Genetic factors associated with the presence and progression of nonalcoholic fatty liver disease: A narrative review

Ruben Hernaez\textsuperscript{a,b,*}

\textsuperscript{a} Department of Medicine, The Johns Hopkins School of Medicine, Baltimore, MD 21287, USA
\textsuperscript{b} Department of Medicine, Georgetown University Hospital/Washington Hospital Center, Washington, DC 20010, USA

Received 28 July 2011; accepted 4 August 2011
Available online 16 November 2011

Abstract Nonalcoholic fatty liver disease (NAFLD) is the most common chronic liver disease in the world. Whereas insulin resistance and obesity are considered major risk factors for the development and progression of NAFLD, the genetic underpinnings are unclear. Before 2008, candidate gene studies based on prior knowledge of pathophysiology of fatty liver yielded conflicting results. In 2008, Romeo et al. published the first genome wide association study and reported the strongest genetic signal for the presence of fatty liver (\textit{PNPLA3}, patatin-like phospholipase domain containing 3; rs738409). Since then, two additional genome wide scans were published and identified 9 additional genetic variants. Whereas these results shed light into the understanding of the genetics of NAFLD, most of associations have not been replicated in independent samples and, therefore, remain undetermined the significance of these findings. This review aims to summarize the understanding of genetic epidemiology of NAFLD and highlights the gaps in knowledge.

© 2011 Elsevier España, S.L. All rights reserved.

Palabras clave

Estudios genéticos de asociación; Enfermedad grasa del hígado no alcohólica; Hígado graso; Progresión de la enfermedad; Prevalencia; Revisión

Factores genéticos asociados con la presencia y progresión de la enfermedad hepática no alcohólica: Una revisión narrativa

Resumen La enfermedad hepática no alcohólica (EHNA) es la enfermedad crónica del hígado más común en el mundo. Si bien se considera que la resistencia a la insulina y la obesidad son factores de riesgo importantes para el desarrollo y progresión de la misma, sus bases genéticas no están claras. Antes de 2008, los estudios sobre genes candidatos basados en conocimientos previos de la fisiopatología de la esteatosis hepática produjeron resultados contradictorios. En 2008, Romeo et al. publicaron el primer estudio de asociación amplia del genoma que presentaba indicios genéticos sólidos de la presencia de esteatosis hepática (\textit{PNPLA3}, dominio de la fosfolipasa-patatina 3; rs738409). Desde entonces, se han publicado dos estudios adicionales de asociación amplia del genoma en los que se identificaron otras nueve variantes genéticas. Si bien estos resultados arrojan luz sobre la genética de la EHNA, la mayoría de las asociaciones...
Nonalcoholic fatty liver disease is common and has a genetic basis

Nonalcoholic fatty liver disease (NAFLD), the most common chronic liver disease, represents a wide spectrum of disease characterized by the presence hepatic steatosis in the absence of significant alcohol consumption or other causes of liver disease. The pathogenic processes leading to steatosis, steatohepatitis (NASH) and fibrosis are multifactorial and involve both environmental and genetic factors. Obesity and type 2 diabetes/insulin resistance are the most common risk factors for the development and progression of NAFLD. Since both obesity and type 2 diabetes continue to rise, it is expected that NAFLD will reach epidemic proportions worldwide. The prevalence of ultrasound-defined NAFLD ranges from 14 to 45%. Extrapolating data from small studies and the current prevalence of obesity and type 2 diabetes, NAFLD was recently estimated to affect more than 30 million people in the U.S. In Spain, Caballeria et al. showed in a recent cross-sectional study a prevalence of 26% in individuals aged 15 and 85 years randomly selected from 25 primary healthcare centers, consistent with prior American estimates.

A genetic underpinning for NAFLD has been suggested by familial aggregation studies, heritability studies, candidate gene studies, genome-wide scans and expression studies. The presence of a genetic basis in NAFLD not only will shed light in the identification of individuals at risk to develop NAFLD and its progression, but also the dissection of NAFLD pathogenesis and the development of new therapies.

The ultimate goal of this review is to provide the big picture of the current understanding of genes associated with NAFLD so the reader is able to understand what the gaps are in the genetic epidemiology of NAFLD. This paper, however, will not assess other forms of fatty liver disease, expression studies or animal studies.

This work is structured in three parts: first, the author will provide some background concepts in the genetic epidemiology of NAFLD and the current understanding of NAFLD pathophysiology; second, the reviewer will highlight the most up-to-date findings in candidate gene studies and genome wide association studies (GWAS) in humans; the review will conclude with suggestions to conduct candidate gene studies. Compared to other reviews, the present work will not elaborate the mechanism for each specific gene but will provide references for the reader for a more careful evaluation.

Study designs and methodological problems in genetic epidemiology of NAFLD

NAFLD is considered a complex disease because it does have a genetic component but with no simple Mendelian pattern of single-gene inheritance such as Wilson disease. In NAFLD, multiple genes, polygenes, environmental factors, age effects, and their interactions, may be involved.

Genetic epidemiology studies can be divided into two broad categories: (a) according to the relatedness amongst participants (family-based versus non-related or population-based); and (b) according to the knowledge of the genetic marker(s) used (hypothesis-driven or candidate gene study versus hypothesis free or genome-wide association study). A candidate gene is defined as gene whose product is considered to play a role in disease pathogenesis. It is important to bear in mind that all studies provide key information and are not mutually exclusive. For instance, family-based studies are typically the first step to determine if a condition is genetic in nature, the model of inheritance, and to understand the proportion of phenotypic variance due to shared genetic factors (heritability). Family-based studies may use linkage analysis to identify the location of a major gene or, association analyses such as the transmission-disequilibrium test, candidate gene approach and a genome-wide association analysis.

Population-based studies, on the other hand, is the mainstream of studies in NAFLD and, until recently, candidate gene studies were the most common design using the "two hit" hypothesis as a framework. Both candidate gene and GWAS rely on the statistical association of genetic markers and the disease (phenotype) of interest. The most commonly used genetic markers are the single nucleotide polymorphisms or SNPs, defined as the change in a single base pair of the DNA common in more than 1% of the population.

The major problem in both candidate gene and GWAS is the presence of false negative and false positive associations. Lack of statistical power leads to false negatives; that is, true associations are not detected in the study. So the researcher needs to estimate the number of participants to be recruited to circumvent such error using power calculators. In GWAS, where there is no hypothesis behind the statistical analysis, the problem is the opposite and SNPs are spuriously associated due to chance with the disease, even after statistical correction for multiple comparisons. Consequently, most of the journals extremely recommend authors to replicate their findings in a different...
Triglycerides are neutral lipids consisting of a glycerol backbone and three long-chain fatty acids. The major routes of free fatty acids (FFAs) are (a) dietary intake; (b) lipolysis from fat reservoirs, and (c) de novo lipogenesis. Animal studies and human inherited diseases have shown increased hepatic steatosis with increased triglyceride dietary intake, increased lipolysis from peripheral tissues, increased de novo lipogenesis, or prevention of triglyceride removal from the liver (i.e. decreased efflux or oxidation). The exact contribution of each pathway for the development of hepatic steatosis has been studied by Donnelly et al. who showed that 59% of hepatic fat derived from circulating FFAs (mainly lipolysis), 26% from de novo lipogenesis and 15% from diet.

Insulin resistance is the key etiopathogenic factor in the development of fatty liver. Whether insulin resistance is a cause of hepatic steatosis or vice versa is a matter of debate and has been reviewed recently. Current evidence suggest that insulin resistance comes first and leads to hepatic steatosis, as shown by animal models and human diseases where hepatic steatosis is not associated with insulin resistance and, conversely, human diseases with inherited insulin resistance leading to hepatic steatosis (e.g. AKT2 mutation). Opposite views, however, suggest that hepatic steatosis, possibly originated by impaired mitochondrial β-oxidation of fatty acids, leads to hepatic insulin resistance which, in turn, increases de novo fatty acid synthesis, and impairs glucogenogenesis, hepatic glucose uptake, and transport of lipoproteins perpetuating increase peripheral insulin resistance.

In 1998, Day and James proposed a 'two-hit model' for the development of steatohepatitis in which the first step was the accumulation of free fatty acids in the liver, and, from there, further insults induced inflammation, fibrosis and, eventually, cirrhosis. Nevertheless, Dr. Day’s group updated their own model and suggested that 'steatosis may be an early adaptive response to hepatocyte stress through which potentially lipotoxic FFAs are partitioned into relatively stable intracellular triglyceride stores’. NAFLD is seen now “as a combination of effects of several fundamental biochemical and immunologic mechanisms of liver injury rather than adhering to a sequential 'two-hit' paradigm” and is being supported by other experts in the field. In Day et al.’s view, “insulin resistance promotes increased hepatic FFA flux (lipolysis, diet and lipogenesis) leading to hepatic steatosis, and is driven by (1) direct hepatocyte lipotoxicity, (2) hepatocellular oxidative stress secondary to free radicals produced during beta- and ω-FFA oxidation, (3) endotoxin/TLR4-induced inflammation, (4) cytokine release, and (5) endoplasmic reticulum (ER) stress”, leading, at the end, to inflammation, cellular damage, and progression to cirrhosis.

### Genetic epidemiology of NAFLD: a systematic approach

To identify most of the available published literature in genetics of NAFLD, the author performed a systematic search using PubMed by applying the following search engine: (polymorphism OR SNP) AND ("steatohepatitis" OR "steatosis" OR NAFLD OR NASH) without language restriction and until July 26th, 2011. The search yielded 224 references; of which 60 corresponded to candidate gene studies and are summarized in Table 1 and three genome wide association studies in Table 2. This search engine is limited by two major factors: on the one hand, it did not include EMBASE database (more sensitive for non-English papers and scientific meetings), and, on the other hand, there is still a risk by publication bias (negative studies are not published).

Genes (or nearby genes) are named using the HUGO Gene Nomenclature Committee (HGNC) and reported identification.

### Candidate genes studies in human NAFLD: hepatic lipid metabolism, insulin resistance and oxidative stress

Numerous candidate gene studies, applying this 'multiple-hit' hypothesis, have studied the effects of genes on NAFLD presence and progression. The reviewer has simplified the prior hypothesis into the following mechanisms (not mutually exclusive): genes involved in hepatic lipid metabolism (synthesis, storage, export, oxidation), genes implicated in insulin signaling (insulin resistance), and, finally, genes involved in oxidative stress and inflammation (and therefore, most likely involved in progression to cirrhosis (Table 1).

The reviewer included, when appropriate, other data from different phenotypes to illustrate the direction of association for a particular gene. The number of symbols represent how many studies have been done for a particular gene.

### Genome wide association studies in NAFLD

Romeo et al. were the first to apply the GWAS method using a phenotype based on magnetic resonance spectroscopy. They studied 9,299 nonsynonymous sequence variations and identified a missense mutation (Ile148→Met148 (1148M)) in patatin-like phospholipase domain-containing (PNPLA) 3 gene (PNPLA3; HGNC: 18590). PNPLA3, highly expressed in adipose tissue and liver, is regulated by insulin through a signaling cascade that includes LXR and SREBP-1c and, therefore, increased with feeding in animal studies. This mutation is, by far, the strongest genetic signal up to date and showed increased odds for fatty liver of 3.26 in the original report. In addition, the PNPLA3 gene could also be responsible for the difference in prevalence of fatty liver disease between ethnic groups. For instance, Mexican-Americans have more prevalence of the high risk allele, whereas African Americans, where fatty liver is known to be less frequent, had a protective variation for such. Future studies confirmed the association between this gene and
<table>
<thead>
<tr>
<th>Gene (HGNC)</th>
<th>SNP</th>
<th>Effect on Steatosis</th>
<th>Inflammation/NASH</th>
<th>Fibrosis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hepatic lipid metabolism</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEMT (8830)</td>
<td>rs7946</td>
<td>↑∅↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MTTP (7467)</td>
<td>rs1800591</td>
<td>∅∅</td>
<td>∅↑↑↑↑</td>
<td></td>
</tr>
<tr>
<td>MTTP^37,45</td>
<td>rs3816873</td>
<td>↑∅</td>
<td></td>
<td></td>
</tr>
<tr>
<td>APOC3 (610)</td>
<td>rs61733139</td>
<td>∅</td>
<td></td>
<td></td>
</tr>
<tr>
<td>APOC3</td>
<td>rs2854116</td>
<td>↑∅</td>
<td></td>
<td></td>
</tr>
<tr>
<td>APOC3</td>
<td>rs2854117</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NR1I2 (7968)</td>
<td>rs7643645</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FABP2 (2556)</td>
<td>rs1799883</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DGAT (2843)</td>
<td>rs1944438</td>
<td>∅</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACSL4 (3571)</td>
<td>rs7887981</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADRB3 (288)</td>
<td>rs4994</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADRB2 (286)</td>
<td>rs1042714</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIPC (6619)</td>
<td>rs1800588</td>
<td>↑↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>APOE (613)</td>
<td>N/A</td>
<td>↑↓</td>
<td>∅</td>
<td></td>
</tr>
<tr>
<td>CLOCK (2082)</td>
<td>rs11932595</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs6843722</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs554483</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs6843722</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs6850524</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs1554483</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Insulin sensitivity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENPP1 (3356)</td>
<td>rs1044498</td>
<td>∅∅</td>
<td>∅∅</td>
<td></td>
</tr>
<tr>
<td>IRS1 (6125)</td>
<td>rs1801278</td>
<td>∅</td>
<td>∅</td>
<td>∅</td>
</tr>
<tr>
<td>ADIPOQ (13633)</td>
<td>rs2241766</td>
<td>∅∅∅↑↑</td>
<td>↑</td>
<td>∅</td>
</tr>
<tr>
<td>ADIPOQ^64-67</td>
<td>rs1501299</td>
<td>↓∅∅↑</td>
<td></td>
<td>∅↑∅</td>
</tr>
<tr>
<td>ADIPOR1 (24040)</td>
<td>rs6666089</td>
<td>∅</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADIPOR2 (24041)</td>
<td>rs767870</td>
<td>↓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPARA (9232)</td>
<td>rs1800206</td>
<td>↑∅</td>
<td></td>
<td>∅</td>
</tr>
<tr>
<td>PPARG (9236)</td>
<td>rs3856806</td>
<td></td>
<td></td>
<td>∅</td>
</tr>
<tr>
<td>PPARA^68,70,71</td>
<td>rs1801282</td>
<td>∅∅∅</td>
<td>∅∅</td>
<td>∅∅</td>
</tr>
<tr>
<td>PPARC1A (9237)</td>
<td>rs8192687</td>
<td></td>
<td></td>
<td>∅</td>
</tr>
<tr>
<td>PPARC1A</td>
<td>rs7943146</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCKR (4196)</td>
<td>rs780094</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4H (6932)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPINK-1 (11244)</td>
<td>N/A</td>
<td></td>
<td></td>
<td>∅</td>
</tr>
</tbody>
</table>

Table 1 Candidate gene studies in nonalcoholic fatty liver disease (last updated 7/27/11).
<table>
<thead>
<tr>
<th>Gene (HGNC)</th>
<th>SNP</th>
<th>Effect on steatosis</th>
<th>Inflammation/NASH</th>
<th>Fibrosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEPR (6554),76 leptin receptor gene</td>
<td>N/A</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEP (6553),77 leptin</td>
<td>rs7799039</td>
<td>↓</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Genes influencing generation of reactive oxidant species, or cytokine genes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNF (11892),36,67,77,80 tumor necrosis factor-alpha</td>
<td>rs180062</td>
<td>↓</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>Tumor necrosis factor-alpha67,79,80</td>
<td>rs361525</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor necrosis factor-alpha67,79</td>
<td>rs1800630</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor necrosis factor-alpha79</td>
<td>rs179964</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor necrosis factor-alpha79</td>
<td>rs1799724</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor necrosis factor-beta79</td>
<td>rs909253</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TNFSF10 (11925)83 tumor necrosis factor (ligand) superfamily, member 10</td>
<td>N/A</td>
<td>↓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IL-6 (6018)40</td>
<td>rs1800795</td>
<td>↑</td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td>CD14 (1628)86</td>
<td>rs2569190</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCLC (4311),43,42 glutamate-cysteine ligase, catalytic subunit</td>
<td>rs4140528</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOD2 (11180),43,44,88 superoxide dismutase 2, mitochondrial</td>
<td>rs4880</td>
<td>↑</td>
<td></td>
<td>↑</td>
</tr>
<tr>
<td>HFE (4886),89 hemochromatosis</td>
<td>rs1800562</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UGT1A1 (12530),90 UDP glucuronosyltransferase 1 family, polypeptide A1</td>
<td>rs1799945</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCP (12519),91,92 uncoupling protein 3 (mitochondrial, proton carrier)</td>
<td>N/A</td>
<td>↓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTGS2 (9506),93 prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)</td>
<td>rs689466</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCB11 (42),94 ATP-binding cassette, subfamily B, member 11</td>
<td>rs2287622</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIF (797),95 macrophage migration inhibitory factor (glycosylation-inhibiting factor), CYP2E1 (2631),96,97 cytochrome P450, family 2, subfamily E, polypeptide 1</td>
<td>rs755622</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SERPINA1 (8941),98 serpin peptidase inhibitor, clade A (alpha-1 antiprotease, antitrypsin), member 1</td>
<td>rs2031920</td>
<td></td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>MTHFR (7436),99 methylenetetrahydrofolate reductase</td>
<td>rs28929474</td>
<td>↑</td>
<td></td>
<td>↑</td>
</tr>
<tr>
<td>IL1B (5992),53 interleukin 1, beta</td>
<td>rs16944</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TLR4 (11850),86 toll-like receptor 4</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFTR/MRP (53),101 ATP-binding cassette, sub-family C, member 2</td>
<td>rs17222723</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>STAT3 (11364),102 signal transducer and activator of transcription 3 acute-phase response factor</td>
<td>rs8187710</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HP (5141),103 haptoglobin</td>
<td>N/A</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGTR1 (336),104 angiotensin II receptor, type 1</td>
<td>rs3772622</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KLF6 (2235),105 Kruppel-like factor 6</td>
<td>rs3750861</td>
<td>↑</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Abbreviations:** N/A: not available; ↑: increase; ∅: no association; ↓: decrease (protective).
the presence of fatty liver disease, including GWAS with liver enzymes\textsuperscript{17} and multiple case-control studies.\textsuperscript{15,16}

Chalasani et al. described in 236 white female biopsy-proven NAFLD patients' five new genetic variants associated with inflammation and fibrosis. They found that the NAFLD activity score (NAS), a pathological tool to measure changes in NAFLD during clinical trials, was associated with the gene \textit{FDFT1} (farnesyl diphosphate farnesyl transferase 1, HGNC: 3629); in addition, they found an association with lobular inflammation for the collagen gene \textit{COL13A1} (collagen, type XIII, alpha 1, HGNC: 2190), a SNP nearby the \textit{PDGFA} gene (platelet-derived growth factor alpha polypeptide, HGNC: 8799); the \textit{LTBP3} (latent transforming growth factor beta binding protein 3, HGNC: 6716), and the \textit{EFCAB4B} (EF-hand calcium binding domain 4B).

The Genetics of Obesity-related Liver Disease (GOLD) Consortium is the last genome wide scan published up to date. We obtained the same association with \textit{PNPLA3} and four additional genetic variants including the \textit{PNPLA3} (HGNC: 18590), namely, \textit{NCAN} (neurocan, HGNC: 2465), \textit{LYPLAL1} (lysophospholipase-like 1, HGNC: 20440); \textit{GCKR} (glucokinase regulatory protein, HGNC: 4196); and the \textit{PPP1R3B} (protein phosphatase 1, regulatory subunit 3b, HGNC: 14942). \textit{GCKR} and \textit{PPP1R3B} are key enzymes in de novo lipogenesis from glucose; \textit{LYPLAL1}-related protein has been predicted to play a crucial part in consecutive steps in triglyceride breakdown. The role of \textit{NCAN}, however, remains to be determined. Interestingly, we found that \textit{PNPLA3} and our other genes had a modest role for lipid metabolism suggesting that these genes, if they are involved in lipid metabolism, exert their effects within the liver through different mechanistic pathways than the observed ones by conventional laboratory lipid measurement.

There is no replication up to date of the aforementioned GWAS findings except for the \textit{PNPLA3} (rs738409). For this gene variant, a recent meta-analysis updated until January 2011 found an association with the presence of fat accumulation (GG homozygous showed 73% higher lipid fat content when compared with CC ones); 3.2-fold greater risk of higher necroinflammatory scores, 3.5-fold risk of NASH, and 3.2-fold greater risk of developing fibrosis when compared with CC homozygous.\textsuperscript{15}

\begin{table}[h]
\centering
\begin{tabular}{lll}
\hline
Gene (HGNC) & SNP & Effect on steatosis & Inflammation/NASH Fibrosis \\
\hline
\textit{PNPLA3} (18590), patatin-like phospholipase domain containing 3\textsuperscript{15,16} & rs738409 & ↑↑ & \\
\textit{PNPLA3} (18590), patatin-like phospholipase domain containing 3\textsuperscript{15} & rs6006460 & ↓ & \\
\textit{FDFT1} (3629),\textsuperscript{14} farnesyl diphosphate farnesyl transferase 1 & rs264524 & ↑ & \\
\textit{COL13A1} (2190),\textsuperscript{14} collagen, type XIII, alpha 1 & rs1227756 & ↑ & \\
\textit{PDGFA} (8799),\textsuperscript{14} platelet-derived growth factor alpha polypeptide & rs343064 & ↑ & \\
\textit{LTBP3} (6716),\textsuperscript{14} latent transforming growth factor beta binding protein 3 & rs1227756 & ↑ & \\
\textit{EFCAB4B} (28657),\textsuperscript{14} EF-hand calcium binding domain 4B & rs887304 & ↑ & \\
\textit{NCAN} (2465),\textsuperscript{16} neurocan & rs2228603 & ↑ & \\
\textit{LYPLAL1} (20440),\textsuperscript{16} lysophospholipase-like 1 & rs1213785 & ↑ & \\
\textit{GCKR} (4196),\textsuperscript{16} glucokinase regulatory protein & rs780094 & ↑ & \\
\textit{PPP1R3B} (14942),\textsuperscript{16} protein phosphatase 1, regulatory subunit 3b & rs4240624 & ↑ & \\
\hline
\end{tabular}
\caption{Genome-wide association studies in NAFLD (last updated 7/27/11).}

\textit{Abbreviations}: N/A: not available; ↑: increase; $\phi$: no association; ↓: decrease (protective).
\end{table}

Future directions: candidate-genes are needed

The typical statement “more research is needed” is clearly shown by the examination of both tables. To better understand the genetic determinants of NAFLD, it is key to replicate prior studies. Whereas GWAS are expensive, require thousands of individuals and strong statistical and population genetics knowledge, the use of candidate gene studies in NAFLD is easy to implement and straightforward. The major caveat for these studies is the lack of power. For instance, assuming a genetic risk ratio of 1.5 for a given polymorphism (high for most of the replicated genetic studies), and a allele frequency around 20%, more than 400 patients and controls are required to give a study 90% power to detect a significant effect at the 5% level.\textsuperscript{27} Day\textsuperscript{27} and Bataller\textsuperscript{23} provide two outstanding reviews to guide the reader in the design of high quality candidate gene studies, step-by-step.

On the other hand, if the researcher is willing to invest his/her time to study new candidate genes, then Day proposes to find them by reviewing: (1) Gene product considered to play a role in the disease; (2) gene is known to be mutated in a familial form of the disease; (3) gene knockout/overexpression in animal models influences disease development; (4) gene lies in a chromosomal region associated with disease in a linkage study; (5) gene expression is altered in microarray studies of tissue from patients with disease; (6) gene is identified in a
phenotype-driven mouse mutagenesis study; and (7) gene identify in genome-wide association studies.

In conclusion, genes play a role in the development and progression of NAFLD; PNPLA3 is the strongest signal up to date but there are other numerous genes that have been described but not formally replicated. The future of genetic epidemiology will require replication and, ultimately, expression studies and animal models to know the molecular role of that particular genetic variant. The understanding of genetic determinants of NAFLD will help to identify individuals at risk and, potentially, new therapies to treat the most common chronic liver disease in the world.

Conflict of interest
The author declare no conflict of interest.

Acknowledgement
The author have been supported by the American Diabetes Association Mentor-Based Program (7-07-MN-08, PI: Dr. Frederick L. Brancati).

References
33. HUGO Gene Nomenclature Committee at the European Bioinformatics Institute. HUGO Gene Nomenclature.


