (CONICET)-University of Buenos Aires, Autonomous City of Buenos Aires, Argentina

Despite more than two decades of extensive research focusing on nonalcoholic fatty liver disease (NAFLD), no approved therapy for steatohepatitis—the severe histological form of the disease—presently exists. More importantly, new drugs and small molecules with diverse molecular targets on the pathways of hepatocyte injury, inflammation, and fibrosis cannot achieve the primary efficacy endpoints. Precision medicine can potentially overcome this issue, as it is founded on extensive knowledge of the druggable genome/proteome. Hence, this review summarizes significant trends and developments in precision medicine with a particular focus on new potential therapeutic discoveries modeled via systems biology approaches. In addition, we computed and simulated the potential utility of the NAFLD polygenic risk score, which could be conceptually very advantageous not only for early disease detection but also for implementing actionable measures. Incomplete knowledge of the druggable NAFLD genome severely impedes the drug discovery process and limits the likelihood of identifying robust and safe drug candidates. Thus, we close this article with some insights into emerging disciplines, such as chemical genetics, that may accelerate accurate identification of the druggable NAFLD genome/proteome. **Clin Mol Hepatol 2020 Sep 10. [Epub ahead of print]**

Keywords: Nonalcoholic fatty liver disease; Genetics; Steatohepatitis; Precision medicine; Complex traits

Corresponding author : Silvia Sookoian
Institute of Medical Research A Lanari, School of Medicine, University of Buenos Aires, Combatientes de Malvinas 3150, Autonomous City of Buenos Aires 1427, Argentina
Tel: +54-11-52873905, Fax: +54-11-52873905
E-mail: ssookoian@intramed.net
https://orcid.org/0000-0001-5929-5470

Carlos J. Pirola
Institute of Medical Research A Lanari, School of Medicine, University of Buenos Aires, Combatientes de Malvinas 3150, Autonomous City of Buenos Aires 1427, Argentina
Tel: +54-11-52873903
E-mail: pirola.carlos@conicet.gov.ar
https://orcid.org/0000-0001-8234-4058

INTRODUCTION

Nonalcoholic fatty liver disease (NAFLD) is a complex disorder that affects a large proportion of the world population of all ages.[1] The disease pathogenesis involves a myriad of factors, including genetic susceptibility and predisposing metabolic comorbidities, such as obesity and type 2 diabetes, as well as environmental exposure and lifestyle, which jointly shape the NAFLD epigenome.[2-4]

Yet, despite more than two decades of extensive research in the field of NAFLD, there is currently no approved therapy for nonalcoholic steatohepatitis (NASH)—the severe histological form of the disease. Moreover, none of the new drugs or small molecules with diverse molecular targets on the pathways of hepatocyte injury, inflammation, and fibrosis can achieve primary efficacy endpoints.[5-10] Diverse factors have been postulated to contribute to the low success rate in NAFLD/NASH drug discovery, including lack of robust animal models needed for preclinical studies, insufficient target engagement or target modulation by the novel drugs, absence or insufficient demonstration of a proof-of-concept in early trials, and/or high false discovery rate (FDR) in phase 2 trials.[11]

Likewise, it has been hypothesized that, as data on the candidate drugs are not only insufficient but are also not corroborated by genetic inactivation, pharmacological inhibition, antisense oligonucleotides, and/or small interfering RNAs, this poses an additional obstacle to achieve consistent and sustained effects on severe histological outcomes, including improvement in fibrosis scores.[11]

The aforementioned pitfalls could potentially be overcome through systems biology analysis, aiming to integrate knowledge of signaling pathways,[12] the genetic information of susceptibility genes,[4] and multiple tissue-specific OMICs-related experiments that include large-scale transcriptomic, proteomic, and metabolomic profiles,[13,14] and more recently metagenomics of the liver tissue.[15]

Furthermore, the approach founded on precision medicine is expected to enhance the effectiveness of novel therapies, including elucidation of predictors of drug response. Hence, this review summarizes significant trends and developments in precision medicine with focus on new potential therapeutic discoveries modeled by systems biology approaches.

THE PATH TOWARDS PRECISION MEDICINE: A SHORT CONCEPTUAL APPRAISAL

The ultimate goal of precision medicine is to develop precision treatment strategies that rely upon a holistic understanding of differences in genetic and underlying molecular pathogenic factors, as well as responses to environmental stressors, among patients. Figure 1 describes milestones in the path towards precision medicine, which include the integration of big data, comprising information from electronic health records of thousands of patients and machine learning strategies, such as artificial intelligence, to assist with the development of algorithms for combining and decoding such complex information. In addition, collections of biological samples in large biobanks linked to patient data increase the likelihood of finding robust disease pathogenesis signatures derived from OMICs state-of-the-art approaches. Knowledge integration and data modeling and analysis are vital processes at the interface with drug discovery. It should be emphasized that time and cost are presently the key limiting factors at these stages. However, as the technological advances progress further, it is expected that, in the near future, the gap between the discovery of potential drugs and clinical validation will be considerably narrowed. Still, even when complex algorithms aimed at multi-scale modeling of OMICs data succeed in identifying a potential drug target, only some of these medications will eventually be included into primary or secondary therapeutic protocols for treating the target disease. At this stage, at least three clinical scenarios will likely emerge. One possibility is that a putative drug "x" demonstrates not only a good safety profile but also succeeds in achieving the objective endpoints in a large group of patients (Fig. 1). Another possibility is that, in a small group of patients, "x" would be contraindicated for diverse reasons, including safety concerns in some vulnerable groups, such as children, pregnant women, or patients suffering from chronic kidney disease, advanced cardiovascular disease, etc. It is also possible that certain patients will require a different kind of drug because their underlying molecular and/or genetic profile does not align with the molecular profile of the drug "x" (Fig. 1).

In addition, pathway-derived drugs may emerge from drug repurposing. A comprehensive algorithm for drug repurposing or repositioning in NASH, which primarily relies on identifying and developing new uses for existing drugs, was recently published.[16] In fact, drugs which had an acceptable safety profile, but failed in achieving the expected response for some diseases, could be used to treat a different condition. It is also noteworthy that a large
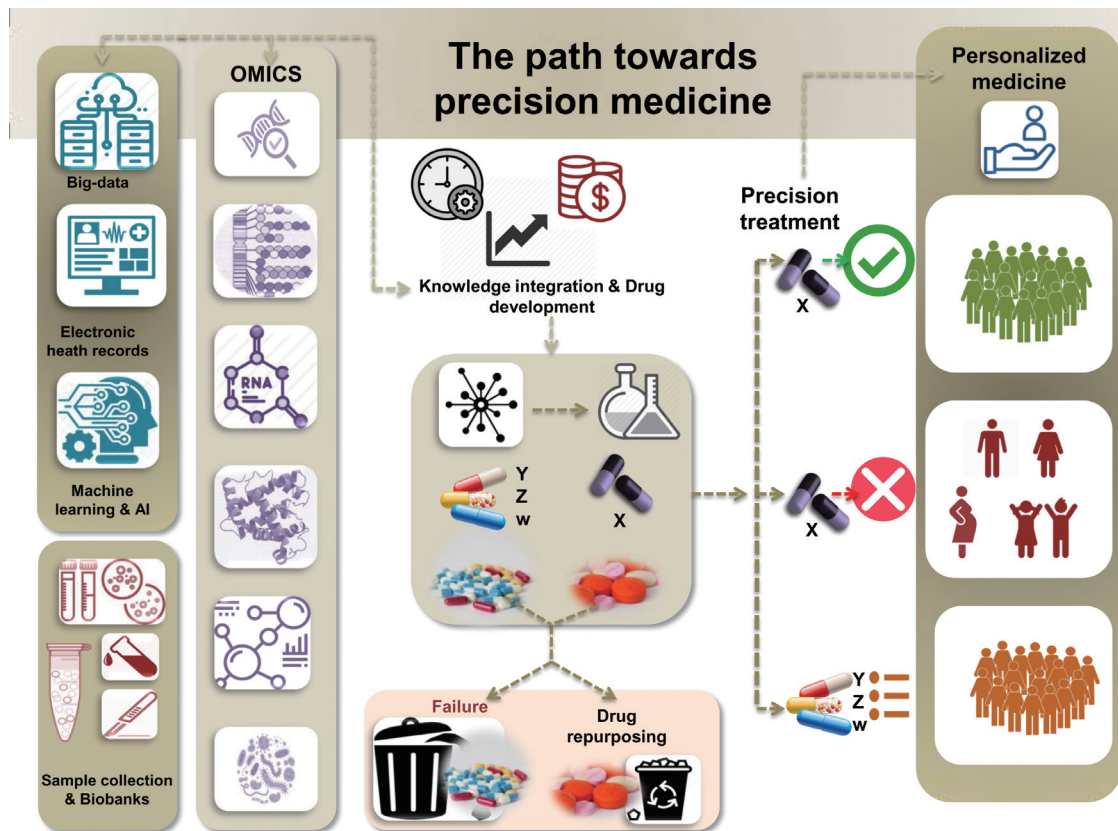
Figure 1. Milestones in the path towards precision medicine. The path involves the integration of knowledge derived from big data, electronic health records, large collections of biological samples in biobanks, and machine learning strategies that are linked to high-throughput OMICs experiments. Strategies pertaining to personalized medicine are also highlighted.

number of discovered drugs fail and never pass preclinical testing. Hence, the ultimate role of precision therapy relies on defining appropriate prediction strategies for implementing patient-based therapies.

PRECISION MEDICINE AND LARGE-SCALE GENOME AND EXOME SEQUENCING DATA

Extant studies on the genetic component of NAFLD indicate that, after 12 years following the discovery that an allele in a patatin-like phospholipase domain containing 3 (*PNPLA3*) variant (rs738409 [G], encoding 148M) was associated with increased hepatic fat[17] and NAFLD disease severity,[18] knowledge of the disease heritability is still incomplete.[3,4,12] The correlation between rs738409 and the risk of developing fatty liver, NASH, and fibrosis is perhaps one of the strongest worldwide-replicated effects for a common variant modifying the individual susceptibility of

NAFLD and NASH (explaining ~5.3% of the total variance).[3,4,18,19] Indeed, available evidence indicates that homozygous carriers with the G-risk allele of rs738409 present 3.24-fold greater risk of higher liver necroinflammatory scores and 3.2-fold greater risk of developing fibrosis when compared with homozygous CC carriers.[19]

Findings yielded by genome-wide association studies (GWAS) as a part of which the heritability of hepatic steatosis was explored at the population level and/or NASH was examined in patients with liver biopsy, consistently show that at least four loci (*PNPLA3*, transmembrane 6 superfamily member 2 [*TM6SF2*], glucokinase regulator [*GCKR*], and hydroxysteroid 17-beta dehydrogenase 13 [*HSD17B13*])[17,20-22] are involved in the genetic susceptibility of the disease (Fig. 2). Conflicting results have been published regarding the rs641738 C/T located in transmembrane channel-like 4 (*TMC4*) exon 1 (p.Gly17Glu) and 500 bases downstream of the membrane bound O-acyltransferase domain containing 7 (*MBOAT7*; *TMC4/MBOAT7*), which were initially de-
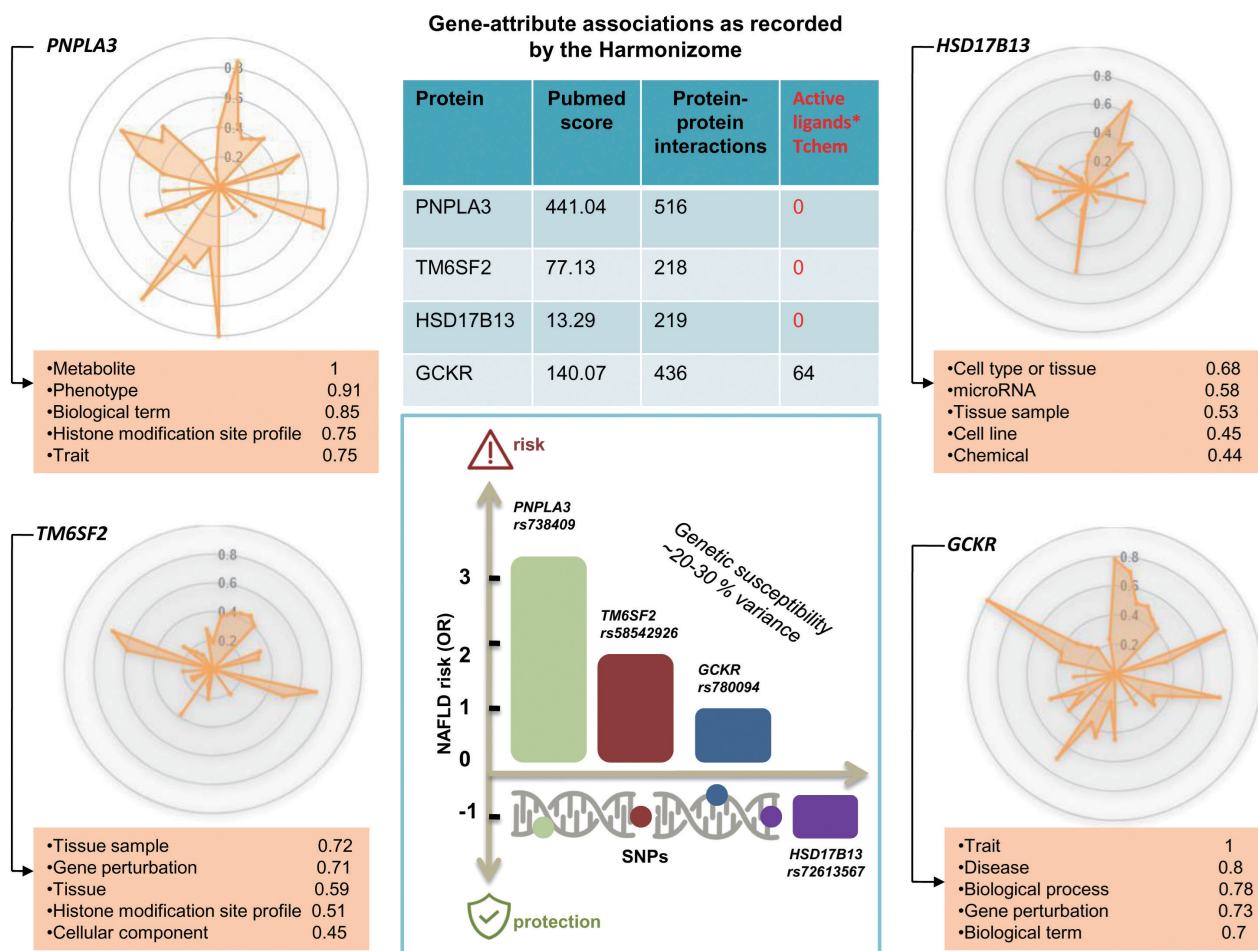
Figure 2. Illumination graph of major genetic modifiers of NAFLD and NASH. Radar plot and knowledge table depicting the variety of information obtained by Pharos (https://pharos.nih.gov/) for *PNPLA3*, *TM6SF2*, *GCKR*, and *HSD17B13*. These radial plots summarize the level of accumulated knowledge about each target. The greater the number of spikes in the plot, the greater the variety, with spike length indicating the quantity of that particular knowledge. The radar chart allows gene-attribute associations as recorded by the Harmonizome[57] to be visualized. The tables below the charts represent the top five knowledge attributes in the illumination graph. The knowledge value property is on a 0–1 scale. *PNPLA3*, patatin-like phospholipase domain containing 3; *TM6SF2*, transmembrane 6 superfamily member 2; *GCKR*, glucokinase regulator; *HSD17B13*, hydroxysteroid 17-beta dehydrogenase 13; NAFLD, nonalcoholic fatty liver disease; OR, odds ratio; SNPs, single nucleotide polymorphisms; NASH, nonalcoholic steatohepatitis.

scribed in Italian population[23] but could not be replicated in other populations around the world,[24-26] including a large cohort of European patients partaking in a GWAS, for whom NAFLD was diagnosed by liver biopsy.[27]

Targeting disease-associated genes has been proven successful in the treatment of numerous human diseases, cancers in particular. In this field, target drugs are associated with the effect of a mutant protein and/or are designed to interfere at the gene or protein level. However, for this goal to be fully realized, full knowledge of the target genes/proteins is required, not only at all molecular levels, but also at the level of chemical-gene/protein interactions, which would ultimately allow identification of potential

active ligands.

Likewise, drug discovery and precision medicine require complete understanding of new technologies as well as gene and protein biology of the selected target. Radar plots depicted in Figure 2 show the discrepancy in the level of knowledge on the four loci reproducibly implicated in the biology of NAFLD and the disease severity. For example, the plots show that, while extensive knowledge on *PNPLA3* at different levels already exists, ranging from metabolic aspects to epigenetic information and traits associations, that pertaining to *HSD17B13*—the newly discovered gene with a putative loss-of-function variant implicated in protective effects against NASH and severe histological stages[20,28]—is limited

at all levels of gene and protein biology (Fig. 2). The rs72613567 insertion/deletion variant, the functional consequence of which is a splice donor variant of the *HSD17B13*,[20] represents an interesting model of a candidate molecule for treating NASH and fibrosis.[29] In fact, although the information about druggable binding domains in *HSD17B13* is scarce due to the lack of an experimental 3D structure, other members of the protein family, such as HSD17B11 with high homology, present putative binding pockets for small molecules.[30]

As indicated in the *PNPLA3* illumination graph, considerable knowledge has been accumulated on the gene and protein, including gene-attributed associations, protein interactions, and high PubMed score; yet, no information on the active ligands of the protein is currently available (Fig. 2).

The protein encoded by *PNPLA3* is a triacylglycerol lipase that mediates triacylglycerol hydrolysis, mostly in adipocytes.[12] The encoded protein, which appears to be membrane-bound, may be involved in the energy usage/storage balance in adipocytes.[12] Figure 3A shows area under the receiver operating characteristics (AUROC; 0.951) for predicted functional processes linked to *PNPLA3*, in-

Predicted biological processes (GO)

	Gene set	Z-score
1	biotin metabolic process (GO:0006768)	7.93450450
2	low-density lipoprotein particle clearance (GO:0034383)	7.72633410
3	regulation of sequestering of triglyceride (GO:0010889)	7.51540267
4	fatty acid transmembrane transport (GO:1902001)	6.59916208
5	white fat cell differentiation (GO:0050872)	6.56176354
6	carnitine shuttle (GO:0006853)	6.33428032
7	brown fat cell differentiation (GO:0050873)	6.21320057
8	* triglyceride catabolic process (GO:0019433)	6.21181214
9	fatty-acyl-CoA biosynthetic process (GO:0046949)	5.95146673
10	* triglyceride biosynthetic process (GO:0019432)	5.80599408

PNPLA3 — AUC=0.951 (Sensitivity vs 1-specificity)

A

Predicted upstream transcription factors (ChEA)*

	Gene set	Z-score
1	TAF7L_23326641_ChIP-Seq_C3H10T1-2_Mouse	5.69863911
2	* RXR_22158963_ChIP-Seq_LIVER_Mouse	4.33561492
3	ESR1_17901129_ChIP-ChIP_LIVER_Mouse	4.22002803
4	PPARA_22158963_ChIP-Seq_LIVER_Mouse	3.56741020
5	* LXR_22158963_ChIP-Seq_LIVER_Mouse	3.48913801
6	GLI1_17442700_ChIP-ChIP_MESCs_Mouse	3.47600926
7	ZNF274_21170338_ChIP-Seq_K562_Hela	3.34868173
8	FOXO1_23066095_ChIP-Seq_LIVER_Mouse	3.09915464
9	* CLOCK_20551151_ChIP-Seq_293T_Human	3.01845043
10	GBX2_23144817_ChIP-Seq_PC3_Human	2.90979636

PNPLA3 — AUC=0.666 (Sensitivity vs 1-specificity)

B

Figure 3. *PNPLA3* predicted functional associations. Predicted biological processes (GO) (A) and upstream transcription factors (ChEA) (B) assessed by the Harmonizome (http://amp.pharm.mssm.edu/Harmonizome/gene/PNPLA3).[57] Tables show the top 10 predictions (the number provided in the first column). Table explanation: If a gene (gene set) shares high correlation with known members of a gene set, it is assigned a high z-score. Known functions/gene set associations are highlighted in green. AUROC is provided by the algorithm available in the ARCHS[4] (massive mining of publicly available RNA-seq data from human and mouse)[58] accessible at https://amp.pharm.mssm.edu/archs4/. Specifically, AUROC shows how well-known annotations are recovered by the ARCHS[4] algorithm. GO, gene ontology; *PNPLA3*, patatin-like phospholipase domain containing 3; AUROC, area under the receiver operating characteristics; ChEA, ChIP enrichment analysis. *From published ChIP-chip, ChIP-seq, and other transcription factor binding site profiling studies.[59]

cluding triglyceride catabolic process (gene ontology [GO]: 0019433) and triglyceride biosynthetic process (GO: 0019432) (prediction was done by the Harmonizome). Importantly, extensive mining of publicly available RNA-seq data acquired through human and mouse experiments (https://amp.pharm.mssm.edu/archs4/) uncovered interesting upstream transcription factors (AUROC, 0.666) in Figure 3B, including *RXR* (retinoid x receptor), *LXR* (liver x receptor, nuclear receptor subfamily 1 group H member 3), and *CLOCK* (circadian locomotor output cycles protein kaput, formally known as circadian clock regulator). The protein encoded by the *CLOCK* plays a central role in the regulation of circadian rhythms. In our previous publications, we reported for the first time that *CLOCK* genetic variation is associated with obesity and NAFLD.[31,32] The haplotype of rs1554483G and rs4864548A was found to be associated with a 1.8-fold risk of overweight status or obesity,[32] whereas rs1554483 was shown to be associated with all histological traits of NASH, including fibrosis.[31] Considered jointly, this body of evidence suggests that the putative cross-talk between *PNPLA3* and *CLOCK* could explain the link among NAFLD genetic susceptibility, the environment, and the circadian regulation of liver metabolism. However, further experiments are required to prove this hypothesis.

TM6SF2, of which rs58542926 C/T (E167K) was initially associated with liver fat accumulation and aminotransferase levels in a large GWAS study[21] and further replicated in subsequent studies,[21,33-35] encodes for a protein involved in lipid metabolism.[12] Diverse areas of molecular knowledge gained on this gene/protein are presented in Figure 2.

Radar plot of *GCKR* presented in Figure 2, of which rs780094 presents a very modest[36] effect (odds ratio [OR], 1.2) on NAFLD biology,[3,12] and shows that considerable gene/protein knowledge, including at least 64 putative ligands, has been already acquired in this domain.[4]

Precision medicine has emerged as a result of comprehensive knowledge of the druggable genome/proteome. Thus, advancing the chemical genetics research, which is based on the screening of low-molecular weight compounds that act by binding to specific receptors/proteins, is crucial to move this promising research domain forward in the right direction. It is worth mentioning that the incomplete knowledge on the druggable genome of NAFLD/NASH severely undermines the drug discovery progress and reduces the chances of having robust and safe drug candidates. Therefore, the substantial gap between the knowledge of NAFLD-predisposing genes and that related to putative protein ligands needs to be urgently addressed.

NAFLD AND THE PUTATIVE CLINICAL BENEFITS OF POLYGENIC RISK SCORES (PRSs)

Estimating the susceptibility risk of a given patient to develop a particular disease and/or to progress into severe disease stages is the ultimate aim of precision medicine. Most researchers concur that the PRS distribution, which is based on the sum of all independent risk single nucleotide polymorphisms (SNPs; ideally weighted by their size effects in a given population), could be approximated by the Gaussian (normal) curve (Fig. 4). PRSs are theoretically designed to explain the relative risk of a disease, as these scores provide information on how a person compares with others with different genetic susceptibility background. However, PRSs do not necessarily follow normal distribution, due to several factors, including differences in the population structure or admixture.

In the case of NAFLD and NASH, PRSs could be conceptually very advantageous not only for allowing early disease detection, but also for implementing timely actionable measures (Fig. 4). For example, invasive diagnostic approaches, such as liver biopsy, as well as early pharmacological intervention, would be advised for high-risk populations (those at the right-tail of the PRS distribution curve pertaining to the relevant population) whereas low-risk individuals (i.e., those on the left-tail of the curve) would be monitored until clinical risk becomes evident (Fig. 4). For those deemed at low or medium risk, which probably applies to the large majority of the affected patients, lifestyle changes would be advised, including regular physical activity and dietary modifications aimed at optimizing body weight and controlling the key metabolic risk factors (lipid traits and glucose metabolism).

Despite these benefits, several important concerns related to the clinical implementation of PRSs also exist, as noted in Figure 4. In particular, use of PRSs in clinical settings will remain impractical until the heritability of NAFLD is fully elucidated, and rare and familial forms of the disease are revealed. Inclusion of additional genetic information will certainly aid in overcoming these issues as the predictive power of PRSs improves and the proportion of individuals at risk diminishes.

For instance, early detection of a rare nonsense *GCKR* mutation (rs149847328, p.Arg227Ter) in a NAFLD patient with associated comorbidities, including morbid obesity and type 2 diabetes, when combined with prompt pharmacological intervention, could potentially prevent or even reverse the disease progression into liver cirrhosis.[37]

Furthermore, the currently available knowledge on the repro-
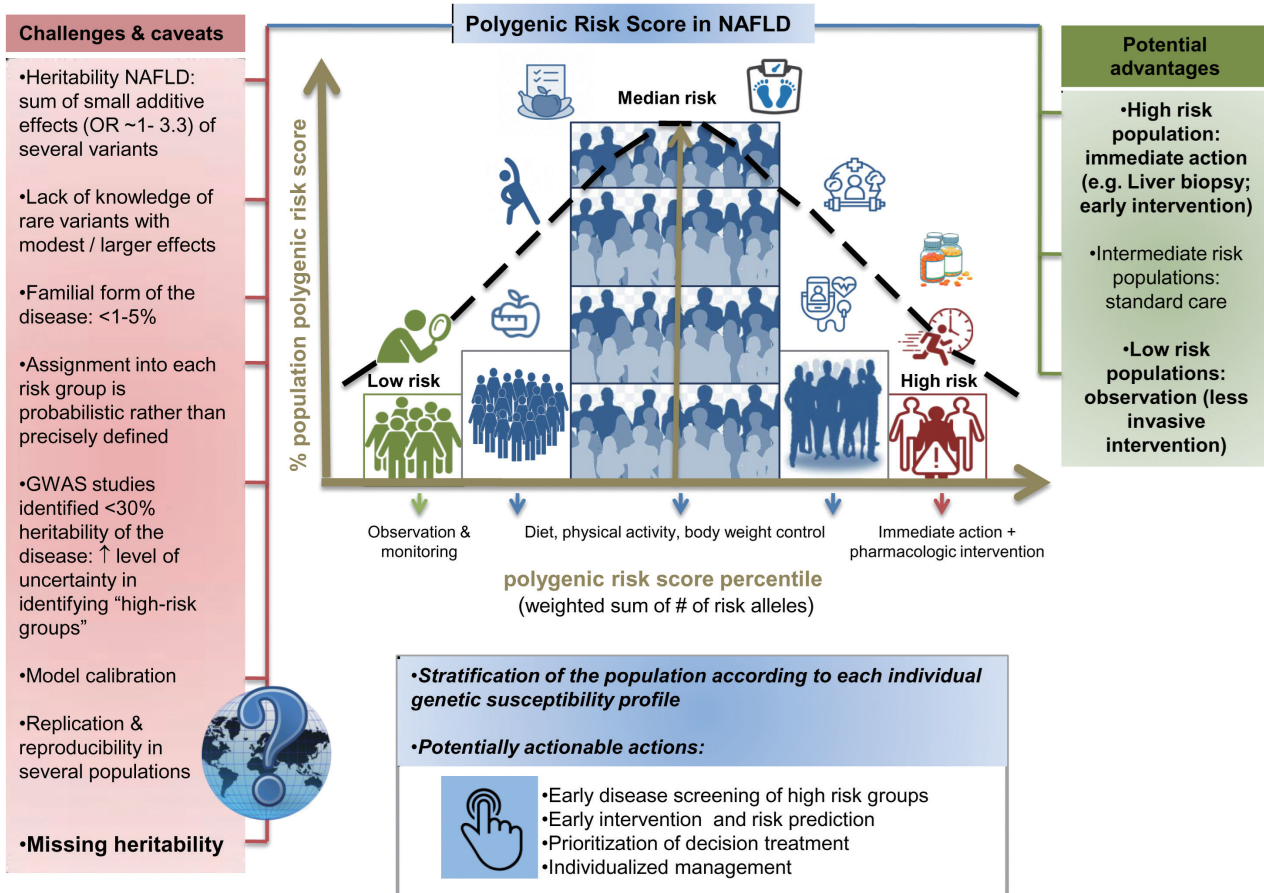
Figure 4. Polygenic risk score in NAFLD: Advantages and challenges. Theoretical frame for a PRS for NAFLD, showing advantages and potential caveats. The figure shows a typical bell-shaped distribution, in which scores pertaining to most individuals will be in the middle, indicating average risk of developing the disease. Those with scores located at the left and right tail of the distribution curve will respectively carry very low and very high risk. NAFLD, nonalcoholic fatty liver disease; PRS, polygenic risk score; OR, odds ratio; GWAS, genome-wide association study.

ducibility and replication of genetic variants of NAFLD across diverse populations around the world is insufficient (Fig. 4), and the effect sizes of most of the variants are yet to be established. Indeed, in the large majority of GWAS focusing on NAFLD, the genetic susceptibility in patients of European ancestry has been examined. This creates a considerable gap in extant knowledge, as SNP-based information on Caucasian and/or people of European descent may not be relevant for inferring the relative risk of NAFLD in non-European populations. Significant efforts have been made, however, to overcome this limitation. For example, Kawaguchi and coworkers conducted a GWAS in Japanese population and demonstrated that patients with NASH are genetically and clinically different from other population subgroups.[38] In addition, as a part of a large population-based GWAS that involved 1,593 patients and 2,816 controls, Chung and coworkers characterized the genetic profile of Korean NAFLD patients.[39] With the

exception of these remarkable examples from Asia, the percentage of non-European ancestry population in NAFLD GWAS studies, including those of African descent and/or other ethnic minorities, is dramatically low.

In addition to these shortcomings, certain technical issues, such as model calibration and calculation algorithms, must be overcome to fully benefit from the PRS implementation.

Genetic markers are already being used as tools for personalizing clinical practice, including treatment decisions.[4] Nevertheless, the utility of genetic variants in NAFLD risk estimation remains inferior to classical predictive or imaging approaches, as explained earlier.[4] In fact, knowledge of population structure and global heterogeneity of variants implicated in the disease progression is rather limited.

Thus, to simulate the potential utility of a NAFLD-PRS we used population-specific distribution information on the four aforemen-

tioned SNPs (*PNPLA3*-rs738409, *TM6SF2*-58542926, *HSD17B13*-rs72613567, and *GCKR*-780094) and the intergenic variant *LYPLAL1*-rs12137855.[22] To compute the NAFLD-PRS, we used the GlobAl Distribution of GEnetic Traits (GADGET) web server available at https://gadget.biosci.gatech.edu/compute/; the formula used to compute the score is based on the original description of Chande et al.[40] The GADGET web server provides access to publicly available genotype data sourced from the 1000 Genomes Project (1KGP) Phase 3 data release and individual trait SNP sets parsed directly from the NHGRI-EBI GWAS Catalog annotations (https://www.ebi.ac.uk/gwas/).[40] Box plots representing NAFLD-PRS distribution by five major continental groups (Africa, Europe, South and East Asia, and America) are shown in Figure 5. According to this information, the overall predicted risk to individuals entailed by the presence of NAFLD-implicated variants in their genomes is about ~0.20. Nevertheless, the PRS or genetic risk score (GRS) reflects disparities in NAFLD risk levels across different populations (Fig. 5), which may be due to the disparities in genetic knowledge or may indicate real differences in the genetic risk.

Thus, our analysis emphasizes the potential caveats of implementing this strategy globally. Of course, our simulation-based approach, which is based on information available in public databases, did not allow us to control for all possible effects, including demographic variables and/or other population risk estimates. Nonetheless, the GRS presented here provides a good starting point for illustrating the current situation, as well as for further investigations aimed at closing the gap in the knowledge needed to generate valuable advances in this field. It is worth noting that, as is the case for almost all human traits, variance in the NAFLD genetic risk within each population is much greater than among continental groups. This fact does not, however, imply that there is no continental group-specific risk profile. Available evidence indicates that certain trends exist at the global level, whereby the lowest GRS is associated with African population, intermediate GRS relates to European and Southeast Asian, and the highest GRS to East Asian and Admixed American groups.
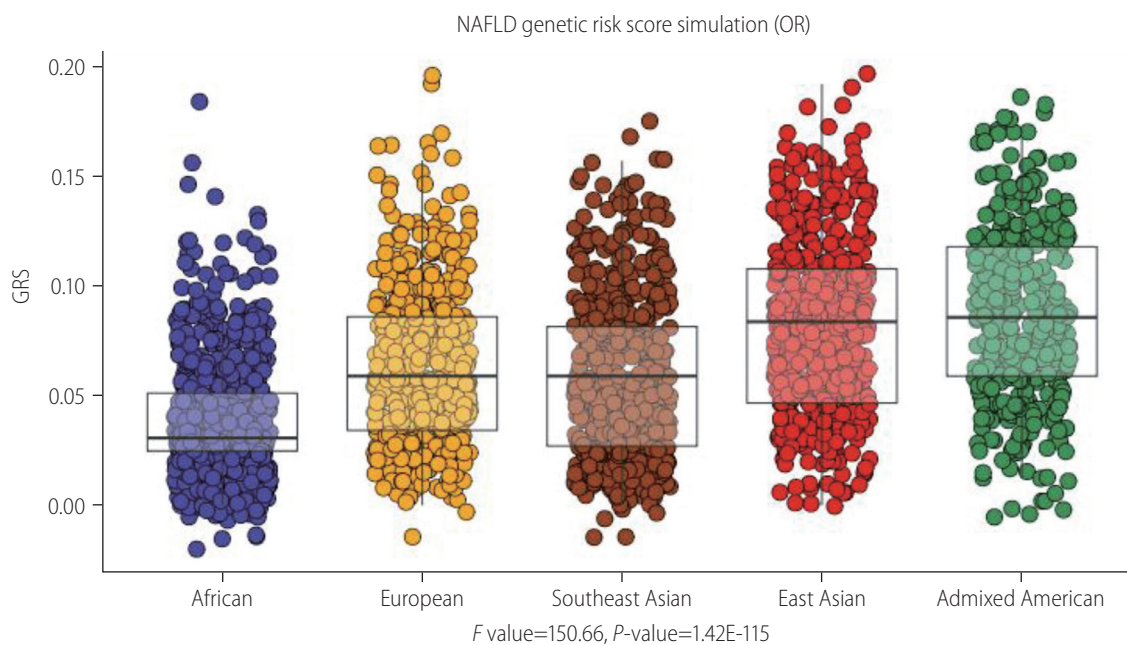
Figure 5. NAFLD PRSs (GRS) across the five major continental population groups. Box plots show population-specific distributions of genetic variants that have been associated with NAFLD in the literature (*PNPLA3*-rs738409, *TM6SF2*-58542926, *HSD17B13*-rs72613567, *GCKR*-780094, and the intergenic variant *LYPLAL1*-rs12137855), as well as medians and standard deviations. Admixed American (n=347, 0.0857±0.0387), African (n=661, 0.0306±0.0265), East Asian (n=504, 0.0837±0.0386), European (n=503, 0.0589±0.0351), Southeast Asian (n=489, 0.0589±0.0364). GRS: the relative risk of developing NAFLD based on the total number of variants associated with the disease the individual carries. The relative genetic risk of NAFLD within the population is shown as log ORs, with *F* and *P* denoting summarized linear regression. The formula by which the GRS was calculated can be found in the original contribution of Chande et al.[40] NAFLD, nonalcoholic fatty liver disease; PRS, polygenic risk score; OR, odds ratio; GRS, genetic risk score; *PNPLA3*, patatin-like phospholipase domain containing 3; *TM6SF2*, transmembrane 6 superfamily member 2; *HSD17B13*, hydroxysteroid 17-beta dehydrogenase 13; *GCKR*, glucokinase regulator.
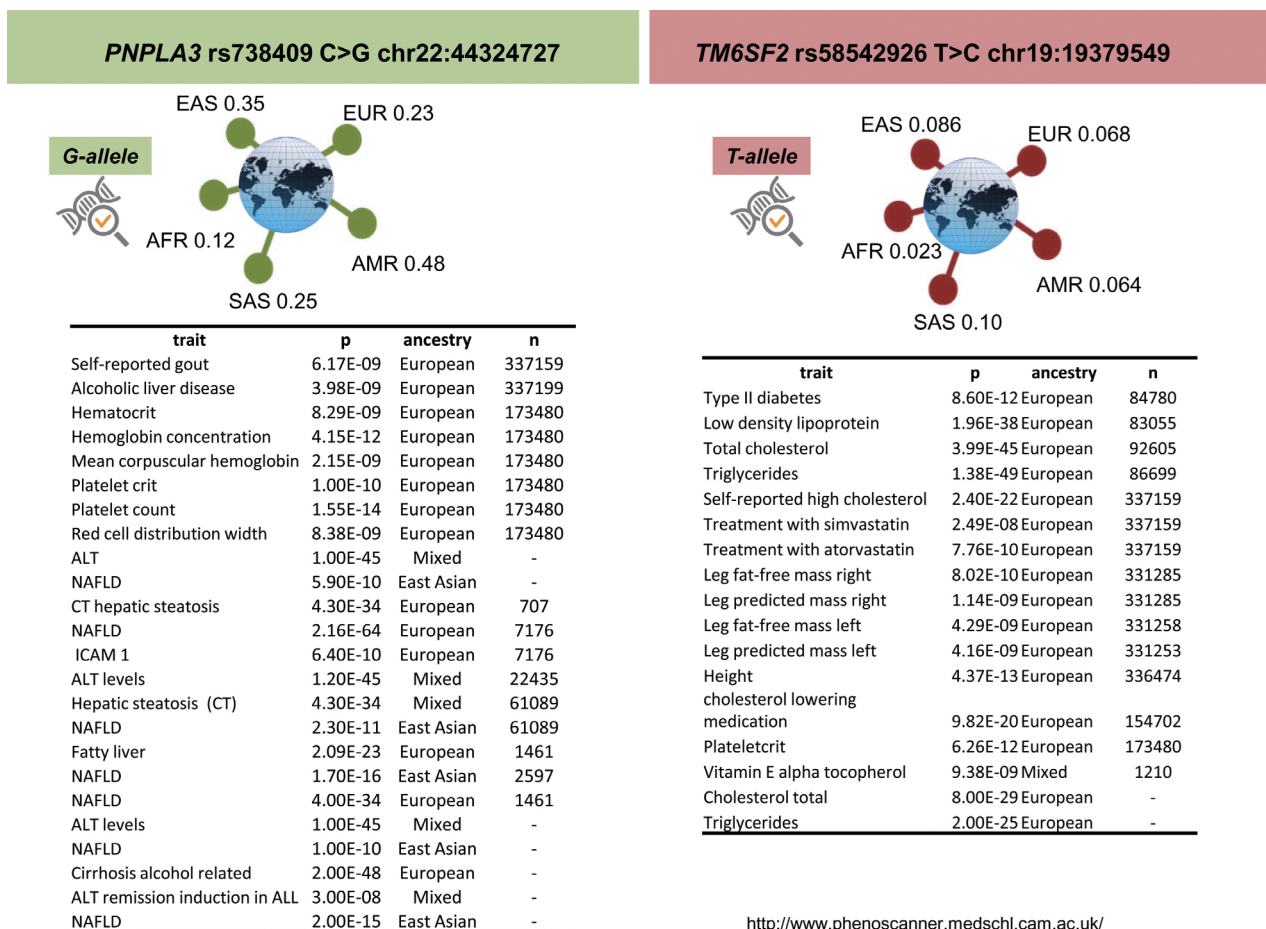
NAFLD GENES, PLEIOTROPIC RELATIONSHIPS, AND PRECISION MEDICINE

Available evidence suggests that genetic factors associated with NAFLD exhibit similar patterns of correlation with genetic factors related to other complex diseases.[4] For instance, findings yielded by large GWAS studies indicate that *PNPLA3*-r738409 and *TM6SF2*-rs58542926 are associated with extra-hepatic traits, including hematological (plateletcrit and count) and lipid traits, and some other interesting pharmacogenetic associations (http://www.phenoscanner.medschl.cam.ac.uk/) (Fig. 6). Importantly, the minor allele frequency of these two variants across different populations supports the prevalent view that the major genetic modi-

fiers of NAFLD are likely ancestry-specific. Therefore, this observation should be specifically examined when designing precision medicine strategies.

The genetic pleiotropy between the aforementioned variants and non-liver related traits includes known NAFLD-associated co-morbidities, such as cardiovascular risk. Phenotypic covariation presents not only significant challenges in clinical practice but also imposes tremendous constraints on identifying novel therapeutic targets. The clinical paradox of *TM6SF2*-rs58542926 C>T is a clear example of that. The C (Glu167) allele has been consistently associated with increased cardiovascular risk,[41] and the T allele (Lys167) is known to be associated with a higher risk for NAFLD and NASH.[21,33,42,43] These opposite effects are dependent

PNPLA3 rs738409 C>G chr22:44324727

EAS 0.35 EUR 0.23 AFR 0.12 AMR 0.48 SAS 0.25 G-allele

trait	p	ancestry	n
Self-reported gout	6.17E-09	European	337159
Alcoholic liver disease	3.98E-09	European	337199
Hematocrit	8.29E-09	European	173480
Hemoglobin concentration	4.15E-12	European	173480
Mean corpuscular hemoglobin	2.15E-09	European	173480
Platelet crit	1.00E-10	European	173480
Platelet count	1.55E-14	European	173480
Red cell distribution width	8.38E-09	European	173480
ALT	1.00E-45	Mixed	-
NAFLD	5.90E-10	East Asian	-
CT hepatic steatosis	4.30E-34	European	707
NAFLD	2.16E-64	European	7176
ICAM 1	6.40E-10	European	7176
ALT levels	1.20E-45	Mixed	22435
Hepatic steatosis (CT)	4.30E-34	Mixed	61089
NAFLD	2.30E-11	East Asian	61089
Fatty liver	2.09E-23	European	1461
NAFLD	1.70E-16	East Asian	2597
NAFLD	4.00E-34	European	1461
ALT levels	1.00E-45	Mixed	-
NAFLD	1.00E-10	East Asian	-
Cirrhosis alcohol related	2.00E-48	European	-
ALT remission induction in ALL	3.00E-08	Mixed	-
NAFLD	2.00E-15	East Asian	-

TM6SF2 rs58542926 T>C chr19:19379549

EAS 0.086 EUR 0.068 AFR 0.023 AMR 0.064 SAS 0.10 T-allele

| trait	p	ancestry	n
Type II diabetes	8.60E-12	European	84780
Low density lipoprotein	1.96E-38	European	83055
Total cholesterol	3.99E-45	European	92605
Triglycerides	1.38E-49	European	86699
Self-reported high cholesterol	2.40E-22	European	337159
Treatment with simvastatin	2.49E-08	European	337159
Treatment with atorvastatin	7.76E-10	European	337159
Leg fat-free mass right	8.02E-10	European	331285
Leg predicted mass right	1.14E-09	European	331285
Leg fat-free mass left	4.29E-09	European	331258
Leg predicted mass left	4.16E-09	European	331253
Height	4.37E-13	European	336474
cholesterol lowering medication	9.82E-20	European	154702
Plateletcrit	6.26E-12	European	173480
Vitamin E alpha tocopherol	9.38E-09	Mixed	1210
Cholesterol total	8.00E-29	European	-
Triglycerides	2.00E-25	European	-

http://www.phenoscanner.medschl.cam.ac.uk/

Figure 6. *PNPLA3*, *TM6SF2*, and pleiotropic relationships. Pleiotropic associations with rs738409 (*PNPLA3*) and rs58542925 (*TM6SF2*) variants, explored by the PhenoScanner web tool available at http://www.phenoscanner.medschl.cam.ac.uk, a database of human genotype-phenotype associations. Associations are based on publicly available results from large-scale genetic association studies; Phenoscanner collated >5,000 genotype-phenotype association datasets. *PNPLA3*, patatin-like phospholipase domain containing 3; EAS, East Asian; EUR, European; AFR, African; SAS, South Asian; AMR, American; n, sample size; ALT, alanine aminotransferase; NAFLD, nonalcoholic fatty liver disease; CT, computed tomography; *TM6SF2*, transmembrane 6 superfamily member 2.

on circulating and liver triglyceride levels, respectively. Consequently, *TM6SF2* does not seem to be a useful drug target because any impact on the protein would eventually lower blood lipids, which will in turn reduce the risk of myocardial infarction, while simultaneously increasing the risk of developing NAFLD.[42]

As explained in previous paragraphs, drug development is a long process characterized by highly uncertain outcomes. At present, 10–15 years typically elapse between target discovery and clinical application. Hence, some computational solutions, including modeling and over-representation analysis grounded in systems biology, may assist with more accurate prediction of drug candidates based on disease-associated genes/proteins. To illustrate this concept, we employed two different strategies. First, we leveraged existing information on the molecular targets (genes/proteins) involved in NAFLD/NASH pathogenesis and performed over-representation analysis using a drug-related functional database. The training set of genes/proteins was obtained by literature-data mining offered by the Genie web server (http://cbdm-01.zdv.uni-mainz.de/~jfontain/cms/?page_id=281)—a tool that computes associations of genes with diseases using biomedical

Over-representation analysis: Enrichment Categories: drug_GLAD4U

Gene-chemical interactions
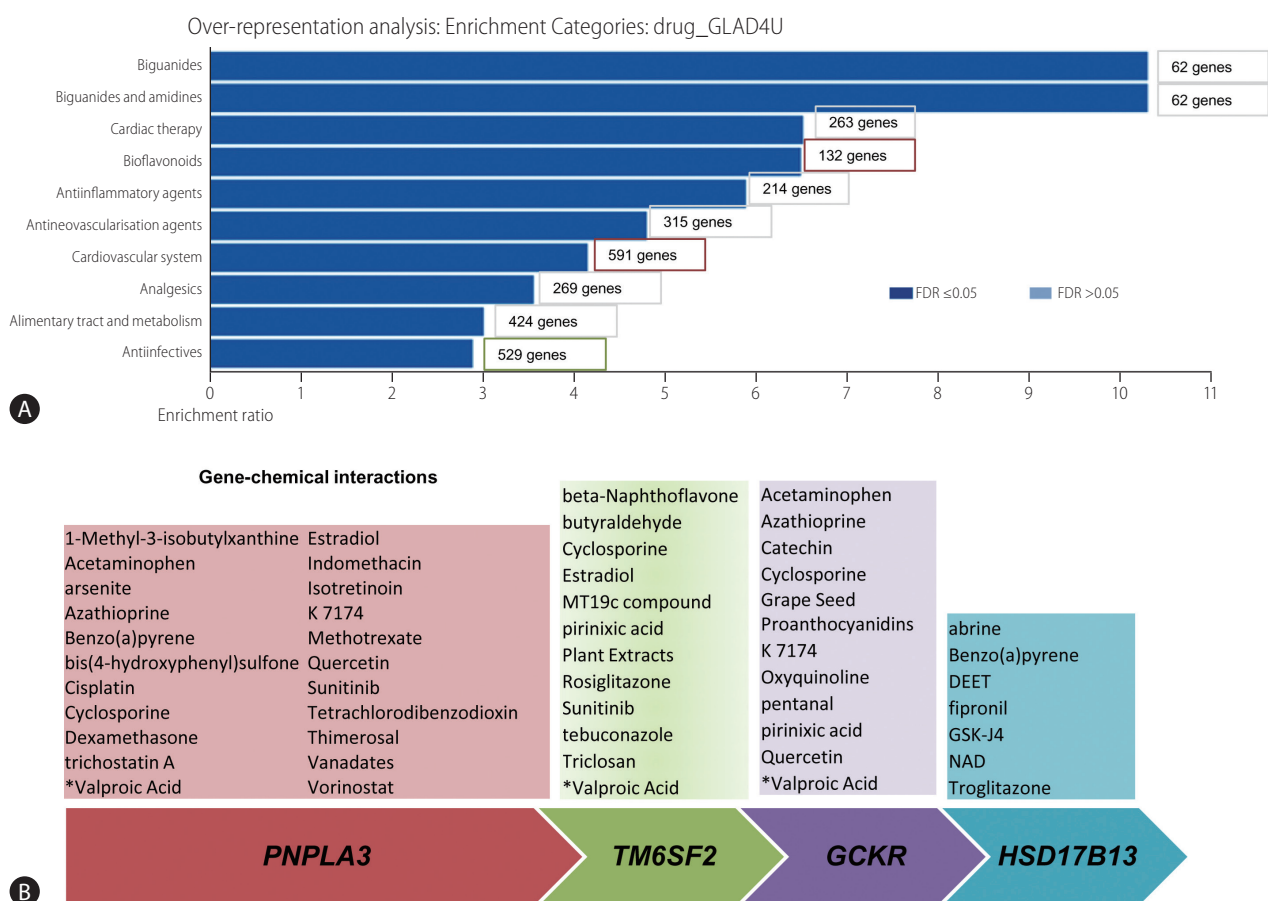
Figure 7. Prediction of genetic-drug/chemical interaction profiles. (A) Over-representation analysis using a drug-related functional database (drug_GLAD4U). The training set of genes/proteins was obtained by literature-data mining offered by the Genie web server. Cut-offs used: *P*<0.01 for abstracts and *P*<0.01 for FDR for genes. The list of retrieved genes/proteins was used to perform over-representation analysis in the web server WebGestalt (WEB-based Gene SeT AnaLysis Toolkit). The bar chart shows ten categories that passed the FDR <0.05, with the gene number denoting the number of genes in the training list that belong to each of the drug categories. (B) Gene-chemical interactions. Gene target prediction was performed using the Comparative Toxicogenomics Database available at (http://ctdbase.org). The list of chemicals was manually curated to restrict interactions based on human data. According to the Comparative Toxicogenomics Database available, chemical-gene and protein interactions are curated from the published literature. Interactions may be retrieved by chemical, interaction type, gene, organism, or Gene Ontology annotation. Tutorial and algorithms by which the list of gene-chemical interactions was done are available at http://ctdbase.org and http://ctdbase.org/documents/ctd_resource_guide.pdf. FDR, false discovery rate; *PNPLA3*, patatin-like phospholipase domain containing 3; *TM6SF2*, transmembrane 6 superfamily member 2; *GCKR*, glucokinase regulator; *HSD17B13*, hydroxysteroid 17-beta dehydrogenase 13; DEET, N,N-diethyl-3-methylbenzamide; NAD, nicotinamide adenine dinucleotide.

literature annotations. Using this approach, 901 abstracts from PubMed were retrieved using the search terms "fatty liver" and "human" (taxonomic identifier 9606; no literature extension by orthology) with the abstracts from the all PubMed database serving as the background set. Supplementary Table 1, intended for online publication only, shows the final ranked list comprising of

Table 1. Prediction of valproic acid pathways

KEGG pathway ID*	Pathway name
hsa:00071	Fatty acid metabolism
hsa:00140	C21-Steroid hormone metabolism
hsa:00232	Caffeine metabolism
hsa:00250	Alanine, aspartate and glutamate metabolism
hsa:00280	Valine, leucine and isoleucine degradation
hsa:00380	Tryptophan metabolism
hsa:00410	Beta-alanine metabolism
hsa:00590	Arachidonic acid metabolism
hsa:00591	Linoleic acid metabolism
hsa:00640/hsa:00650	Propanoate metabolism/butanoate metabolism
hsa:00830	Retinol metabolism
hsa:00980	Metabolism of xenobiotics by cytochrome P450
hsa:00982/hsa:00983	Drug metabolism
hsa:03320	PPAR signaling pathway
hsa:04012	ErbB signaling pathway
hsa:04062	Chemokine signaling pathway
hsa:04080	Neuroactive ligand-receptor interaction
hsa:04110	Cell cycle
hsa:04270	Vascular smooth muscle contraction
hsa:04310	Wnt signaling pathway
hsa:04340	Hedgehog signaling pathway
hsa:04510	Focal adhesion
hsa:04610	Complement and coagulation cascades
hsa:04660/hsa:04662	T cell receptor signaling pathway/B cell receptor signaling pathway
hsa:04722	Neurotrophin signaling pathway
hsa:04810	Regulation of actin cytoskeleton
hsa:04910	Insulin signaling pathway

Prediction was performed by the Supertarget tool available at http://bioinformatics.charite.de/supertarget/index.php?site=drugs.
KEGG, Kyoto Encyclopedia of Genes and Genomes; PPAR, peroxisome proliferator-activated receptor gamma and alpha.
*Pathway IDs correspond to KEGG.

938 genes/proteins.

The aforementioned list of retrieved genes/proteins was used to perform over-representation analysis in the webserver WebGestalt (WEB-based Gene SeT AnaLysis Toolkit)—a functional enrichment analysis tool available at http://www.webgestalt.org/. Drug terms were downloaded from ParmGKB by the WebGestalt, and individual drug terms associated with genes were inferred using the GLAD4U option. Accordingly, the drug enrichment results based on the NAFLD training set are shown in Figure 7A. The bar chart shows ten categories that passed the FDR <0.05, whereby cardiac therapy and cardiovascular system were overrepresented by the largest number of genes (263 and 591, respectively). Other expected drug categories are anti-inflammatory agents and biguanides, which are probably justified by the significant enrichment of genes/proteins associated with inflammation and glucose metabolism in the training set. In fact, the top ten genes/proteins of the training list were *ADIPOQ* (adiponectin), *PNPLA3*, *PPARG* and *PPARA* (peroxisome proliferator-activated receptor gamma and alpha), *FGF21* (fibroblast growth factor 21), *RBP4* (retinol binding protein 4), *GPT* (glutamic-pyruvic transaminase), *SREBF1* (sterol regulatory element binding transcription factor 1), *LEP* (leptin), and *NR1H3* (nuclear receptor subfamily 1 group H member 3, also known as liver X receptor alpha). Lastly, the analysis revealed several anti-infection drugs, which were also expected because the training set presents several genes/proteins associated with immune response. Hence, the chart may be useful for inferring drug classes or drug compounds that could be repurposed for the treatment of NASH.

Our second strategy was based on using the four genes largely and reproducibly associated with NAFLD and NASH (*PNPLA3*, *TM6SF2*, *GCKR*, and *HSD17B13*) to predict curated chemical–gene/protein interactions in the Comparative Toxicogenomics Database available at (http://ctdbase.org) (Fig. 7B). The aim of this approach was to infer and/or uncover potential associated disease mechanisms from genetic predisposing factors that can yield biologically informative insights. Notably, some drugs were consistently found to interact with three of the loci, for example valproic acid (valproate) (Fig. 7B).

Valproate appears to impact on fatty acid metabolism, and the use of valproate has been linked to the development of obesity and probably NAFLD.[44] Furthermore, valproate acts as a direct histone deactylase (HDAC) inhibitor.[45] While tissue-specific DNA methylation in NAFLD and NASH, including 5-hydroxymethylcytosine (5-hmC) has been previously studied,[46-48] the role of other epigenetic mechanisms, including acetylation and deacetylation

of histones, remains to be fully ascertained.[2] Likewise, the use of valproate has been associated with the inhibition of mitochondrial beta-oxidation and peroxisomal stimulation in rodent livers,[49] which reinforces the concept that the progression of NAFLD into severe clinical and histological forms involves mitochondrial dysfunction.[48,50-52] Table 1 provides a complete list of predicted valproic acid-KEGG pathways, which involve alanine, aspartate and glutamate metabolism, arachidonic acid metabolism, retinol metabolism, PPAR and insulin signaling pathway, regulation of actin cytoskeleton, and hedgehog signaling pathway, among many other pathways relevant to the NAFLD pathogenesis.[12] In addition, the list of valproate-linked pathways contains propanoate and butanoate metabolism that has been linked to the NASH-associated tissue microbiome.[12]

Another compound that appears to be linked to three of the genes (*PNPLA3*, *TM6SF2*, and *GCKR*) is quercetin, an antioxidant phenolic heterocyclic compound that is a specific quinone reductase 2 (QR2) inhibitor (Fig. 7B). Collectively, these genetic-chemical interactions provide valuable information about cellular and biological mechanisms of disease, suggesting that this strategy may be used as a complement in the target-based drug development as well.

LIMITATIONS OF THE SYSTEMS BIOLOGY APPROACH

Some limitations of the systems biology approach must be highlighted, including the restricted possibility of adequately addressing the gender dimension of the disease. Sexual dimorphism is observed not only in the prevalence of NAFLD but also in the disease pathogenesis and diverse histological outcomes.[53,54] Although sex differences substantially contribute to the biology of NAFLD, there is limited information on the sex-specific genetic architecture of the disease. We have addressed this aspect by meta-regression analysis of studies that assessed the effect of rs738409 on NAFLD, and we found a negative correlation between the male proportion in the studied populations and the effect of the SNP on liver fat content,[19] suggesting that sexual dimorphism might be involved in the impact of the variant on NAFLD development. Likewise, a recent study that involved large-scale analysis of transcriptomic profiles from human livers suggested the sexually dimorphic nature of NASH and its link with fibrosis and responses to drugs.[55]

Unrevealing the genetic mechanisms that contribute to sex-spe-cific NAFLD risk should be urgently addressed in further candidate gene association or GWAS studies of NAFLD. Elucidating the sex-specific genetic architecture of NAFLD represents an important area for future research. Besides, there are many other important genetic and epigenetic factors,[2,46,48] including mitochondrial genetics,[51,56] that play a substantial role in the disease biology, and that deserve a more detailed analysis.

CONCLUSIONS

The utility of the systems biology approach for accelerating the NASH drug discovery process remains to be established. Still, to attain its full potential, refined genomic strategies must be implemented to increase knowledge of genetic susceptibility across all ancestry groups. Although theoretically powerful, PRSs need to be validated and built, not only on validated GWAS-variants but also on robust and global genetic information, preferably specific to each ethnic group. Finally, disciplines such as chemical genetics must be used in tandem with traditional drug discovery approaches to accelerate the progress of precision medicine in NASH and to reveal the druggable NASH genome/proteome.

Authors' contribution

CJP and SS designed the study, performed the analyses, analyzed and interpreted the data, and prepared and wrote the manuscript. Both authors have read and approved the final manuscript.

Conflicts of Interest

The authors have no conflicts to disclose.

SUPPLEMENTARY MATERIAL

Supplementary material is available at Clinical and Molecular

Hepatology website (http://www.e-cmh.org).

REFERENCES

1. Brunt EM, Wong VW, Nobili V, Day CP, Sookoian S, Maher JJ, et al. Nonalcoholic fatty liver disease. Nat Rev Dis Primers 2015;1:15080.

2. Pirola CJ, Sookoian S. Epigenetics factors in nonalcoholic fatty liver disease. Expert Rev Gastroenterol Hepatol. 2020 Jun 1. doi: 10.1080/17474124.2020.1765772.

3. Sookoian S, Pirola CJ. Genetic predisposition in nonalcoholic fatty liver disease. Clin Mol Hepatol 2017;23:1-12.

4. Sookoian S, Pirola CJ. Genetics of nonalcoholic fatty liver disease: from pathogenesis to therapeutics. Semin Liver Dis 2019;39:124-140.

5. Chalasani N, Abdelmalek MF, Garcia-Tsao G, Vuppalanchi R, Alkhouri N, Rinella M, et al. Effects of belapectin, an inhibitor of galectin-3, in patients with nonalcoholic steatohepatitis with cirrhosis and portal hypertension. Gastroenterology 2020;158:1334-1345.e5.

6. Garcia-Tsao G, Bosch J, Kayali Z, Harrison SA, Abdelmalek MF, Lawitz E, et al. Randomized placebo-controlled trial of emricasan for non-alcoholic steatohepatitis-related cirrhosis with severe portal hypertension. J Hepatol 2020;72:885-895.

7. Harrison SA, Abdelmalek MF, Caldwell S, Shiffman ML, Diehl AM, Ghalib R, et al. Simtuzumab is ineffective for patients with bridging fibrosis or compensated cirrhosis caused by nonalcoholic steatohepatitis. Gastroenterology 2018;155:1140-1153.

8. Harrison SA, Wong VW, Okanoue T, Bzowej N, Vuppalanchi R, Younes Z, et al. Selonsertib for patients with bridging fibrosis or compensated cirrhosis due to NASH: results from randomized phase III STELLAR trials. J Hepatol 2020;73:26-39.

9. Harrison SA, Goodman Z, Jabbar A, Vemulapalli R, Younes ZH, Freilich B, et al. A randomized, placebo-controlled trial of emricasan in patients with NASH and F1-F3 fibrosis. J Hepatol 2020;72:816-827.

10. Loomba R, Lawitz E, Mantry PS, Jayakumar S, Caldwell SH, Arnold H, et al. The ASK1 inhibitor selonsertib in patients with nonalcoholic steatohepatitis: a randomized, phase 2 trial. Hepatology 2018;67:549-559.

11. Ratziu V, Friedman SL. Why do so many NASH trials fail? Gastroenterology. 2020 May 18. doi: 10.1053/j.gastro.2020.05.046.

12. Sookoian S, Pirola CJ, Valenti L, Davidson NO. Genetic pathways in nonalcoholic fatty liver disease: insights from systems biology. Hepatology 2020;72:330-346.

13. Friedman SL, Neuschwander-Tetri BA, Rinella M, Sanyal AJ. Mechanisms of NAFLD development and therapeutic strategies. Nat Med 2018;24:908-922.

14. Pirola CJ, Sookoian S. Multiomics biomarkers for the prediction of nonalcoholic fatty liver disease severity. World J Gastroenterol 2018;24:1601-1615.

15. Sookoian S, Salatino A, Castaño GO, Landa MS, Fijalkowky C, Garaycoechea M, et al. Intrahepatic bacterial metataxonomic signature in non-alcoholic fatty liver disease. Gut 2020;69:1483-1491.

16. Sookoian S, Pirola CJ. Repurposing drugs to target nonalcoholic steatohepatitis. World J Gastroenterol 2019;25:1783-1796.

17. Romeo S, Kozlitina J, Xing C, Pertsemlidis A, Cox D, Pennacchio LA, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 2008;40:1461-1465.

18. 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;50:2111-2116.

19. 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 nonalcoholic fatty liver disease. Hepatology 2011;53:1883-1894.

20. Abul-Husn NS, Cheng X, Li AH, Xin Y, Schurmann C, Stevis P, et al. A protein-truncating HSD17B13 variant and protection from chronic liver disease. N Engl J Med 2018;378:1096-1106.

21. Kozlitina J, Smagris E, Stender S, Nordestgaard BG, Zhou HH, Tybjærg-Hansen A, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. Nat Genet 2014;46:352-356.

22. Speliotes EK, Yerges-Armstrong LM, Wu J, Hernaez R, Kim LJ, Palmer CD, et al. Genome-wide association analysis identifies variants associated with nonalcoholic fatty liver disease that have distinct effects on metabolic traits. PLoS Genet 2011;7:e1001324.

23. Mancina RM, Dongiovanni P, Petta S, Pingitore P, Meroni M, Rametta R, et al. The MBOAT7-TMC4 variant rs641738 increases risk of nonalcoholic fatty liver disease in individuals of european descent. Gastroenterology 2016;150:1219-1230.e6.

24. Koo BK, Joo SK, Kim D, Bae JM, Park JH, Kim JH, et al. Additive effects of PNPLA3 and TM6SF2 on the histological severity of nonalcoholic fatty liver disease. J Gastroenterol Hepatol 2018;33:1277-1285.

25. Lin YC, Chang PF, Chang MH, Ni YH. Genetic determinants of hepatic steatosis and serum cytokeratin-18 fragment levels in Taiwanese children. Liver Int 2018;38:1300-1307.

26. Sookoian S, Flichman D, Garaycoechea ME, Gazzi C, Martino JS, Castaño GO, 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;8:5097.

27. Anstee QM, Darlay R, Cockell S, Meroni M, Govaere O, Tiniakos D, et al. Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically-characterised cohort. J Hepa-

tol 2020;73:505-515.

28. Pirola CJ, Garaycoechea M, Flichman D, Arrese M, San MJ, Gazzi C, et al. Splice variant rs72613567 prevents worst histologic outcomes in patients with nonalcoholic fatty liver disease. J Lipid Res 2019;60:176-185.

29. Sookoian S, Arrese M, Pirola CJ. Genetics meets therapy? Exomewide association study reveals a loss-of-function variant in 17-beta-hydroxysteroid dehydrogenase 13 that protects patients from liver damage and nonalcoholic fatty liver disease progression. Hepatology 2019;69:907-910.

30. Sookoian S, Pirola CJ. Liver tissue microbiota in nonalcoholic liver disease: a change in the paradigm of host-bacterial interactions. Hepatobiliary Surg Nutr. 2020. doi: 10.21037/hbsn-20-270.

31. 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;13:4242-4248.

32. Sookoian S, Gemma C, Gianotti TF, Burgueno A, Castano G, Pirola CJ. Genetic variants of clock transcription factor are associated with individual susceptibility to obesity. Am J Clin Nutr 2008;87:1606-1615.

33. Dongiovanni P, Petta S, Maglio C, Fracanzani AL, Pipitone R, Mozzi E, et al. Transmembrane 6 superfamily member 2 gene variant disentangles nonalcoholic steatohepatitis from cardiovascular disease. Hepatology 2015;61:506-514.

34. Sookoian S, Pirola CJ. Nonalcoholic fatty liver disease and metabolic syndrome: shared genetic basis of pathogenesis. Hepatology 2016;64:1417-1420.

35. Sookoian S, Pirola CJ. Meta-analysis of the influence of TM6SF2 E167K variant on plasma concentration of aminotransferases across different populations and diverse liver phenotypes. Sci Rep 2016;6:27718.

36. 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;30:21-27.

37. Pirola CJ, Flichman D, Dopazo H, Fernández GT, San MJ, Rohr C, 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;2:1030-1036.

38. Kawaguchi T, Shima T, Mizuno M, Mitsumoto Y, Umemura A, Kanbara Y, et al. Risk estimation model for nonalcoholic fatty liver disease in the Japanese using multiple genetic markers. PLoS One 2018;13:e0185490.

39. Chung GE, Lee Y, Yim JY, Choe EK, Kwak MS, Yang JI, et al. Genetic polymorphisms of PNPLA3 and SAMM50 are associated with nonalcoholic fatty liver disease in a Korean population. Gut Liver 2018;12:316-323.

40. Chande AT, Wang L, Rishishwar L, Conley AB, Norris ET, Valderra-

ma-Aguirre A, et al. GlobAl Distribution of GEnetic Traits (GADGET) web server: polygenic trait scores worldwide. Nucleic Acids Res 2018;46:W121-W126.

41. Mahdessian H, Taxiarchis A, Popov S, Silveira A, Franco-Cereceda A, Hamsten A, 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;111:8913-8918.

42. 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;62:1742-1756.

43. Sookoian S, Castaño GO, Scian R, Mallardi P, Fernández GT, Burgueño AL, et al. Genetic variation in transmembrane 6 superfamily member 2 and the risk of nonalcoholic fatty liver disease and histological disease severity. Hepatology 2015;61:515-525.

44. Farinelli E, Giampaoli D, Cenciarini A, Cercado E, Verrotti A. Valproic acid and nonalcoholic fatty liver disease: a possible association? World J Hepatol 2015;7:1251-1257.

45. Phiel CJ, Zhang F, Huang EY, Guenther MG, Lazar MA, Klein PS. Histone deacetylase is a direct target of valproic acid, a potent anticonvulsant, mood stabilizer, and teratogen. J Biol Chem 2001;276:36734-36741.

46. Pirola CJ, Gianotti TF, Burgueño AL, Rey-Funes M, Loidl CF, Mallardi P, et al. Epigenetic modification of liver mitochondrial DNA is associated with histological severity of nonalcoholic fatty liver disease. Gut 2013;62:1356-1363.

47. Pirola CJ, Scian R, Gianotti TF, Dopazo H, Rohr C, Martino JS, et al. Epigenetic modifications in the biology of nonalcoholic fatty liver disease: the role of DNA hydroxymethylation and TET proteins. Medicine (Baltimore) 2015;94:e1480.

48. Sookoian S, Rosselli MS, Gemma C, Burgueño AL, Fernández GT, Castaño GO, et al. Epigenetic regulation of insulin resistance in nonalcoholic fatty liver disease: impact of liver methylation of the peroxisome proliferator-activated receptor gamma coactivator 1alpha promoter. Hepatology 2010;52:1992-2000.

49. Veitch K, Draye JP, Van HF. Inhibition of mitochondrial beta-oxidation and peroxisomal stimulation in rodent livers by valproate. Biochem Soc Trans 1989;17:1070-1071.

50. Sanyal AJ, Campbell-Sargent C, Mirshahi F, Rizzo WB, Contos MJ, Sterling RK, et al. Nonalcoholic steatohepatitis: association of insulin resistance and mitochondrial abnormalities. Gastroenterology 2001;120:1183-1192.

51. Sookoian S, Flichman D, Scian R, Rohr C, Dopazo H, Gianotti TF, et al. Mitochondrial genome architecture in non-alcoholic fatty liver disease. J Pathol 2016;240:437-449.

52. Sookoian S, Castaño GO, Scian R, Fernández GT, Dopazo H, Rohr C, et al. Serum aminotransferases in nonalcoholic fatty liver disease are a signature of liver metabolic perturbations at the amino acid and

Krebs cycle level. Am J Clin Nutr 2016;103:422-434.

53. Lonardo A, Nascimbeni F, Ballestri S, Fairweather D, Win S, Than TA, et al. Sex differences in nonalcoholic fatty liver disease: state of the art and identification of research gaps. Hepatology 2019;70:1457-1469.

54. Lonardo A, Suzuki A. Sexual Dimorphism of NAFLD in Adults. Focus on clinical aspects and implications for practice and translational research. J Clin Med 2020;9:1278.

55. Vandel J, Dubois-Chevalier J, Gheeraert C, Derudas B, Raverdy V, Thuillier D, et al. Hepatic molecular signatures highlight the sexual dimorphism of Non-Alcoholic SteatoHepatitis (NASH). Hepatology. 2020 May 11. doi: 10.1002/hep.31312.

56. Pirola CJ, Garaycoechea M, Flichman D, Castaño GO, Sookoian S. Liver mitochondrial DNA damage and genetic variability of Cytochrome

b - a key component of the respirasome - drive the severity of fatty liver disease. J Intern Med. 2020 Jul 7. doi: 10.1111/joim.13147.

57. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, et al. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database (Oxford) 2016;2016:baw100.

58. Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, et al. Massive mining of publicly available RNA-seq data from human and mouse. Nat Commun 2018;9:1366.

59. Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'ayan A. ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics 2010;26:2438-2444.