Genetics in non-alcoholic fatty liver disease: The role of risk alleles through the lens of immune response

Running title: Genetics in NAFLD

Silvia Sookoian 1*, Carlos J Pirola 2*

1 Clinical and Molecular Hepatology. Centro de Altos Estudios en Ciencias Humanas y de la Salud (CAECIHS), Universidad Abierta Interamericana, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina.

2 Systems Biology of Complex Diseases. Centro de Altos Estudios en Ciencias Humanas y de la Salud (CAECIHS), Universidad Abierta Interamericana, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina.

* Co-corresponding authorship

Correspondence: Co-Corresponding Authors

Silvia Sookoian, M.D., Ph.D (Orcid ID: 0000-0001-5929-5470) and Carlos J. Pirola, Ph.D. (Orcid ID: 0000-0001-8234-4058)

Av. Montes de Oca 745, 2 piso, (1270AAH) CABA. Argentina.

E-mails: ssookoian@intramed.net and pirola.carlos@conicet.gov.ar

Phone: 54 11 4301-5240-5323

SS and CJP should be considered joint senior authors

Abbreviations

GCKR, glucokinase regulator

GWAS, genome-wide association study

HSD17B13, hydroxysteroid 17-beta dehydrogenase 13

MBOAT7, membrane bound O-acyltransferase domain containing 7

NAFLD, non-alcoholic fatty liver disease

NASH, non-alcoholic steatohepatitis
PNPLA3, patatin-like phospholipase domain containing 3
SNP, single nucleotide polymorphism
TM6SF2, transmembrane 6 superfamily member 2

Conflicts of Interest
The authors have no conflicts to disclose

Authors' contribution
C.J.P Concept of the work, manuscript writing and approval. S.S. Concept of the work, manuscript writing and approval.

Funding
This study was partially supported by grants PICT 2018-889, PICT 2019-0528, PICT 2018-00620, PICT 2020-799 (Agencia Nacional de Promoción Científica y Tecnológica, FONCyT), CONICET Proyectos Unidades Ejecutoras 2017, PUE 0055.
ABSTRACT
The knowledge of the genetic component of non-alcoholic fatty liver disease (NAFLD) has grown exponentially over the last 10-15 years. This review summarizes current evidence and the latest developments in the genetics of NAFLD and non-alcoholic steatohepatitis (NASH) from the immune system's perspective.

Activation of innate and/or adaptive immune response is an essential driver of NAFLD disease severity and progression. Lipid and immune pathways are crucial in the pathophysiology of NAFLD and NASH. Here, we highlight novel applications of genomic techniques, including single-cell sequencing and the genetics of gene expression, to elucidate the potential involvement of NAFLD/NASH-risk alleles in modulating immune system cells.

Together, our focus is to provide an overview of the potential involvement of the NAFLD/NASH-related risk variants in mediating the immune-driven liver disease severity and diverse systemic pleiotropic effect/s.

Key words: NAFLD; NASH; genetics; PNPLA3; HSD17B13; immune system
The global trends in the prevalence and incidence of NAFLD represent a significant public health challenge. The disease prevalence has reached alarming figures not only in adults but also in the children's population.\(^1\) Knowledge of the genetic component of non-alcoholic fatty liver disease (NAFLD) has grown exponentially over the last 10-15 years.\(^3\) With this knowledge it becomes possible to translate information of risk alleles and its effects on the disease biology into clinical applications.\(^6\) Most importantly, knowledge on the genetic component of NAFLD may be leveraged to identify individuals at risk and/or to estimate the risk of severe histological outcomes, including NASH (non-alcoholic steatohepatitis)-fibrosis, cirrhosis, and hepatocellular carcinoma.\(^6\)

While NAFLD is a disorder characterized by excess accumulation of fat in hepatocytes, in up to 40% of individuals, there are additional findings of portal and lobular inflammation and hepatocyte injury which characterize the severe histological forms of the disease.\(^3\) Therefore, activation of the immune system is a key feature of the disease severity and progression.\(^3\) Furthermore, NAFLD progressive clinical forms, including NASH-fibrosis, NASH-cirrhosis, and eventually hepatocellular carcinoma, are the main drivers of liver disease-associated mortality worldwide.\(^1\)

Although remarkable progress has been made in understanding the disease biology, it remains unclear how to link NAFLD/NASH-associated variants with immune-specific cells mechanistically and how to explain the role of genetics in immune-driven disease progression. In this review, we summarize current evidence and the latest developments in the field of genetics of NAFLD and NASH—the disease' severe histological form, from the perspective of the role of risk alleles in modulating gene expression of cells of the immune system. Our focus is to provide an overview of the potential involvement of the NAFLD/NASH-related risk variants in mediating the immune-driven disease severity.
A short overview of variants influencing the risk and protection against NAFLD and the histological disease severity

Genetic discoveries in the field of NAFLD have mainly been motorized by the use of genome-wide (GWAS) \(^9, 10\), exome-wide (EWAS) \(^11\), more recently, phenome-wide (PHEWAS) association studies using electronic Health Records \(^12\), as well as high-throughput sequencing technologies that allowed refining and mapping the discovered variants \(^13\).

Most relevant and replicated targets associated with the genetic component of NAFLD are illustrated in Figure 1, which depicts the primary protein function and subcellular localization. Notably, major candidate gene variants function in metabolic pathways (Figure 1).

Figure 2 summarizes the most replicated variants associated with NAFLD and NASH, including the global minor allele frequency, the variant's most severe consequence, the variant functionality, and the variant effect on the disease traits. It is interesting to point out that most of the variants associated with NAFLD and NASH are mapped to coding regions of the genome facilitating the variants' functional assessment (Figure 2).

The variants and single nucleotide polymorphisms (SNPs) identified in GWAS, EWAS, and PHEWAS that were further replicated in extensive studies across the world as being associated with the NAFLD phenotype and the disease severity (NASH and NASH fibrosis), explain only approximately 30–50% estimated heritability of the disease. The effect of each SNP on NAFLD and disease-associated traits is relatively modest (Figure 2).

However, the effect of rs738409 C/G variant located in PNPLA3 (patatin-like phospholipase domain containing 3) on the risk of NAFLD and the disease progression is probably the strongest effects for a common variant modifying the genetic susceptibility of NAFLD and NASH (explaining ~5.3% of the total variance) \(^14\). The evidence indicates that homozygous carriers of 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

The rs58542926 C/T variant located in TM6SF2 (Transmembrane 6 Superfamily Member 2) that was initially associated with liver fat accumulation and aminotransferase levels in a large GWAS study and further replicated in subsequent candidate gene association studies, encodes for a protein involved in lipid metabolism. The rs58542926 is an important modifier of blood lipid traits in different populations. As a challenge for personalized medicine, the C-allele, which has an overall frequency as high as 93%, is associated with higher blood lipids, whereas the T allele confers a moderate risk for NAFLD (carriers of the risk allele present approximately ~2.2% higher lipid fat content).

Likewise, the rs72613567 insertion/deletion variant in HSD17B13 (hydroxysteroid 17-beta-dehydrogenase 13), the functional consequence of which is a splice donor variant of the HSD17B13, presents the mayor protective effect against NAFLD and severe histologic outcomes.

The modest effects on NAFLD risk of the rs780094 in GCKR (glucokinase regulator) — ~OR 1.2 and rs641738 located in TMC4 (transmembrane channel-like 4) exon 1 (p.Gly17Glu) and 500 bases downstream of the MBOAT7 (TMC4/MBOAT7) — ~OR 1.17, are also highlighted in Figure 2.

In addition, the genetic architecture of NAFLD and NASH involves rare variants in other loci, for example, the recently discovered p.P426L loss-of-function variant (rs143545741 C>T) located in autophagy-related 7 (ATG7). Furthermore, a rare nonsense mutation (rs149847328, p.Arg227Ter) in the glucokinase regulator (GCKR) has also been recently reported in an adult patient with NAFLD, morbid obesity, and type 2 diabetes. The p.Arg227Ter was associated with a rapidly progressive histological form of the disease.
Besides, the genetic component of NAFLD and NASH involves mutations in genes of the oxidative phosphorylation (OXPHOS) chain of the mitochondrial DNA (mtDNA) 25, 26, and variants in long noncoding RNAs (lncRNAs), which have a remarkable role in transcriptional and epigenetic regulation 27, 28. Moreover, we reported that deregulated expression of a particular lncRNA, metastasis-associated lung adenocarcinoma transcript 1 (MALAT1), stratifies patients into the histologic phenotypes associated with NAFLD severity 28. MALAT1 up-regulation seems to be a common molecular mechanism in immune-mediated chronic inflammatory liver damage, which suggests that convergent pathophenotypes (inflammation and fibrosis) share similar molecular mediators leading to cancer 28.

**Novel aspects of genetics in NAFLD: Gene variants and interaction effects**

The nonsynonymous rs738409 variant in PNPLA3 is regarded as the major genetic component of NAFLD and NASH 9, 14, 15. The risk effect of this variant on developing fatty liver is the strongest ever reported for a common variant modifying the genetic susceptibility of NAFLD (5 % of the total variance) 14, 15. A recent two-stage (discovery and replication) GWAS that included NAFLD patients characterized by liver biopsy confirmed the rs738409 variant in PNPLA3 as a risk factor for the full histological spectrum in patients of European ancestry 29. Likewise, this large GWAS confirmed important contributions from variants in TM6SF2 (rs58542926) and HSD17B13 (rs72613567) but not MBOAT7 (rs641738) in the disease biology 29.

Like many other complex diseases, NAFLD results from the interaction between genes and environmental factors 5-7. Hence, in addition to individual genetic susceptibility, other important factors contribute to the phenotypic expression of NAFLD and NASH, including dietary patterns and food.

There have been attractive studies focused on gene-diet interaction effect/s, for example, a recent study assessing a gene-diet interaction among rs738409, nutrient intake, and liver
histology severity. Vilar-Gomez et al. showed that PNPLA3 rs738409 G-allele might modulate the effect of specific dietary nutrients on the risk of fibrosis in patients with NAFLD.

Other studies explored gene-gene interaction effects, which are also known as epistasis. For example, Vilar-Gomez et al. found that the protection conferred by HSD17B13 rs72613567 A-allele on severe histological outcomes may be limited to selected subgroups of individuals. Specifically, the protective effects of rs72613567 A-allele on the risk of inflammation and fibrosis seem to be notably stronger in women, persons aged 45 or older, individuals with diabetes, or those with BMI ≥ 35, even after adjusting for the other relevant confounders.

Other human studies have explored the direct effect of the PNPLA3 rs738409 on developing liver fibrosis in relation to liver histologic traits. Specifically, Vilar-Gomez et al. recently reported that a large proportion of the indirect effect of rs738409 on fibrosis severity is mediated through portal inflammation.

Finally, recent studies highlighted the influence of genetic variants, including variants influencing the risk and protection against NAFLD-histological severity (PNPLA3-rs738409, TM6SF2-rs58542926, MBOAT7-rs641738, and HSD17B13-rs72613567) and a variant influencing macronutrient intake (FGF21-rs838133), on the liver microbial DNA composition. For example, Pirola et al. found that members of the Gammaproteobacteria class were significantly enriched in carriers of the rs738409 and rs58542926 risk-alleles, including Enterobacter and Pseudoalteromonas genera, respectively.

GWAS on NAFLD and variants in immune-related loci

The analysis of the GWAS catalog using the EMBL-EBI dataset (EMBL's European Bioinformatics Institute) shows interesting associations between variants in immune-related loci and NAFLD (Table 1). The human major histocompatibility complex on chromosome 6p21 has been associated with susceptibility to many liver diseases. GWAS confirmed the
potential association of many variants in HLA genes and NAFLD and variants in interleukin 36 alpha (IL36A) and beta (IL36B) (Table 1).

To obtain a more comprehensive view of the overlapping between NAFLD and immune system-associated genes, we search the literature with the query "NAFLD" and "immune system" using the web-based platform Genie (available at cbdm.mdc-berlin.de/tools/genie) 34. Using a cutoff of 0.01 for abstracts and a false discovery rate (FDR) <0.01 for genes, we retrieved 941/983 and 975/1524 abstracts/gene, corresponding to NAFLD and the immune system, respectively. Two hundred fifty-eight genes were associated with both NAFLD and the immune system (Figure 3a). In Supplemental Figure 1, we show that some of the 258 overlapping genes are expressed preferentially in cells of the immune systems, for example, MPO (myeloperoxidase), a major component of neutrophil azurophilic granules. In contrast, certain genes, such as C3 (complement C3), SERPINA1 (serpin family A member 1, a serine protease inhibitor), or KART18/19 (keratin 18 and 19, intermediate filament chain keratins), are expressed in different adult tissues, including liver, heart, ovary, lung, or colon (Supplemental Figure 1). Only a few are expressed in any cells, for example, KRT8, HSPD1 (heat shock protein family D member 1) or HSPA5 (heat shock protein family D member 5, encoding a mitochondrial protein which may function as a signaling molecule in the innate immune system).

Both gene groups were significantly enriched in anti-apoptotic, cell communication, and signal transduction biological processes (Figure 3b). As expected, the molecular function characterizing NAFLD-the immune system-shared genes are significantly similar, i.e., ligand-dependent nuclear receptor, chemokine, growth factor, cytokine, and receptor activities (Figure 3c). Finally, Figure 3d, shows shared genes-associated transcription factors (TF). As novel findings, on the one hand, we found BACH1, which encodes a TF that belongs to the Cap’n’collar (CNC) type of basic region leucine zipper factor family (CNC-bZip) associated
with cancer metastasis. On the other hand, we found \textit{NFIC}, whose encoded protein belongs to the \textit{CTF/NF-I} family. These are dimeric DNA-binding proteins that function as cellular TFs and as replication factors for adenoviruses, which also play a role in cancer cell proliferation and metastasis through an epithelial-to-mesenchymal transition process.

Finally, results from a recent study using multicellular liver culture that recapitulates many key features of NAFLD suggested a potential causal link between elevated interleukin 6 (IL6)/STAT3 activity and rs738409-mediated susceptibility to NAFLD. Park et al. showed that dampening IL6-STAT3 activity alleviated the rs738409-G risk allele mediated risk of NAFLD. This effect was attributed to elevated IL6-STAT3 activity in liver cultures carrying the rs738409 G-risk allele that increased NF-kB activity. This finding has clinical implications. For instance, network-based druggability assessment for STAT3, which examines the structure or the protein-protein interaction around the target, suggests that STAT3 is a good drug target presenting a ligand-based druggability score of 97%. In addition, this finding is particularly relevant in the light of the association between NAFLD-predisposing risk factors, including obesity and insulin-resistance, and \textit{STAT3} gene variants.

Interestingly, from the above-described approach of clustering NAFLD and the immune system-associated genes, we retrieved a long list of potential drugs to target the disease (data not shown). Among obvious repurposed drug candidates, such as non-steroid anti-inflammatory drugs, statins, antidiabetic drugs, etc. auranofin emerged. Hwangbo H et al. reported that auranofin, ameliorates the characteristics of NAFLD through the inhibition of NLRP3 inflammasome and Lee et al. recently found that auranofin attenuates hepatic steatosis and fibrosis in NAFLD via NRF2 and NF-kappaB signaling pathways.
Risk alleles in common variants associated with NAFLD/ NASH and gene regulation of immune system: eQTLs

Activation of the immune system, including innate and or adaptive immune response, is an essential driver of the disease severity and progression. While various immune-responsive cells are involved in the pathogenesis of NASH, including T cells and natural killer T cells, classical effectors of NASH-linked inflammation are Kupffer cells and recruited macrophages. In addition, the infiltrated immune cells play several roles in the liver of NASH patients, including the release of cytokines, chemokines, and eicosanoids, among other inflammatory factors.

Analysis of genetic pathways in NASH shows that the immune system is significantly enriched with the sub-pathway "innate immune system" and "cytokine signaling in the immune system". However, much remains to be understood in how risk alleles modify the immune system.

The genomic tools, including GWAS complemented by expression quantitative trait locus (eQTL) analyses, are powerful instruments for understanding how disease-linked variants regulate the expression of quantitative molecular phenotypes across diverse tissues.

GWAS of complex diseases, including NAFLD and NASH, showed that some gene variants are implicated in the susceptibility of multiple traits—a phenomenon known as pleiotropy. This feature involves not only the rs738409 variant in PNPLA3 but also variants in TM6SF2, HSD17B13, and MBOAT7 that are associated with diverse laboratory measurements related to hematological traits.

In addition, the rs738409 has been associated with the soluble intercellular adhesion molecule 1 (sICAM-1) concentration in a large GWAS involving 22,435 healthy women from the Women's Genome Health Study. ICAM-1 is an endothelium and cells of the immune system-derived inflammatory marker. This finding is particularly relevant as previous studies demonstrated that NAFLD is associated with elevated circulating levels of sICAM-1 and
abnormal liver expression of ICAM-1[44]. Furthermore, it was found that liver ICAIM-1 expression levels are significantly correlated with liver lobular inflammatory infiltrate and the severity of necroinflammatory activity[44].

Another important aspect is the exploration of the influence of genetic variation on gene expression across tissues and cell types. For example, Table 2 shows associations of major variants in NAFLD-NASH genes with gene expression levels in non-liver tissues, of which information has been extracted from PhenoScanner, a curated database holding publicly available results from large-scale genome-wide association studies[45, 46].

The rs738409 is associated with adipose tissue and blood expression levels of SAMM50, which plays a crucial role in the maintenance of the structure of mitochondrial cristae, the proper assembly of the mitochondrial respiratory chain complexes, and/or the maintenance of mtDNA[47]. In addition, the rs738409 is associated with whole blood expression levels of FAM89B (Family With Sequence Similarity 89 Member B), which negatively regulates TGFβ-induced signaling—a key factor involved in the regulation of immune response[48].

The rs58542926 in TM6SF2 is associated with blood expression levels of CXCL9 (C-X-C Motif Chemokine Ligand 9)—a member of the chemokine superfamily that encodes secreted proteins involved in immunoregulatory and inflammatory processes, and expression levels of CXCL16 (C-X-C Motif Chemokine Ligand 16), which is involved in several processes, including positive regulation of cell growth, response to interferon-gamma, and response to tumor necrosis factor.

The rs641738 in MBOAT7 is associated with whole blood expression levels of LILRP1 (leukocyte Immunoglobulin-Like Receptor Pseudogene) — also known as leukocyte-expressed receptors of the immunoglobulin superfamily.
Risk alleles in common variants associated with NAFLD/NASH and its relationship with immune system cells types

In the last few years, novel molecular approaches allowed the differentiation between eQTLs in 'bulk' samples of different tissues and “single cells” eQTLs. The difference is that eQTLs from bulk samples represent the average gene expression across all cells in a given tissue. Conversely, eQTLs using single-cell sequencing technology (scRNA-seq) allow the cell-specific gene expression signature (cell type-specific eQTLs).

Although technological advances illuminate the pathophysiology of NAFLD, how the major genetic variants associated with the risk (rs738409) and protection (rs72613567) against NAFLD and NASH affect gene expression of specific immune cells remains largely unknown. To gain insights into this aspect, we explored the DICE database (database of immune cell expression, eQTLs, and epigenomics), which helped to reveal the effects of disease risk-associated genetic polymorphisms on specific immune cell types (https://dice-database.org).

Figure 4 shows differential gene expression of PNPLA3 and HSD17B13 across specific immune cell types. We found very modest levels of PNPLA3 expression in T cell, CD8, naive [activated], and T cell, CD4, naive [activated] (Figure 4a). In addition, we explored the genetic variants directly associated with PNPLA3 gene expression level (SNP located within +/- 1 Mb of the TSS) or eQTLs, and we found three single nucleotide polymorphisms in chromosome 22 influencing T cell, CD4, memory TREG, including rs5766088, rs9626589, and rs9626589. Conversely, we found significant levels of HSD17B13 expression across a variety of immune cells, including B cell, naïve monocyte, classical T cell, CD4, naïve TREG, cell, CD8, naïve T cell, CD4, naïve natural killer (NKO cell, CD56dim CD16+T cell, CD4, TH1/17T cell, CD4, TH1T cell, CD4, TH2T cell, CD4, TFHT cell, CD4, TH1 monocyte, non-classical, and T cell, CD4, memory TREG (Figure 3b). More importantly, these cells are not only relevant effectors of cytotoxicity, but also, these findings are aligned with our previous results on the effect/s of
the splice variant rs72613567 in HSD17B13 on the liver transcriptome. Specifically, we found that the most significant changes in the liver gene expression are enriched by biological pathways related to the immune system, including antigen presentation and interferon-related processes, cytokine signaling, and signal transduction\textsuperscript{19}.

More recent studies on multi-tissue single-cell transcriptomics allowed a broader understanding of the genetic architecture of complex diseases concerning the cross-talk between genetic variants and immune cells\textsuperscript{50}. Dominguez et al. profiled immune cell populations isolated from a wide range of donor-matched tissues, generating nearly 360,000 single cell transcriptomes\textsuperscript{50}. Using data from the study of Dominguez et al.\textsuperscript{50} that can be freely retrieved from the Single Cell Portal, we explored the distribution of PNPLA3, HSD17B13, STAT3, and IL6 expression across tissues and immune cell types using (Figure 5). On the one hand, we observed that expression levels of PNPLA3 are generally very modest across cell types compared to HSD17B13 levels, which present higher levels of expression in innate lymphoid cells, myeloid, mast, and progenitor cells, and megakaryocytes (Figure 5a) — despite that the percentage of gene-expressing cells is relatively low. On the other hand, MBOAT7 presents a relatively high level of expression in myeloid, mast, and progenitor cells with more than 50% of cells expressing the gene (Figure 5a). As expected, STAT3 presents not only very high levels of expression across diverse immune cells in all conditions but also significant levels of expression in the liver tissue (Figure 5b).

Remarkable differences in gene expression across different age groups are also present in Figure 5c, the biological meaning of which remains unknown. However, these differences might explain differences in disease outcomes and sexual dimorphism (Figure 5d).

\textit{Conclusion}
Recent findings based on GWAS, single cells transcriptomics, and analysis of eQTLs may prime future studies that can help to understand the functional basis of shared loci between NAFLD and NASH and immune-mediated mechanisms of the disease severity.

Likewise, while translating GWAS / EWAS / and PHEWAS signals into clinical applications has been slow, genetic knowledge in NAFLD and NASH may significantly improve disease management and monitoring. The accumulated genetic knowledge is now being used to predict disease outcomes and personalized medicine in the field of NAFLD \(^8, 51, 52\) and for repurposing drugs and or select potential actionable targets to treat the disease \(^4, 53, 54\).
REFERENCES


(25) Pirola CJ, Garaycoechea M, Flichman D, Castano GO, Sookoian S. Liver mitochondrial DNA damage and genetic variability of Cytochrome b - a key


(37) Park J, Zhao Y, Zhang F, Zhang S, Kwong AC, Zhang Y, et al. IL6/STAT3 axis dictates the PNPLA3-mediated susceptibility to non-alcoholic fatty liver disease. J Hepatol 2022; S0168-8278(22)030537.

(38) Gianotti TF, Sookoian S, Gemma C, Burgueno AL, Gonzalez CD, Pirola CJ. Study of Genetic Variation in the STAT3 on Obesity and Insulin Resistance in Male Adults. Obesity 2008;16(7):1702-1707.


(53) Kim HY. Recent advances in nonalcoholic fatty liver disease metabolomics. Clin Mol Hepatol 2021;27(4):553-559.


**FIGURE LEGEND**

**Figure 1:** Most relevant and replicated targets associated with the genetic component of NAFLD.

Figure depicts primary protein function and subcellular localization. Information was retrieved from UniProt, a comprehensive and freely accessible resource of protein sequence and functional information available at [https://www.uniprot.org](https://www.uniprot.org).
Figure 2: Summary of variants influencing the risk and protection against NAFLD and the histological disease severity

Figure depicts the most replicated variants associated with NAFLD and NASH, including the global minor allele frequency, the most severe consequence of the variant, the linked variant functionality. In addition, figure highlights the main effect/s on the risk and or protection against NAFLD and NASH.
Figure 3: Overlapping between gene variants associated with NAFLD and the immune system.

This figure provides a functional enrichment analysis of the genes and proteins associated with NAFLD and the immune system, the information of which has been extracted from heterogeneous genomic and proteomic resources using the FunRich program (55, 56).

The literature was searched with the query "NAFLD" and "immune system" by using the genie platform Genie (available at cbdm.mdc-berlin.de/tools/genie/) (34) with a cutoff of 0.01 for abstracts and an FDR <0.01 for genes.

Pathways were ranked according to the $p$ value, whereby $p < 0.05$ was considered statistically significant. % indicates the percentage of altered genes in the whole pathway.

A: Venn diagram of the corresponding gene lists with overlapping genes obtained by the FunRich program, a user-friendly tool, which performs functional enrichment analysis on the generated datasets (available at http://www.funrich.org) (55).

B: Enrichment of genes of both lists in Biological Processes.

C: Enrichment of genes of both lists in Molecular Function.

D: Associated Transcription factors.
Figure 4: Gene expression levels of PNPLA3 and HSD17B13 in immune cells

Differential gene expression of PNPLA3 (a) and HSD17B13 (b) across cell types as calculated by the DESeq package (version 1.6.3). Cells are sorted based on median gene expression from highest to lowest. Squares in the upper diagonal matrix indicate results from pair-wise comparisons of two cell types on the x- and y-axis. Figure shows 2 log2 fold-change. Findings were retrieved from The DICE (Database of Immune Cell Expression, Expression quantitative trait loci (eQTLs) and Epigenomics) project, available at https://dice-database.org/
Figure 5: Analysis of NAFLD/NASH-risk alleles and cross-tissue immune cell expression

Information was retrieved from Single cell portal available at https://singlecell.broadinstitute.org/single_cell/study/SCP1845/ (50). Panels depict annotation of cell population type (A), organ/tissue distribution (B), age (C), and sex of donors (D).

Scaling is relative to each gene's expression across all cells in a given annotation selection, i.e., cells associated with each column label in the dot plot. Gene targets were arbitrarily selected, including major NAFLD/NASH related loci and two immune-related genes (IL6 and STAT3).

ILCT: innate lymphoid cells; pDCs: plasmacytoid dendritic cells, which are a unique subset of dendritic cells specialized in secreting high levels of type I interferons; myeloid: myeloid cells are granulocytic and phagocytic leukocytes that traverse blood and solid tissues mast: mast cells are immune cells of the myeloid lineage; B: B lymphocytes, also called B cells; mast: mast cells are immune cells of the myeloid lineage; progenitor: the common B- and T-cell progenitor can be found in the bone marrow; ery: erythroid cells; mgk: megakaryocytes/platelets; MNP/RT doublets cells: mononuclear phagocytes; B/T doublets: B and T cells stuck together as a "doublet".
### Table 1: GWAS catalog and associations between variants in immune-related loci and NAFLD susceptibility

<table>
<thead>
<tr>
<th>Mapped loci</th>
<th>Variant ID</th>
<th>P value</th>
<th>Study accession</th>
<th>Chromosome location</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>IL36A, IL36B</em></td>
<td>rs28946269</td>
<td>9 x 10^{-6}</td>
<td>GCST008468</td>
<td>2:113011237</td>
</tr>
<tr>
<td><em>HLA-DRB5, HLA-DRB9</em></td>
<td>rs7748270</td>
<td>4 x 10^{-12}</td>
<td>GCST90094908</td>
<td>6:32480822</td>
</tr>
<tr>
<td><em>HLA-DRB1, HLA-DQA1</em></td>
<td>rs5021727</td>
<td>3 x 10^{-9}</td>
<td>GCST90094908</td>
<td>6:32610856</td>
</tr>
<tr>
<td><em>HLA-DQA1, HLA-DRB1</em></td>
<td>rs9271325</td>
<td>2 x 10^{-8}</td>
<td>GCST90094908</td>
<td>6:32614736</td>
</tr>
<tr>
<td><em>HLA-DQA1, HLA-DRB1</em></td>
<td>rs9271406</td>
<td>2 x 10^{-8}</td>
<td>GCST90094908</td>
<td>6:32619811</td>
</tr>
<tr>
<td><em>HLA-DQA1</em></td>
<td>rs2213287</td>
<td>5 x 10^{-9}</td>
<td>GCST90094908</td>
<td>6:32637731</td>
</tr>
<tr>
<td><em>HLA-DQA1</em></td>
<td>rs9272699</td>
<td>2 x 10^{-8}</td>
<td>GCST90094908</td>
<td>6:32641452</td>
</tr>
</tbody>
</table>

Source: [https://www.ebi.ac.uk/](https://www.ebi.ac.uk/) EMBL’s European Bioinformatics Institute
Table 2: Variants in NAFLD-NASH associated genes and regulation of gene expression in immune related loci

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Gene expression</th>
<th>Sample size</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP ID: rs738409  hg19_coordinates: chr22:44324727 (PNPLA3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adipose subcutaneous</td>
<td>SAMM50</td>
<td>385</td>
<td>9.60E-07</td>
</tr>
<tr>
<td>Whole blood</td>
<td>SAMM50</td>
<td>5257</td>
<td>5.67E-106</td>
</tr>
<tr>
<td>Whole blood</td>
<td>SAMM50</td>
<td>31300</td>
<td>3.15E-18</td>
</tr>
<tr>
<td>Whole blood</td>
<td>FAM89B</td>
<td>31300</td>
<td>5.49E-06</td>
</tr>
<tr>
<td>SNP ID: rs58542926  hg19_coordinates: chr19:19379549 (TM6SF2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole blood</td>
<td>CXCL9</td>
<td>5257</td>
<td>3.13E-06</td>
</tr>
<tr>
<td>Whole blood</td>
<td>CXCL16</td>
<td>28533</td>
<td>5.71E-06</td>
</tr>
<tr>
<td>SNP ID: rs641738  hg19_coordinates: chr19:54676763 (TMC4/MBOAT7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole blood</td>
<td>LILRP1</td>
<td>5417</td>
<td>8.87E-06</td>
</tr>
</tbody>
</table>

**Supplemental Figure 1.** Tissue-specific expression pattern of the 258 overlapping genes associated with both NAFLD and the immune system from figure 3 as described in the main text. The heat plot was also obtained by the FunRich program described in the legend of Figure 3 (55).