GENETIC MODIFIERS OF NON-ALCOHOLIC FATTY LIVER DISEASE PROGRESSION

Quentin M. Anstee, Ann K. Daly & Christopher P. Day.

Liver Research Group, Institute of Cellular Medicine, The Medical School, Newcastle University, Framlington Place, Newcastle-upon-Tyne, NE2 4HH, U.K.

Keywords: NAFLD, NASH, Steatohepatitis, Gene, Polymorphism

Corresponding Author/Reprint Requests

Professor Christopher P. Day
Pro-Vice Chancellor & Provost of Medical Sciences,
Medical Sciences Faculty Office,
The Medical School,
Newcastle University,
Framlington Place,
Newcastle-upon-Tyne, NE2 4HH, Great Britain.
Telephone: +44 (0) 191 222 7003
Fax: +44 (0) 191 222 6621
Email: c.p.day@ncl.ac.uk

Acknowledgements

QMA is the recipient of a Clinical Senior Lectureship Award from the Higher Education Funding Council for England (HEFCE). QMA, AKD & CPD are recipients of an MRC Programme Grant to identify and study novel genetic modifiers of liver disease and are members of the FLIP research consortium funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement Health-F2-2009-241762.
ABSTRACT

Non-alcoholic fatty liver disease (NAFLD) is now recognised as the most common cause of liver dysfunction worldwide. However, whilst the majority of individuals who exhibit features of the metabolic syndrome including obesity and insulin resistance will develop steatosis, only a minority progress to steatohepatitis, fibrosis and cirrhosis. Subtle inter-patient genetic variations and environment interact to determine disease phenotype and influence progression. A decade after the sequencing of the human genome, the comprehensive study of genomic variation offers new insights into the modifier genes, pathogenic mechanisms and is beginning to suggest novel therapeutic targets. We review the current status of the field with particular focus on advances from recent genome-wide association studies.

111 words (150 words max.)
1. INTRODUCTION
Non-alcoholic fatty liver disease (NAFLD) represents a spectrum of liver disease encompassing simple fatty infiltration of the liver parenchyma (steatosis), fat and inflammation (non-alcoholic steatohepatitis; NASH) and cirrhosis, in the absence of excessive alcohol consumption (typically a threshold of <20g/day for women and <30g/day for men is adopted)(1, 2)..< Population studies show that NAFLD is strongly associated with obesity, insulin resistance/type II diabetes mellitus and dyslipidaemia and it is considered by many to be the hepatic manifestation of the metabolic syndrome(3-5). It is now recognised that NAFLD is the most common cause of liver dysfunction in developed countries(1). Estimates vary between populations however one large European study found NAFLD present in 94% of obese patients (BMI >30 kg/m$^2$), 67% of overweight patients (BMI >25 kg/m$^2$), and 25% of normal weight patients(6). The overall prevalence of NAFLD in type 2 diabetics ranges from 40-70%(6). As only a minority of patients with NAFLD progress to more advanced disease characterised by inflammation, fibrosis, cirrhosis and hepatocellular carcinoma (HCC), NAFLD is best considered a complex disease trait where subtle inter-patient genetic variations and environment interact to determine disease phenotype and progression(7, 8). Here we will review the current understanding of genetic modifiers of NAFLD/NASH with particular focus on data from human studies (Figure 1).

2. PATHOGENESIS
Our current understanding of disease pathogenesis has been achieved through clinic based research and the translational study of specific animal models(9, 10). The initiating events in NAFLD/NASH relate to the development of obesity and insulin resistance. Together, these promote hepatic free fatty acid (FFA) flux which provides the appropriate milieu for NAFLD/NASH to develop. Importantly, the visible steatosis that has for some time been considered the ‘first hit’ in the pathogenesis of NAFLD/NASH is now recognised by many investigators to be an epiphenomenon reflecting these changes in hepatocyte FFA flux and associated cellular stress responses. Recognition of this means that steatosis should now be considered an early adaptive response to hepatocyte stress through which potentially lipotoxic FFAs are partitioned into relatively stable intracellular triglyceride stores. This was elegantly demonstrated by silencing of hepatic DGAT2 expression, a key enzyme mediating the conversion of FFA to triglyceride(11). Rather than ameliorating steatohepatitis, the consequent reduction in hepatocyte triglyceride synthesis was associated with a greater level of fatty acid oxidation, particularly through Cyp2E1, leading to greater oxidative stress, cellular damage and higher serum transaminase levels(11). Discussion of pathogenesis and
the genetic modifiers of progressive NAFLD should now consider the combined effects of several fundamental biochemical and immunological mechanisms of liver injury rather than adhering to a sequential ‘two-hit’ paradigm. These effects include: (1) Direct hepatocyte lipotoxicity; (2) Hepatocellular oxidative stress secondary to free radicals produced during β-and ω-FFA oxidation; (3) Endotoxin/TLR4 induced Kupffer cell cytokine release; (4) Cytokine release; and (5) Endoplasmic reticulum (ER) stress. Consequent cellular damage triggers a mixture of immune mediated hepatocellular injury and both necrotic and apoptotic cell death pathways[12-14]. If these persist for some time, these processes lead to stellate cell activation, collagen deposition and hepatic fibrosis[15]. In depth discussion of these mechanisms falls outside the scope of this review, however this framework will be adopted in our discussion of genetic modifiers.

3. TECHNIQUES FOR INVESTIGATING THE GENETIC BASIS OF NAFLD/NASH

Until recently, the primary approach to the genetic study of complex disease traits was through case-control disease-association studies in man where candidate genes were selected on the basis of a putative role of their encoded proteins in disease pathogenesis. Being reliant on an a priori hypotheses for gene selection, these studies were limited to known candidates and therefore were not able to explore the potential role of other less obvious genes that may have an equal or greater influence on disease susceptibility. In the ten years since the publication of the draft human genome there have been significant advances in our understanding of genomic variation. Several million single nucleotide polymorphisms (SNPs) have been described across individuals from diverse ethnic backgrounds[16-18]. This has paved the way for the development of SNP genotyping arrays and genome-wide association studies (GWAS) that have allowed the majority of common (minor allele frequency >5%) variability in the human genome to be simultaneously surveyed. The availability of high throughput next-generation sequencing offers the prospect of even more comprehensive characterisation of genomic variability, particularly the more rare genetic variants that are not effectively identified by GWAS but which may have a relatively large effect on disease risk. Greater discussion of these techniques falls outside the scope of this review but they are discussed more fully elsewhere [19-21].

The utility of the GWAS approach was demonstrated by one of the earliest reports describing simultaneous study of seven different complex diseases[22]. Since then, the number of GWAS studies in the literature has expanded exponentially. Amongst these, several liver-related diseases and traits have been studied. Factors that influence variation in biochemical liver function tests[23], drug induced liver injury[24, 25], gallstone disease[26],
primary biliary cirrhosis[27-29], NAFLD[30-32], hepatitis B persistence[33] and hepatitis C treatment response[34, 35] have been identified. A welcome consequence of wider adoption of non-hypothesis driven GWAS techniques is that the loci identified are frequently novel and would not previously have been implicated in disease pathogenesis. However, as often neither biological function nor pathogenic mechanisms are known, such associations require further detailed study both to determine activity and to validate causality. In addition, candidate gene association studies have examined modifiers of disease progression and fibrosis[36, 37].

4. GENOME WIDE ASSOCIATION STUDIES IN NAFLD/NASH RESEARCH

To date, three GWAS scale studies have been reported in this field[30-32]. Each has captured new data and provided additional insights into the role of genomic variation in NASH pathogenesis (Table 1).

4.1. The First GWAS – Romeo et al

The first NASH related GWAS was a genome-wide survey of non-synonymous sequence variation encompassing 9229 SNPs across a mixed population of Hispanic, African American and European ancestry derived from the Dallas Heart Study[30, 38]. Although not based on direct assessment of steatosis in liver biopsy samples, the non-invasive proton magnetic resonance spectroscopy ($^1$H-MRS) technique used to assess hepatic steatosis is widely adopted in both human and murine studies[39, 40]. The striking results of this study clearly identified the patatin-like phospholipase domain-containing 3 gene ($PNPLA3$), also known as adiponutrin, as a strong modifier of NASH pathogenesis. The study demonstrated that the $PNPLA3$ I148M variant, induced by a cytosine to guanine nucleotide transversion mutation (SNP rs738409) which results in an isoleucine to methionine amino acid change at codon 148, was associated with increased hepatic fat burden ($p=5.9 \times 10^{-10}$)[30]. The variant allele was most common in Hispanics (minor allele frequency 0.49), which is the group most susceptible to NAFLD and where carriage was also found to be associated with increased ALT and AST levels. Minor allele frequency was lower in people of European descent (0.23) and lowest in African-Americans (0.17), the group found to have the lowest levels of hepatic triglyceride accumulation. Furthermore, a gene dosage effect was observed with heterozygote carriage of the I148M minor allele conferring increased hepatic triglyceride content over wild-type and homozygotes having even greater levels (2-fold more than non-carriers). The same study also reported a second sequence variation in $PNPLA3$ (rs6006460) causing a serine to isoleucine change in codon 453 (S453I). In contrast to the
I148M variation, S453I had the opposite ethnic distribution and was associated with reduced hepatic triglyceride levels[30, 41]. Together, these two variants were able to account for 72% of the ethnic variation in steatosis observed in the study population.

4.2. The Second GWAS – Chalasani et al

A second NAFLD GWAS, which was not limited to non-synonymous genetic variation, has recently been reported[31]. This used a relatively small all-female cohort of 236 biopsy-confirmed NAFLD patients and involved genotyping for 325,000 SNPs. Following adjustment for age, BMI, diabetes, waist:hip ratio and HbA1c levels in multivariate analysis, an association was identified between severity of histological NAFLD activity score and SNP rs2645424 on chromosome 8 in the gene encoding farnesyl diphosphate farnesyl transferase 1 (FDFT1) ($p=6.8 \times 10^{-7}$), an enzyme with a role in cholesterol biosynthesis. Other associations with fibrosis (chromosome 7, rs343062; $p=2.7 \times 10^{-8}$) and lobular inflammation were seen (including a SNP in the collagen gene COL13A1 ($p=2.0 \times 10^{-7}$)). In addition loci associated with raise ALT were also identified. No associations with steatosis, ballooning degeneration or portal inflammation were identified and it is perhaps surprising that no association for any feature of NAFLD with PNPLA3 was seen in this study. Whilst the associations detected in this study are potentially interesting, they remain to be independently replicated in larger patient cohorts. A recent study in a UK cohort of 340 NAFLD patients failed to find any association for the FDFT1 rs2645424 SNP with severity of either fibrosis or steatosis[42].

4.3. The Third GWAS – Speliotes et al

The most recent large-scale study of genetic variation in NAFLD adopted a two-stage approach[32]. In a first exploratory stage, computerised tomography (CT) measurement of hepatic steatosis was examined in a meta-analysis of GWAS data across individuals drawn from several large population studies (Age/Gene/Environment Susceptibility-Reykjavik Study, Old Order Amish Study, Family Heart Study and Framingham Heart Study). Genotypes for 2.4 million SNPs in over 7,100 individuals were available. SNPs from 45 loci associated with hepatic lipid content ($p<10^{-3}$) were identified and association validity tested using a cohort of 592 patients with biopsy proven NASH derived from the NIH NASH CRN cohort. Five SNPs associated with aspects of the NAFLD phenotype in or near the genes PNPLA3 (rs738408), neurocan (NCAN; rs2228603), protein phosphatase 1, regulatory (inhibitor) subunit 3B (PPP1R3B; rs4240624), glucokinase regulator (GCKR; rs780094) and lysophospholipase-like 1 (LYPLAL1; rs12137855) were identified. The rs738408 PNPLA3 SNP is in strong linkage disequilibrium with the rs738409 SNP previously identified[30].
PNPLA3, NCAN and PPP1R3B were most strongly associated with CT measured steatosis at p<5x10^{-8} in the initial study. This was also confirmed in the histology validation study for PNPLA3 and NCAN but not PPP1R3B. Two additional SNPs in or near GCKR and LYPLAL1 were shown to be associated with histological steatosis. NCAN, GCKR, LYPLAL1 and PNPLA3 were associated with histologically assessed lobular inflammation and/or fibrosis. In addition, variants in or near NCAN, GCKR and PPP1R3B were associated with altered serum lipid levels and variants near GCKR and PPP1R3B affected glycemic traits. This large study confirmed the strong association for PNPLA3 rs738408 with steatosis shown in the first GWAS study (section 4.1) and by candidate gene association studies (see 5.2.1 below). The other associations are novel though not as strong as for PNPLA3. LYPLAL1 codes for a protein with a complementary function to the PNPLA3 gene product in triglyceride breakdown so the observed association is biologically plausible. The NCAN gene product may have a role in cell adhesion whereas GCKR codes for a regulator of glucose metabolism and both are therefore interesting genes in terms of NAFLD susceptibility though further studies are needed to identify the precise causative SNPs and the underlying mechanisms affecting disease severity.

5. GENETIC MODIFIERS OF NAFLD PATHOGENESIS & PROGRESSION

5.1. Genetic Modifiers of Metabolic Syndrome Risk

When considering genetic modifiers of NAFLD one should first consider those genes that influence aetiological factors causally linked to the disease. These are exemplified by the associated conditions of insulin resistance and obesity for NAFLD. Rather than being simple environmental challenges, these factors are themselves, at least in part, genetically determined.

5.1.1. Obesity

In terms of obesity related genes[43], initial candidate studies were based on discoveries from mutant mouse models and focussed on genes related to leptin signalling which links obesity with NAFLD[9, 44, 45], however the effect of these genes has proved to be uncommon in human disease[46]. Identification of the FTO gene by GWAS analysis of a cohort of obese patients has provided a new candidate[47]. Translational validation by several groups using targeted and random mutagenesis techniques has provided exciting insights into pathogenic mechanisms, however no direct association with fatty liver has yet been described[48, 49]. The role of epigenetic imprinting as an influence on obesity and fatty liver phenotypes has also been explored in models of NAFLD[50].
5.1.2. Insulin Resistance

The relationship between insulin resistance and progression of NAFLD is complex. Whilst multiple loci from GWAS have been associated with type 2 diabetes and insulin resistance[51], few have so far carried through to NAFLD. To date, the polymorphisms in ENPP1/PC-1 (ectoenzyme nucleotide pyrophosphate phosphodiesterase 1), and IRS-1 (insulin receptor substrate-1) have been studied in candidate gene studies. A large study involving 702 pooled biopsy-proven NAFLD cases from Italy and the UK found that the SNPs in both ENPP1 (K121Q) and IRS-1 (Q972R) were independently associated with fibrosis scores >1 and insulin resistance[52]. However, a second smaller and arguably underpowered study on ENPP1 did not find a significant effect on fibrosis[53].

The adiponectin gene may also be relevant to insulin resistance and the metabolic syndrome more generally and has been well studied in relation to type 2 diabetes. A study of 119 patients with NAFLD showed that homozygosity for the variant form of a SNP at position 45 of exon 2 of this gene was a risk factor for severe fibrosis but not for NASH[54]. A more recent large Finnish study with two separate validation cohorts examined the association of polymorphisms within adiponectin receptors 1 and 2 (ADIPOR1 and ADIPOR2) as well as three peroxisome proliferator activated receptors (PPARA, PPARD and PPARG) with ¹H-MRS measured steatosis and concluded that only the ADIPOR2 SNP (rs767870) was associated with hepatic fat accumulation[55]. In a Chinese population, an association with the C161T PPARG SNP, reduced plasma adiponectin levels and NAFLD has been described. No association was found with the Gly482Ser SNP in the related protein PPARGC1A[56].

5.2. Genetic Modifiers of Hepatic Fatty Acid Flux and Triglyceride Levels

Steatosis occurs when the rate of import or synthesis of fatty acids by hepatocytes exceeds the rate of export or catabolism[57, 58]. Given the recent evidence that esterification of FFA to triglyceride is a protective mechanism which limits hepatocellular exposure to FFA induced lipotoxicity[11], polymorphisms in genes that mediate synthesis, storage and export of triglyceride are likely candidate modifiers for NASH severity and progression. The majority of NAFLD associated genes so far identified are from traditional candidate gene allele-association studies based on existing understanding of hepatocyte lipid handling.
As already discussed, interest in the role of PNPLA3 in the pathogenesis of NASH started with its identification as a genetic modifier of steatosis risk in a GWAS study. Later the same year a GWAS seeking factors that influence variation of common clinical chemistry indices reported that the I148M allele was one of two loci associated with raised serum ALT levels in populations of European and Indian-Asian descent (chromosome 10: APN1-ERLIN1-CHUK and chromosome 2: PNPLA3-SAMM50)[23]. The association of PNPLA3 with NAFLD/NASH has been independently replicated in both adult[59-62] and paediatric[62-65] cohorts, as has its association with raised ALT/AST levels[66] (Table 2). There is also evidence from biopsy based studies that carriers of the I148M variant exhibit more severe steatohepatitis with greater levels of fibrosis[60, 62]. Illustrating the convergence in our understanding of the pathogenesis of alcoholic liver disease and NAFLD, studies in alcohol dependent patients have linked advanced fibrosis or cirrhosis with carriage of the I148M allele[67]. This has been independently validated in two additional European cohorts[68, 69]. Further, there is evidence that PNPLA3 influences steatosis, fibrosis and risk of developing HCC in patients with chronic hepatitis C infection[70].

The PNPLA3 gene on chromosome 22 encodes a 481 amino acid protein that is closely related to adipose triglyceride lipase (ATGL/PNPLA2), the major triglyceride hydrolase in adipose tissue[41, 71]. PNPLA3 has been primarily associated with severity of hepatic lipid accumulation however its precise physiological role in vivo remains incompletely characterised[41]. Current evidence would suggest that the influence of PNPLA3 on hepatic steatosis is not through affecting insulin resistance as assessed by hyperinsulinaemic, euglycaemic clamp[59, 61] and plasma insulin response to oral glucose tolerance testing[72]; or with the broader features of the metabolic syndrome such as BMI, dyslipidaemia and overt type 2 diabetes mellitus[73]. Rather, consistent with the concept of NASH as a complex disease trait, the evidence suggests that PNPLA3 variation sensitises the liver to environmental stressors. Thus, in European cohorts I148M carriage was only found to associate with elevations of AST/ALT in the presence of obesity[72].

The challenge remains to better understand the role of PNPLA3 and hence the mechanisms through which the I148M variant exhibits its metabolic effects. Studies have to some extent been hampered by inter-species differences in gene expression pattern. In man PNPLA3 is found in adipose tissue but is most strongly expressed in the liver[74, 75], whilst in normal mice gene expression is at low levels in liver tissue but is abundant in white and brown adipose tissue[76, 77]. In both species hepatic expression is increased after feeding and is also raised in obese humans[59] and in Ob mice[77] whilst fasting reduces PNPLA3
expression [41, 77]. Recent in vitro studies with recombinant adiponutrin expressed in HUH-7 cells have shown that this enzyme can hydrolyze emulsified triglyceride and that the I148M polymorphism is associated with substantially reduced enzymatic activity[78]. Structural analysis of PNPLA3 indicates that the I148M mutation does not directly alter the enzyme’s highly conserved S47/D166 catalytic dyad. Instead, the I148M amino acid change lies within the hydrophobic substrate-binding groove and prevents substrate access to the active site[78]. Mechanistic studies demonstrate that postprandial PNPLA3 expression is controlled by an insulin mediated feed-forward loop through the LXR/RXR heterodimer and the transcription factor SREBP-1c[74]. These effects are responsive to post-translational control by the types of fatty acids present[74]. Specifically, selected saturated (palmitate, C16:0), monounsaturated (oleate, C18:1) and polyunsaturated fatty acids (linoleic acid, C18:2) led to increased PNPLA3 protein expression whilst very long chain fatty acids (arachadonic acid, C20:4 and eicosapentanoic acid, C20:5) did not effect PNPLA3 expression[74]. It has also been suggested that, in addition to lipolytic activity, PNPLA3 may also exhibits lipogenic transacetylase activity although this has not been uniformly replicated[78, 79]. These apparent mixed enzymatic actions, coupled with the transcriptional regulation of PNPLA3 by feeding and the specific fatty acid profile present suggests that the action of PNPLA3 could vary according tissue and metabolic milieu. Whilst partially purified PNPLA3 can catalyse triglyceride hydrolysis in vitro, expression patterns and regulation have led some investigators to suggest that the enzyme may be primarily involved in lipid remodelling rather than catabolism[74].

5.2.2. Microsomal Triglyceride Transfer Protein (MTTP)

Microsomal Triglyceride Transfer Protein (MTTP) mediates the synthesis and secretion of very-low density lipoprotein in the liver and intestine. A loss of function frame-shift mutation in MTTP causes abetalipoproteinemia (OMIM #200100) however, whilst this is associated with marked hepatocyte triglyceride accumulation, it is much less frequently associated with progressive steatohepatitis or fibrosis[80]. A guanine to thiamine transversion at position -493 in the promoter region has been associated with reduced gene transcription, lower MTTP levels and hepatocyte failure of triglyceride excretion. Evidence has been presented that NAFLD patients homozygous for the low-activity G allele have increased steatosis and higher histological NASH grade compared to heterozygous patients or patients homozygous for the high-activity T allele[81, 82]. Unfortunately, these data should be interpreted with caution as histological data on NASH activity was only available on a small cohort of 63 patients[81] and a recent larger study of 131 biopsy proven NAFLD cases from Brazil failed to show a significant effect for the same polymorphism[83].
5.2.3. *Phosphatidylethanolamine methyltransferase (PEMT)*

Phosphatidylethanolamine methyltransferase (PEMT) catalyses *de novo* synthesis of phosphatidylcholine which is needed for very-low-density lipoprotein synthesis. Two studies have reported an association between NAFLD and a non-synonymous *PEMT* exon 8 guanine-to-adenine transversion. The larger study included 107 biopsy-proven NASH patients from Japan and demonstrated that the consequent loss-of-function V175M amino acid change was more common in NASH patients compared to controls[84]. Further, V175M positive NASH patients tended to have lower BMI, suggesting they were genetically predisposed to develop NASH. A second biopsy-based American study also suggested that V175M was more common among NASH cases compared to controls[85].

5.2.4. *Apolipoprotein C3 (ApoC3)*

An association with degree of hepatic steatosis and two SNPs, rs2854116 (T-455C) and rs2854117 (C-482T), located within the promoter region of Apolipoprotein C3 (ApoC3) has been reported in cohorts of Asian-Indian and non-Asian ethnicity[86]. Both SNPs cause increased *ApoC3* expression and were already recognised as mediators of postprandial hypertriglyceridaemia in humans and murine studies via inhibition of lipoprotein lipase activity. It has been suggested that the increased circulating chylomicron-remnant load was preferentially taken up by hepatic scavenger receptors leading to steatosis[86]. However, the association of these SNPs with steatosis severity was not confirmed in two recent studies. The first of these failed to show any association between these SNPs with either hepatic triglyceride content or insulin resistance in approximately 2,000 Americans of African, European and Hispanic ethnicity[87]. The second, involving 758 biopsy proven NAFLD cases of European ethnic origin, failed to show any relationship for these *APOC3* SNPs with steatosis or fibrosis severity or with NASH[88]).

5.2.5. *Transcription Factors & Nuclear Receptors*

Following the observation that mice deficient in the *NR1I2/PXR* (pregnane X receptor) gene develop steatosis[89], a recent study has focussed on its role in lipid homeostasis and NAFLD. PXR is a transcription factor with a well established role in regulation of hepatic detoxification mechanisms[90]. It may also influence lipid homeostasis through effects on the fatty acid translocase gene (*CD36*) and several lipogenic enzymes[89]. Two SNPs (rs7643645 and rs2461823) were found to be significantly associated with aspects of the NAFLD phenotype in a group of 188 patients compared with 102 healthy controls and were
also a predictor of disease severity[91]. One SNP (rs7643645) is believed to lie within a binding site for the transcription factor \textit{HNF-4} and so may affect expression of the genes such as \textit{CYP3A4} and \textit{ABCB1}. Effects on lipid homeostasis are therefore also possible but the findings for NAFLD need to be confirmed independently. SNPs in other genes encoding proteins involved in regulating intra-hepatic FFA flux and triglyceride synthesis, storage and export are also attractive candidates as NAFLD risk factors. In addition to those described above, published candidate gene association studies support a role for several other genes including Transcription factor 7-like 2 (\textit{TCF7L2})[92], Apolipoprotein E[93, 94] and \textit{MRP2} (\textit{ABCC2})[95] in hepatic fat accumulation.

\textbf{5.3. Genetic Modifiers of Progression to Steatohepatitis}

\textbf{5.3.1. Genes Influencing Oxidative Stress}

Genetic modifiers of oxidative stress fall into two broad categories, those that encode proteins involved in the generation of reactive oxygen species (ROS) or those that mediate cellular antioxidant defence. The former group includes genes that are specific to a given aetiology, as well as those that may sensitise hepatocytes to oxidative stress damage irrespective of aetiology (e.g. HFE). The latter group includes the main mitochondrial ROS scavenger manganese-dependent superoxide dismutase (SOD2) and genes that influence the relative abundance of anti-oxidant reduced glutathione stores.

Given that liver iron deposition promotes oxidative stress, the HFE gene is an obvious candidate in disease pathogenesis for NAFLD. The evidence of a role for HFE in NAFLD is mixed. An initial Australian study demonstrated that 31% of 51 NASH patients carried at least one copy of the C282Y \textit{HFE} mutation compared to only 13% of controls[96]. This is supported by a second study in 126 patients with biopsy proven NASH which also reported that the C282Y mutation was associated more advanced disease including bridging fibrosis or cirrhosis[97]. However, another study of 263 consecutive patients with NAFLD[98], and two further small European studies[99, 100], found the prevalence of the C282Y and H63D mutations to be identical to the control population.

In the face of the increased hepatocyte FFA flux encountered in NAFLD, mitochondrial β- and extra-mitochondrial β- and ω- fatty acid oxidation are major sources of ROS. An alanine-to-valine substitution in codon 16 (A16T) of the SOD2 mitochondrial targeting sequence increases MnSOD activity which leads to increased generation of peroxide and hence cellular damage. This polymorphism has been associated with advanced hepatic fibrosis in
NAFLD in both Japanese[81] and European[101] cohorts. The European study used both case-control and intra-familial association methodology to report a consistent association between this SNP and fibrosis in NAFLD and also demonstrated a gene dosage effect where the presence of significant fibrosis increased with the number of valine (T) alleles. In the family study, the valine allele was transmitted on 47 of 76 possible occasions (62%), whereas the C allele was transmitted on only 29 of 76 occasions (38%; p=0.038) [101]. Multivariate analysis using a cohort of more than 500 patients with biopsy proven NASH showed susceptibility to advanced fibrotic disease was determined by SOD2 genotype (OR 1.56 (95%CI 1.09-2.25), p=0.014), PNPLA3 genotype (p=0.041), type 2 diabetes mellitus (p=0.009) and histological severity of NASH (p=2.0x10^{-16})[101]. Cellular oxidative stress defence is largely mediated by glutathione which may be conjugated to xenobiotics and ROS. Glutamate-cysteine ligase (gamma-glutamylcysteine synthetase), composed of two subunits encoded by the genes GCLC and GCLM, is the rate limiting step in synthesis of glutathione. Murine studies have demonstrated that absence of GCLC causes steatosis and liver failure[26]. A recent study of NAFLD in Brazil involving 131 patients examined the effects of the C-129T promoter region polymorphism involving the GCLC gene and demonstrated that carriers of the mutant allele were overrepresented in a steatohepatitis arm compared to those with simple steatosis[83].

5.3.2. Genes Influencing Endotoxin Response & TLR4

In recent years there has been mounting interest in the role of the gut flora-liver axis in the pathogenesis of fatty liver disease with evidence supporting a role for endotoxin-mediated cytokine release in the pathogenesis of NAFLD from both human and murine studies. The identification of promoter region polymorphisms in genes encoding endotoxin receptors has provided investigators with a new set of candidate genes for study. CD14, a lipopolysaccharide receptor expressed on monocytes, macrophages and neutrophils enhances toll-like receptor-4 (TLR4) endotoxin receptor signalling. Genetic data supporting a role for TLR4 polymorphisms rs4986791 and rs4986790 in the pathogenesis of hepatitis C related fibrosis in man is available[102, 103]. In addition, a spontaneous null mutation in TLR4 present in C3H/J mice has provided a useful tool with which to explore the role of TLR4/endotoxin in NAFLD pathogenesis in the laboratory[104]. One preliminary study in NAFLD found no association with polymorphisms in either TLR4 or the NOD2 receptor for bacterial cell wall peptidoglycan but did reported an association with a cytosine-to-thymine polymorphism at position -159 in the 5’ promoter region of CD14[105]. Carriage of the CD14 minor allele is associated with increased expression of both soluble and membrane bound
Genetic Modifiers of Non-Alcoholic Fatty Liver Disease Progression

CD14[106]. Once again, larger studies are required to validate these findings.

5.3.3. Genes Influencing Cytokine Activity

Investigation of the role of TNFα and other cytokines in NASH was first stimulated by the apparent similarity between NASH and the effects of pro-inflammatory cytokines such as TNFα[107]. The TNFα guanine-to-adenine promoter polymorphism at position -238 has been associated with NASH[108]. In a separate study, DNA from 102 Japanese patients was screened for SNPs in TNFα and serum TNF receptor-2 levels assayed. sTNF-2 levels were significantly higher in NASH patients than patients with simple steatosis or healthy controls. The carrier frequencies of polymorphisms at -1031C and -863A in the promoter region were significantly higher in NASH than steatosis, however no significant difference between those with NAFLD and the control population was detected[109]. Whilst interesting, this small study requires replication before any firm conclusions on the role of TNFα as a NASH susceptibility gene can be made. In a small study of 114 NAFLD cases, including 59 who had undergone liver biopsy, an IL-6 promoter region G-174C SNP was found to be more common in cases of NASH[53].

5.4. Genetic Modifiers of Fibrogenesis and Disease Progression

The majority of research into hepatic fibrogenesis has been conducted in chronic hepatitis C infected cohorts however, as stellate cell activation and collagen deposition are a final common pathway of chronic liver injury, the genes identified remain strong candidates for a role in NAFLD-related fibrosis[110]. Candidates suggested by these studies include transforming growth factor (TGF)-β1, connective tissue growth factor, matrix metalloproteinase 3, PPARα, DDX5, CPT1A and various fibrogenic adipocytokines including angiotensin II. One of the few pro-fibrotic polymorphisms independently validated in multiple human viral hepatitis studies and in animal models is carriage of the pro-coagulant Factor V Leiden mutation (R506Q; OMIM #612309)[37, 111, 112]. Give the evidence that the metabolic syndrome is a pro-coagulant state and that fibrosis progression in NAFLD is associated with the presence of greater thrombotic risk[113], it is tempting to speculate that polymorphisms related to coagulation activation and signalling may also influence fibrogenesis[114] however gene association evidence is currently lacking.

The transcription factor nuclear factor-κB (NF)-κB promotes survival of hepatic myofibroblasts and hepatic fibrogenesis. It is recognised that angiotensin II mediates this
effect through activation of IkB kinase (IKK) phosphorylation of the NF-kB subunit RelA at Ser 536[115]. Five of 12 SNPs studied in the angiotensin II receptor 1 (ATGR1) gene were shown to associate with steatohepatitis in a cohort of patients with biopsy proven NASH[116]. The strongest association was with SNP rs3772622 (OR=1.95, 95%CI 1.49-2.55; p=1.2x10^{-6}) with the other four SNPs being found to lie within the same linkage block. This SNP was also shown to be significantly associated with degree of hepatic fibrosis[116]. Another report indicates that obese patients possessing both the high TGF\(\beta\)1 and angiotensinogen producing SNPs may be more susceptible to advanced fibrosis. However this study involved very small patient numbers and is yet to be replicated[117]. These findings are consistent with reports that show benefit of angiotensin receptor blockers such as losartan in ameliorating fibrosing steatohepatitis both in vitro and in vivo and demonstrate how knowledge of the genetic modifiers of disease may inform development of potential therapies[115, 118].

Kruppel-like factor (KLF6) is a ubiquitously expressed transcription factor which was found to be expressed by activated stellate cells soon after injury[119]. Evidence suggests that KLF6 expression is increased in rat models of NASH and regulates expression of several key genes that mediate fibrogenesis, making it an attractive candidate modifier of fibrosis severity in human NAFLD[120]. A SNP affecting mRNA splicing (rs3750861) has been shown to be functionally significant[121]. Carriage of this SNP has been shown to associate with more mild hepatic fibrosis in three separate European NAFLD cohorts[36].

6. CONCLUSIONS

Like the majority of clinically important diseases, NAFLD is a complex disease trait where genetic factors and environmental influences combine to determine disease phenotype and progression. The relative importance of these factors will vary between populations depending on background modifier genes and lifestyle choices/challenges. During the last decade the advent of the GWAS and whole-genome sequencing technologies has provided investigators with the tools to comprehensively study variability within the human genome. The advent of genome-led tailored therapy based on genetic risk assessment in the clinic remains some way off. However, our greater understanding of how genetic variation modifies disease phenotype offers new insights into the underlying physiological processes and pathophysiological mechanisms. With this opportunity comes the challenge of translating these findings into tangible therapeutic benefits.
<table>
<thead>
<tr>
<th>Study</th>
<th>SNP</th>
<th>Gene Symbol</th>
<th>Gene Name/Location</th>
<th>Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Romeo, 2008 [30]</td>
<td>rs738409</td>
<td>PNPLA3</td>
<td>Patatin-like phospholipase domain-</td>
<td>MRI measured steatosis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>containing 3</td>
<td></td>
</tr>
<tr>
<td>Chalasani, 2010</td>
<td>rs2645424</td>
<td>FDFT1</td>
<td>Farnesyl diphosphate farnesyl transferase 1</td>
<td>Histological NASH activity score</td>
</tr>
<tr>
<td>[31]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs343062</td>
<td>-</td>
<td>Chromosome 7</td>
<td>Histological fibrosis</td>
</tr>
<tr>
<td></td>
<td>rs1227756</td>
<td>COL13A1</td>
<td>Collagen, type XIII, alpha 1</td>
<td>Histological lobular inflammation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs6591182</td>
<td>-</td>
<td>Chromosome 11</td>
<td>Histological lobular inflammation</td>
</tr>
<tr>
<td></td>
<td>rs887304</td>
<td>EFCAB4B</td>
<td>EF-hand calcium binding domain 4B</td>
<td>Histological lobular inflammation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs2499604</td>
<td>-</td>
<td>Chromosome 1</td>
<td>Serum ALT</td>
</tr>
<tr>
<td></td>
<td>rs6487679</td>
<td>PZP</td>
<td>Pregnancy zone protein</td>
<td>Serum ALT</td>
</tr>
<tr>
<td></td>
<td>rs14212001</td>
<td>-</td>
<td>Chromosome 18</td>
<td>Serum ALT</td>
</tr>
<tr>
<td></td>
<td>rs2710833</td>
<td>-</td>
<td>Chromosome 4</td>
<td>Serum ALT</td>
</tr>
<tr>
<td>Speliotes, 2011</td>
<td>rs738408</td>
<td>PNPLA3</td>
<td>Patatin-like phospholipase domain-</td>
<td>CT measured steatosis &amp; histological NAFLD</td>
</tr>
<tr>
<td>[32]</td>
<td></td>
<td></td>
<td>containing 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs2228603</td>
<td>NCAN</td>
<td>Neurocan</td>
<td>CT measured steatosis &amp; histological NAFLD</td>
</tr>
<tr>
<td></td>
<td>rs4240624</td>
<td>PPP1R3B</td>
<td>Protein phosphatase 1, regulatory (inhibitor) subunit 3B</td>
<td>CT measured steatosis (not validated in histology cohort)</td>
</tr>
<tr>
<td></td>
<td>rs780094</td>
<td>GCKR;</td>
<td>Glucokinase regulator</td>
<td>Histological NAFLD</td>
</tr>
<tr>
<td></td>
<td>rs12137855</td>
<td>LYPLAL1</td>
<td>Lysophospholipase-like 1</td>
<td>Histological NAFLD</td>
</tr>
<tr>
<td>Study</td>
<td>Study Design</td>
<td>Cohort Studied</td>
<td>Associated Phenotype</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>--------------</td>
<td>---------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Yuan, 2008 [23]</td>
<td>GWAS</td>
<td>Indian-Asian &amp; European</td>
<td>↑ ALT (serum biochemistry only)</td>
<td></td>
</tr>
<tr>
<td>Romeo, 2008 [30]</td>
<td>GWAS</td>
<td>USA based European-American; African-American &amp; Hispanic</td>
<td>↑ ¹H-MRS Liver TG content; ↑ ALT/AST in Hispanics</td>
<td></td>
</tr>
<tr>
<td>Kotronen, 2009 [59]</td>
<td>Candidate Gene</td>
<td>Finnish</td>
<td>↑ ¹H-MRS Liver TG content; No association with Insulin Resistance</td>
<td></td>
</tr>
<tr>
<td>Sookoian, 2009 [60]</td>
<td>Candidate Gene</td>
<td>Argentinean</td>
<td>↑ Biopsy proven steatosis; ↑ Risk of histological progression &amp; inflammation</td>
<td></td>
</tr>
<tr>
<td>Kantartzis, 2009 [61]</td>
<td>Candidate Gene</td>
<td>German</td>
<td>↑ ¹H-MRS Liver TG content; ↑ ALT &amp; ↑ AST</td>
<td></td>
</tr>
<tr>
<td>Romeo, 2010 [72]</td>
<td>Candidate Gene</td>
<td>Italian (BMI&gt;30kg/m²)</td>
<td>↑ ALT if Obese</td>
<td></td>
</tr>
<tr>
<td>Kollerits, 2010 [66]</td>
<td>Candidate Gene</td>
<td>European</td>
<td>↑ ALT &amp; ↑ AST</td>
<td></td>
</tr>
<tr>
<td>Romeo, 2010 [63]</td>
<td>Candidate Gene</td>
<td>Italian (Obese children/adolescents)</td>
<td>↑ ALT &amp; ↑ AST</td>
<td></td>
</tr>
<tr>
<td>Speliotes, 2010 [73]</td>
<td>Candidate Gene</td>
<td>USA based European-American</td>
<td>↑ Risk of histological progression &amp; inflammation</td>
<td></td>
</tr>
<tr>
<td>Speliotes, 2011 [32]</td>
<td>GWAS</td>
<td>Meta-analysis European and USA based cohorts</td>
<td>↑ CT Liver TG content; ↑ Risk of histological progression &amp; inflammation</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Summary of Genetic Modifiers of Progressive NAFLD
7. REFERENCES


Genetic Modifiers of Non-Alcoholic Fatty Liver Disease Progression


[43] K.A. Fawcett, I. Barroso, The genetics of obesity: FTO leads the way, Trends Genet, 26 266-274.


[70] L. Valenti, M. Rumi, E. Galmozzi, A. Aghemo, B. Del Menico, S. De Nicola, P. Dongiovanni, M. Maggioni, A.L. Fracanzani, R. Rametta, M. Colombo, S. Fargion, Patatin-like phospholipase domain-
Genetic Modifiers of Non-Alcoholic Fatty Liver Disease Progression


[76] M. Hoekstra, Z. Li, J.K. Kruijt, M. Van Eck, T.J. Van Berkel, J. Kuiper, The expression level of non-alcoholic fatty liver disease-related gene PNPLA3 in hepatocytes is highly influenced by hepatic lipid status, J Hepatol, 52 (2010) 244-251.


