# Northumbria Research Link

Northumbria
University
NEWCASTLE

UniversityLibrary

*Article*

# A *PDCD1* Role in the Genetic Predisposition to NAFLD-HCC?

Nardeen Eldafashi [1,2,†], Rebecca Darlay [3,†], Ruchi Shukla [4,†], Misti Vanette McCain [1], Robyn Watson [1], Yang Lin Liu [1], Nikki McStraw [4], Moustafa Fathy [2], Michael Atef Fawzy [2], Marco Y. W. Zaki [1,2], Ann K. Daly [1], João P. Maurício [1], Alastair D. Burt [1], Beate Haugk [5], Heather J. Cordell [3], Cristiana Bianco [6], Jean-François Dufour [7,8], Luca Valenti [6,9], Quentin M. Anstee [1,10] and Helen L. Reeves [1,10,*]

1   Translational and Clinical Research Institute, Faculty of Medical Sciences, The Medical School, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK; nardeen.rafat@mu.edu.eg (N.E.); Misti.McCain@newcastle.ac.uk (M.V.M.); Robyn.Watson@newcastle.ac.uk (R.W.); ejayliu@hotmail.com (Y.L.L.); marcozaki@mu.edu.eg (M.Y.W.Z.); a.k.daly@newcastle.ac.uk (A.K.D.); J.Mauricio2@newcastle.ac.uk (J.P.M.); alastair.burt@newcastle.ac.uk (A.D.B.); quentin.anstee@newcastle.ac.uk (Q.M.A.)
2   Biochemistry Department, Faculty of Pharmacy, Minia University, Minia 61519, Egypt; mostafa_fathe@minia.edu.eg (M.F.); michael.fawzy777@mu.edu.eg (M.A.F.)
3   Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, International Centre for Life, Newcastle upon Tyne NE1 3BZ, UK; rebecca.darlay@newcastle.ac.uk (R.D.); heather.cordell@newcastle.ac.uk (H.J.C.)
4   Biosciences Institute, Faculty of Medical Sciences, The Medical School, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK; Ruchi.Shukla@newcastle.ac.uk (R.S.); nikkimartinamcstraw@hotmail.co.uk (N.M.)
5   Department of Cellular Pathology, Royal Victoria Infirmary, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle NE1 4LP, UK; beate.haugk@nhs.net
6   Translational Medicine, Department of Transfusion Medicine and Hematology, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, 20122 Milan, Italy; cristiana.bianco@policlinico.mi.it (C.B.); luca.valenti@unimi.it (L.V.)
7   University Clinic for Visceral Surgery and Medicine, University Hospital of Bern, 3010 Bern, Switzerland; jf.dufour@svmed.ch
8   Hepatology, Department of Biomedical Research, University of Bern, 3012 Bern, Switzerland
9   Department of Pathophysiology and Transplantation, Università degli Studi di Milano, Fondazione IRCCS Cà Granda Ospedale Maggiore Policlinico, 20122 Milan, Italy
10  The Liver Unit, Freeman Hospital, Freeman Road, Newcastle upon Tyne Hospitals NHS Foundation Trust, Heaton NE7 7DN, UK
*   Correspondence: Helen.Reeves@newcastle.ac.uk; Tel.: +01-91-208-4423
†   Joint 1st author.

**Simple Summary:** Many more people are dying each year from primary liver cancers arising in obesity-related fatty liver disease. Often these cancers are a consequence of fatty liver disease progression, with inflammation, scarring and cirrhosis. Less often, cancers develop in the presence of fat without cirrhosis. Evidence from animal models suggests the immune response to fat is important. We have explored genetic variations in candidate immunoregulatory genes. Our study of nearly one-thousand patients with fatty liver disease, comparing 391 with cancers to 594 without, indicates that genetic variation in a gene (*PDCD1*) that codes for the T cell receptor PD-1 may be important. Inherited variations that affect function of immunoregulatory proteins like PD-1 may underpin why some patients with fatty liver disease—whether they have cirrhosis or not—are more likely to develop liver cancer.

**Abstract:** Obesity and non-alcoholic fatty liver disease (NAFLD) are contributing to the global rise in deaths from hepatocellular carcinoma (HCC). The pathogenesis of NAFLD-HCC is not well understood. The severity of hepatic steatosis, steatohepatitis and fibrosis are key pathogenic mechanisms, but animal studies suggest altered immune responses are also involved. Genetic studies have so far highlighted a major role of gene variants promoting fat deposition in the liver (*PNPLA3* rs738409; *TM6SF2* rs58542926). Here, we have considered single-nucleotide polymorphisms (SNPs) in candidate immunoregulatory genes (*MICA* rs2596542; *CD44* rs187115; *PDCD1* rs7421861 and rs10204525), in 594 patients with NAFLD and 391 with NAFLD-HCC, from three European centres.

Associations between age, body mass index, diabetes, cirrhosis and SNPs with HCC development were explored. *PNPLA3* and *TM6SF2* SNPs were associated with both progression to cirrhosis and NAFLD-HCC development, while *PDCD1* SNPs were specifically associated with NAFLD-HCC risk, regardless of cirrhosis. *PDCD1* rs7421861 was independently associated with NAFLD-HCC development, while *PDCD1* rs10204525 acquired significance after adjusting for other risks, being most notable in the smaller numbers of women with NAFLD-HCC. The study highlights the potential impact of inter individual variation in immune tolerance induction in patients with NAFLD, both in the presence and absence of cirrhosis.

## 1. Introduction

Commonly arising on a background of cirrhosis, the geographic variations in global incidence and mortality attributed to hepatocellular carcinoma (HCC) have largely reflected differences in the prevalence of chronic viral hepatitis, namely, hepatitis B (HBV) and hepatitis C (HCV) [1]. Alcohol-related liver disease (ARLD) has also been recognised as a significant contributor to risk, particularly in Western nations [1]. With over hundreds of thousands of new cases and deaths per year, for a cancer where the major risk factors are known and many cases potentially preventable or treatable if detected at an early stage, the needs to better understand this and exploit opportunities to intervene and improve outcomes are widely recognised. In some countries with a particularly high incidence of HCC, population-level interventions have had an impact, with reductions in HCC mortality consequent to HBV immunisation [2] or early HCV-HCC detection and treatment [3]. The introduction of effective antiviral therapies for both HBV and HCV have also heralded major scientific advances [4,5]. It is disappointing, therefore—that in contrast to the majority of other cancers—HCC deaths globally have continued to increase [6].

Epidemiological studies in the Western nations have highlighted a close association between HCC and the rising prevalence of obesity, associated with the metabolic syndrome [7]. Non-alcoholic fatty liver disease (NAFLD) is the liver manifestation of this syndrome, and while primary care management targeting diabetes and hypertension [8], as well as public health campaigns to reduce smoking [9], have had a significant impact on cardiovascular deaths, the prevalence, morbidity and mortality consequent to NAFLD and NAFLD-HCC have escalated. NAFLD is estimated to affect 25% of Western adults [10]. Although the majority of those have steatosis without significant inflammation and fibrosis, the prevalence is so high that even with only a minority of patients developing progressive disease, NAFLD is now one of the commonest causes of cirrhosis and HCC [7]. In patients with known NAFLD cirrhosis who are fit for therapeutic intervention, HCC surveillance with bi-annual liver ultrasound is widely endorsed, although views of its cost effectiveness and benefit are frequently debated [11]. One of the consequences of ineffective or failed surveillance is that NAFLD-HCC is more often detected at an advanced stage, where older age and metabolic syndrome associated comorbidities further limit treatment options. There is a pressing need to develop a better understanding of why some patients with NAFLD develop HCC—both in the presence or absence of cirrhosis—if we are to identify clinically useful biomarkers that will inform intervention strategies or more cost-effective approaches to surveillance and early detection.

Previous studies have recognised that NAFLD progression to advanced fibrosis or cirrhosis is associated with diabetes, overweight, male sex and advanced age, as well as a genetic predisposition [12,13]. Although there have not been genome wide association studies (GWAS) reporting genetic variables associated with NAFLD-HCC risk, single-nucleotide polymorphisms (SNPs) in a number of candidate genes associated with NAFLD development have also been studied in the context of NAFLD-HCC. SNPs in genes that

promote fat accumulation in hepatocytes as well as reportedly promoting NAFLD-HCC include rs738409 in patatin-like phospholipase domain-containing 3 (*PNPLA3*) [13,14] and rs58542926 in transmembrane-6 superfamily member-2 (*TM6SF2*) [7,15–17]. The sensitivities and specificities of individual SNPs such as these for determining HCC risk in NAFLD is poor. However, their clinical utility may be improved using combinations of risk associated SNPs. A polygenic risk score (PRS) combining genetic variations in *PNPLA3* and *TM6SF2*, with others in membrane bound O-acyltransferase domain containing 7 (*MBOAT7*), glucokinase regulator (*GCKR*) and 17β-hydroxysteroid dehydrogenase type 13 (*HSD17B13*), has recently been proposed [18].

While increased liver fat stores appear to increase the risk of HCC development [18], data from animal models and human studies have also highlighted a role of the immune environment in the modulation of cancer risk in NAFLD [7]. Here, in the hope of enhancing our understanding of NAFLD-HCC in a clinically translatable fashion, we have explored SNPs in a number of candidate HCC immunoregulatory genes, in addition to those in *PNPLA3* and *TM6SF2*, in patients with NAFLD and NAFLD-HCC. The SNPs selected were based on previous reports and included rs2596542 in Major Histocompatibility Complex class I polypeptide related sequence A (*MICA*) [19], rs187115 in Cluster of differentiation 44 (*CD44*) [20–22], as well as rs7421861 and rs10204525 in programmed cell death-1 (*PDCD1*) which encodes PD-1 [23–33]. *PDCD1* rs10204525 C < T is common in Asian populations, with functional implications for patients with HBV. The data presented here highlight genetic variability in *PDCD1* associated with HCC risk in NAFLD in European patients. The clinical implications are discussed.

## 2. Materials and Methods

### 2.1. Patients

This study included a total of 594 NAFLD Caucasian patients derived from the European NAFLD Registry [34] who were considered as controls, and 391 patients with HCC attributed to NAFLD. The primary cohort was from the Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle, UK (controls 416, HCC 198). Recruitment of patients with HCC was from 1 January 2004 until 3 December 2019, with a minimum follow-up period of 12 months, until 31 December 2020. There were two replication cohorts, from Inselspital Hospital, Bern, Switzerland (76 controls, 84 HCC), and Milan, Italy (102 controls with severe fibrosis, 109 HCC), as approved by each institution. All patients signed a consent form. The Berne patients have been recruited since 2010, with follow up ongoing until death or transplantation. For the Milan patients, those with cancer were recruited between 2010 and 2016. The Milan control patients were those with NAFLD and severe fibrosis or cirrhosis, recruited between 2018 and 2020. There was no subsequent follow up of the Milan patients for research purposes. Demographic data included age and sex, with body mass index (BMI) and presence of type 2 diabetes (T2DM) recorded. Each of the controls had a documented history of NAFLD, with an otherwise negative liver screen (HBV and HCV serology, serum ferritin, autoantibody screen, and alpha-1-antitrypsin level). Patients classed as having NAFLD-HCC were men or women with evidence of a fatty liver on biopsy or imaging or having the presence of at least T2DM or BMI > 30, with an otherwise negative liver screen, drinking <21 or 14 units of alcohol per week, respectively, for at least 5 years prior to their first presentation with liver disease. The diagnosis of HCC was established through noninvasive assessment or histologically according to EASL clinical practice guidelines [1].

### 2.2. Genotyping Strategy

DNA was prepared as previously described [13]. SNPs in the candidate genes (*PNPLA3* rs738409, *TM6SF2* rs58542926, *MICA* rs2596542, *CD44* rs187115, *PD-1* rs7421861 and rs10204525), were genotyped by taqman assay (#4351379, Applied Biosystems Inc., Waltham, MA, USA) according to the manufacturers protocol or determined by GWAS as reported previously [35]. Taqman assay details are summarised in Table S1.

## 2.3. Statistical Analysis

Associations between SNPs and HCC status were analysed by logistic regression in PLINK v1.9 [36], with age, sex, cirrhosis and diabetes status included as additional covariates. Multivariate analyses were expressed as odds ratio (OR) with 95% confidence intervals (CI). Logistic regression was repeated conditioned on the genotypes of both rs738409 and rs58542926.A subset of non-cirrhotic individuals were analysed separately by similar methods. A fixed-effect meta-analysis was performed with METAv1.7 [37] on Newcastle, Berna and Milan cohorts using the log Odds Ratios (lnORs) and standard errors from the individual cohort analyses. Haplotype analysis for *PDCD1* SNPs was carried out using software UNPHASED, as previously described [38]. In the patients with liver cancer, associations were also explored with clinicopathological variables and survival, by Kaplan Meier and Cox regression analyses using SPSS version 25 (IBM Corp., Armonk, NY, USA) licensed to Newcastle University.

## 2.4. Power Calculation

A power calculation of sample size to evaluate the effect of genetic variation on HCC risk in NAFLD was performed using the "case–control for discrete traits" option in GPC software (http://zzz.bwh.harvard.edu/gpc/ (accessed on 3 March 2021)). To perform this, the prevalence of NAFLD-HCC in a NAFLD population similar to our own was estimated as 0.5%, based epidemiological studies data [39], but also from HCC incidence as reported in large cohorts included in recent therapeutic trials for NASH [40,41]. The minor allele frequencies used were those of our own NAFLD control population. These power calculations are summarised in Supplementary Table S2. In the combined cohorts, there was an estimated 80% power to detect a significant allelic relative risk of $\geq 1.3$, falling to 55–60% in the Newcastle cohort alone. At a level of relative risk $\geq 1.4$, the power was 95% in our combined cohort and 76–80% in the Newcastle cohort.

Graphical abstract was designed using BioRender software (https://biorender.com/ accessed on 2 February 2021; Toronto, ON, Canada).

## 3. Results

### 3.1. The Newcastle Patient Cohorts Clinical Information

#### 3.1.1. The NAFLD Control and NAFLD-HCC Patient Characteristics

The NAFLD control patients included all patients, regardless of age, sex or stage of liver disease, attending the NAFLD specialist clinic in Newcastle who consented to take part in clinical research studies. The cohort was representative of the patients seen. The patients with HCC were also representative, again including all patients who consented to the use of their tissues and data for research. The progression of NAFLD to HCC in Newcastle is known to be associated with increasing age, male sex, T2DM and the presence of cirrhosis—each of which were highly significantly different between the two groups, as summarised in Table 1. Notably, although increasing BMI is well established as a risk factor for NAFLD progression to cirrhosis, BMI was significantly reduced in NAFLD-HCC cases compared to NAFLD controls. The lower BMI may reflect older age and/or sarcopenia, or be a consequence of cancer development, rather than being a predisposing feature—falling particularly in those with more advanced TNM stage (Table S3A). Note, however, that the BMI was lower across all TNM stages relative to controls, particularly in the non-cirrhotic (NC) cases (Table S3B).

#### 3.1.2. The Newcastle NAFLD-HCC Cohort, Comparing Those with and without Cirrhosis

The features of the NAFLD-HCC cases are summarised in Supplementary Table S4. In 61.1%, HCC arose in the presence of cirrhosis. As previously reported, the median age of individuals with NC NAFLD-HCC was significantly greater [42]. Tumours were also more advanced in terms of size, with more patients presenting with TNM stage III or IV disease. Notably though, 48% of patients with NC NAFLD-HCC presented with a single large tumour, with only 6.5% classed as TNM Stage II, compared to 35.5% and 36.4% for cirrhotic

NAFLD-HCC cases, respectively. These differences in TNM stage were highly significant. Furthermore, the NC NAFLD-HCC cases had relatively preserved liver function, as shown by a number of the parameters (albumin, bilirubin, ascites, encephalopathy, prothrombin time and Childs Pugh Stage) presented in Supplementary Table S4. Consequently, despite their advanced age and 36.4% receiving only supportive care, NC NAFLD-HCC patients were twice as likely to be treated surgically. Survival in NAFLD-HCC cases was mostly determined by stage (tumour number, size and portal vein invasion) at presentation rather than the presence of cirrhosis, as summarised in Supplementary Table S5. Age, ECOG PST and tumour differentiation were also significant in multivariate Cox Regression analyses.

**Table 1.** Demographic characteristics of Newcastle NAFLD and NAFLD-HCC cohorts.

| Phenotype | Group | NAFLD ($n$ = 416) | NAFLD/HCC ($n$ = 198) | $p$-Value [1] |
|---|---|---|---|---|
| Age (Mean $\pm$ SD) | | 52.97 $\pm$ 0.58 | 72.21 $\pm$ 0.65 | <0.0001 |
| Gender | male (%) | 232 (55.8) | 157 (79.3) | <0.0001 |
| | female (%) | 184 (42.8) | 41 (20.7) | |
| BMI (Mean $\pm$ SD) | | 35.03 $\pm$ 0.28 | 31.96 $\pm$ 0.44 | <0.0001 |
| Diabetes | no (%) | 198 (48.1) | 60 (30.3) | <0.0001 |
| | yes (%) | 214 (51.9) | 138 (69.7) | |
| Cirrhosis | no (%) | 353 (84.9) | 77 (38.9) | <0.0001 |
| | yes (%) | 63 (15.1) | 121 (61.1) | |

[1] $p$-values estimated by Mann–Whitney or Chi Square tests for continuous or categorical datasets respectively.

Raised BMI and T2DM were expectedly common in the NAFLD-HCC cohort overall. The lower BMI in NC NAFLD-HCC cases has been mentioned above, but also of note, T2DM was less prevalent, being present in 59.7% of NC cases compared to 76.0% of cirrhotic cases. While these features are widely associated with cancer risk, it is perhaps not so surprising that they were less striking in the NC NAFLD-HCC cases, given these are also strongly associated with progression to cirrhosis, and by definition the NC cases do not have cirrhosis. The differences do suggest, however, that the key genes and/or pathways in obesity and T2DM promoting NAFLD progression to cirrhosis, versus those promoting NAFLD progression to HCC, are not necessarily the same. The likelihood that other pro-carcinogenic pathways may play a relatively larger role in NAFLD patients without cirrhosis has been previously suggested [43].

The diagnosis of HCC was confirmed histologically in 93/198 cases (47.0%). Histopathology features were often heterogenous even within the biopsies, but the more common architectural type, as well as any subtype described [44], was attributed where possible ($n$ = 90). Apart from fibrolamellar cases being restricted to NC only, the distribution of other subtypes was similar in cirrhotic versus NC cancers. Trabecular architecture was common, followed by solid/compact tumours. In some cases, features of both hepatocellular and cholangiocarcinoma were evident, and these were classed as mixed tumours. In over 60% no subtype was reported. In those with a subtype noted, steatohepatitic HCC (SH-HCC) was the most common, present in 20–25% of cases regardless of cirrhosis. In this series, there were no cases described as macrotrabecular massive, chromophobe or neutrophil-rich.

*3.2. The Newcastle Patients' Cohort Genotype Data*

3.2.1. PNPLA3 rs738409 C > G and TM6SF2 rs58542926 C > T Genotyping Data

All of the genotypes—for fat and immunoregulatory genes—were in Hardy–Wienberg equilibrium in the Newcastle cases and controls cohort. *PNPLA3* rs738409 and *TM6SF2* rs58542926 are two well-studied SNPs where the minor alleles are reported to be associated with the development of HCC in fatty liver attributed to NAFLD or ARLD [13,45]. The allele frequency data for these two SNPs in the Newcastle NAFLD and NAFLD-HCC cohorts is shown in the first two column clusters of Figure 1 and Supplementary Table S6. In Europe, 78% of people carry at least one copy of the *PNPLA3* rs738409 "C" allele (https://www.ncbi.nlm.nih.gov/ (accessed on 18 January 2021)), with just 5% being homozygous for the G allele. Figure 1 shows that the percentage of patients carrying the C allele falls quite markedly in patients with NAFLD, with and without HCC in Newcastle. Notably, >20% of those with NAFLD-HCC were homozygous for the variant G allele. The differences in *PNPLA3* genotype assessed by logistic regression were significant, with statistical tests assessing dominant and recessive models of inheritance shown in Supplementary Table S7. For *TM6SF2* rs58542926 (yellow bars Figure 1), both individuals heterozygous or homozygous for the variant T allele were significantly elevated in patients with NAFLD-HCC, although—as expected—the minor allele frequency *TM6SF2* rs58542926 was lower than that of *PNPLA3*.
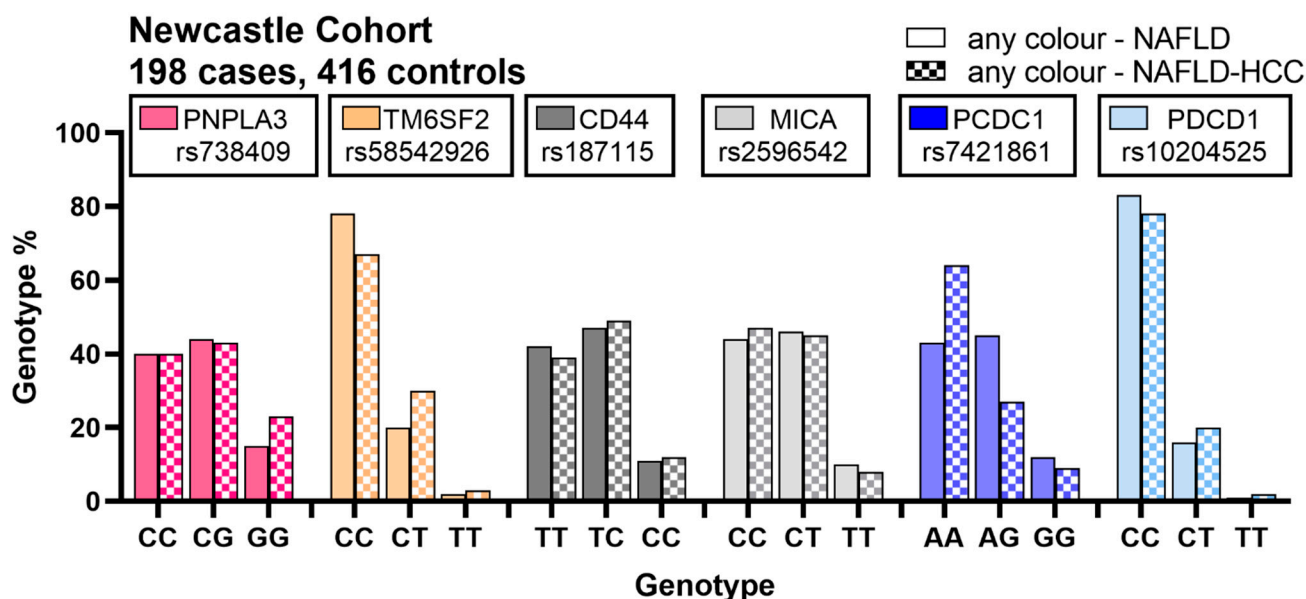


**Figure 1.** Allele frequencies of candidate SNPs in Newcastle cohorts.

Figure 1 The figure shows the allelic frequency (wild type, heterozygous mutant, homozygous mutant) of different candidate SNPs in fat-regulatory genes (*PNPLA3* and *TM6SF2*) and immune-regulatory genes (*CD44*, *MICA* and *PDCD1*) in NAFLD (*n* = 416; solid bars) and NAFLD-HCC (*n* = 198; dotted bars) patients from Newcastle cohort.

3.2.2. The Newcastle Patients' Cohort Genotype Data—Immunoregulatory Genes

There was no evidence that any of the candidate immunoregulatory genes were associated with the development of NAFLD, when comparing the minor allele frequency (MAF) to controls from the European NAFLD GWAS [34]. The MAF for each of the SNPs (bar rs10204525 which failed quality control) are shown in Supplementary Table S8. There were no statistically significant differences in genotypes for either *CD44* rs187115 or *MICA* rs2596542 when comparing the NAFLD versus NAFLD-HCC cases (grey bars Figure 1; Supplementary Table S7). Considering the SNPs tested in *PDCD1* however, there was quite a difference in allele frequencies for rs7421861 A > G, with the wild type A allele being

significantly more common in patients with NAFLD-HCC. This was in keeping with the variant allele protecting against the development of HCC in NAFLD, as the wild type A allele was more common in the patients with cancer. The *PDCD1* rs10204525 T allele was not common and was not significantly different in NAFLD-HCC versus NAFLD controls.

### 3.2.3. The Newcastle Cohort-Multivariate Analyses Exploring Associations with NAFLD-HCC

For each of the fat or immunoregulatory SNPs, we went on to explore their association with the development of HCC in NAFLD, independently of the other established risk factors for NAFLD-HCC—namely, age, gender, the presence of T2DM and the presence of cirrhosis. These data are summarised in Table 2. Of note, while both *PNPLA3* and *TM6SF2* variants are widely acknowledged as promoting NAFLD-HCC, neither was independently significant of all the other risk factors in this single centre cohort. The proposed association of *PDCD1* rs7421861, with the variant allele less common in NAFLD-HCC, remained highly significant when including each of the other risk factors. Notably, the *PDCD1* rs10204525 C > T SNP, acquired significance after adjusting for the other risk factors. Inspection of the data (Supplementary Table S9) indicated that the *PDCD1* rs10204525 T allele was more common in Newcastle cohort females with NAFLD-HCC (15/41; 36.6%) compared to female NAFLD controls (30/184; 16.8%, $p$ = 0.003, Chi Square).

**Table 2.** Allelic analyses, including multivariate regression and conditioning on the fat-regulatory genes.

| Gene | | Identity | $p$ Value | OR | Conditioned on *PNPLA3* + *TM6SF2* | |
|------|------|----------|-----------|-----|------------------|-----|
| | | | | | $p$ Value | OR |
| *PNPLA3* | C > G | rs738409 | 0.01750 * | 1.33 (1.05–1.68) | NA | NA |
| *TM6SF2* | C > T | rs58542926 | 0.00531 ** | 1.60 (1.15–2.22) | NA | NA |
| *MICA* | T > C | rs2596542 | 0.37940 | 0.89 (0.68–1.16) | 0.3741 | 0.89 (0.68–1.16) |
| *CD44* | C > T | rs187115 | 0.42710 | 1.11 (0.86–1.44) | 0.4155 | 1.11 (0.86–1.45) |
| *PDCD1* | A > G | rs7421861 | 0.00014 *** | 0.59 (0.45–0.78) | 0.000465 *** | 0.61 (0.47–0.81) |
| *PDCD1* | C > T | rs10204525 | 0.16470 | 1.31 (0.90–1.91) | 0.1838 | 1.29 (0.88–1.91) |
| Age, Sex, Cirrhosis, T2DM | | | | | | |
| *PNPLA3* | C > G | rs738409 | 0.06432 | 1.49 (0.98–2.26) | NA | NA |
| *TM6SF2* | C > T | rs58542926 | 0.48770 | 0.82 (0.47–1.44) | NA | NA |
| *MICA* | T > C | rs2596542 | 0.84010 | 0.95 (0.60–1.52) | 0.9079 | 0.97 (0.61–1.56) |
| *CD44* | C > T | rs187115 | 0.37700 | 1.23 (0.77–1.96) | 0.3132 | 1.27 (0.80–2.04) |
| *PDCD1* | A > G | rs7421861 | 0.00152 * | 0.49 (0.31–0.76) | 0.001514 ** | 0.49 (0.31–0.76) |
| *PDCD1* | C > T | rs10204525 | 0.02212 * | 2.11 (1.11–3.99) | 0.007101 ** | 2.49 (1.28–4.86) |

$p$-values * < 0.05; ** < 0.01; *** < 0.001.

Considering the possibility of a risk *PDCD1* haplotype, Linkage Disequilibrium analyses between the two *PDCD1* SNPs revealed a D′ value of 1 an $R^2$ of 0.0487329. The low $R^2$ value reflects the low frequency of the rs10204525 T allele (0.1) relative to the rs7421861 A allele (0.7). However, when the rarer rs10204525 T allele was present, it was always inherited with the more common rs7421861 A allele. These data and haplotype analyses are summarised in Supplementary Table S10, supporting the *PDCD1* rs7421861 "A" allele as being the one associated with greatest risk. It would appear, although numbers are small, that the relatively minor contribution of the *PDCD1* rs10204525 T allele, always inherited with the rs7421861 A allele, may be more relevant in women in the Newcastle cohort.

We went on to condition the analyses of the immunoregulatory SNPs on the *PNPLA3* and *TM6SF2* genotypes. The analyses yielded similar results, also shown in Table 2. In the conditioned univariate analysis, the protective effect of *PDCD1* rs7421861 minor allele remained strikingly significant, after adjusting for age, sex, cirrhosis and T2DM. For *PDCD1* rs10204525, where the presence of the variant allele acquired significance after adjusting for age, sex, cirrhosis and T2DM, the odds ratio and significance after conditioning on the fat regulatory genes was enhanced.

3.2.4. The Newcastle Cohort—Genotype Associations within the Cirrhotic Versus Non-Cirrhotic NAFLD-HCC Cases

Differences in genotype associations within cases classed as cirrhotic or NC were explored, and these are summarised in Table 3. Although the *PNPLA3* rs738409 variant G allele is undoubtedly more frequent in cancers compared to controls, there was also a significant difference between the cirrhotic versus the non-cirrhotic cases with NAFLD-HCC (red text Table 3, *p* < 0.0001). As *PNPLA3* rs738409 is associated with advanced NAFLD [35], this was not so surprising. Viewing the groups with the non cancer cases also broken down into those with and without cirrhosis confirmed that the most striking associations of both the *PNPLA3* and the *TM6SF2* variants were with the presence of cirrhosis, rather than with the presence of cancer. As previously reported, the HCC risk associated with these two SNPs is largely mediated by fibrosis, although in larger cohorts it is possible to discern an impact partially independent of fibrosis [18]. In contrast, the prevalence of the *PDCD1* alleles associated with elevated cancer risk (rs7421861 A allele and *PDCD1* rs10204525 T allele, green text Table 3) was clearly associated with HCC, in both the presence and absence of cirrhosis.

**Table 3.** Genotype distribution between cirrhotic and non-cirrhotic (NC) NAFLD and NAFLD-HCC patients.

| | | NAFLD Control | | | NAFLD-HCC | | |
|---|---|---|---|---|---|---|---|
| | | Total (%) | NC (%) | Cirrhotic (%) | Total (%) | NC (%) | Cirrhotic (%) |
| *PNPLA3* rs738409 | CC | 170 (40.9) | 156 (44.2) | 14 (22.2) | 67 (33.84) | 35 (46.1) | 32 (26.2) |
| | CG | 184 (44.2) | 150 (42.5) | 34 (54.0) | 85 (42.93) | 29 (38.2) | 56 (45.9) |
| | GG | 62 (14.9) | 47 (13.3) | 15 (23.8) | 46 (23.23) | 12 (15.8) | 34 (27.9) |
| *TM6SF2* rs58542926 | CC | 323 (77.6) | 281 (79.6) | 42 (66.7) | 132 (66.7) | 57 (74) | 75 (62.0) |
| | CT | 85 (20.4) | 67 (19) | 18 (28.6) | 60 (30.3) | 19 (24.7) | 41 (33.9) |
| | TT | 8 (2) | 5 (1.4) | 3 (4.8) | 6 (3) | 1 (1.3) | 5 (4.1) |
| CD44 rs187115 | TT | 175 (42.1) | 150 (42.5) | 25 (39.7) | 79 (39.9) | 27 (35.1) | 52 (43) |
| | CT | 197 (47.3) | 165 (46.7) | 32 (50.8) | 96 (48.5) | 40 (51.9) | 56 (46.3) |
| | CC | 44 (10.6) | 38 (10.8) | 6 (9.5) | 23 (11.6) | 10 (13) | 13 (10.7) |
| *MICA* rs2596542 | CC | 183 (44) | 159 (45) | 24 (38.1) | 92 (46.9) | 31 (41.3) | 61 (50.4) |
| | CT | 190 (45.7) | 158 (44.8) | 32 (50.8) | 89 (45.4) | 37 (49.3) | 52 (43) |
| | TT | 43 (10.3) | 36 (10.2) | 7 (11.1) | 15 (7.7) | 7 (9.3) | 8 (6.6) |
| *PDCD1* rs7421861 | AA | 180 (43.3) | 155 (43.9) | 25 (39.7) | 126 (63.6) | 50 (64.9) | 76 (62.8) |
| | AG | 189 (45.4) | 159 (45) | 30 (47.6) | 53 (26.8) | 22 (28.6) | 31 (25.6) |
| | GG | 47 (11.3) | 39 (11) | 8 (12.7) | 19 (9.6) | 5 (6.5) | 14 (11.6) |
| *PDCD1* rs10204525 | CC | 345 (82.9) | 289 (81.9) | 56 (88.9) | 154 (78.6) | 56 (73.7) | 98 (81.7) |
| | CT | 66 (15.9) | 59 (16.7) | 7 (11.1) | 38 (19.4) | 18 (23.7) | 20 (16.7) |
| | TT | 5 (1.2) | 5 (1.4) | 0 (0) | 4 (2) | 2 (2.6) | 2 (1.7) |

We went on to analyse genotypes in a subgroup restricted to controls and cases without cirrhosis. These analyses, including 353 NC NAFLD controls and 78 NC NAFLD-HCC patients, are summarised in Supplementary Table S11. In the NC cases, the *PDCD1* rs7421861 G allele was again protective against HCC. Furthermore, significance was retained after including age, gender and T2DM by regression, as well as after conditioning on *PNPLA3* and *TM6SF2*.

There were no overarching genotype associations with any particular architectural or histological subtype of HCC, although there were notable subtype associations within the NC HCC cases. Of 12 classed as SH-HCC, nine were homozygous and three heterozygous for the *PDCD1* rs7421861 "A" allele. Notably, 11/12 also carried the *CD44* rs187115 C allele. The clear cell subtype was rare, but both cases were homozygous for these named *PDCD1* and *CD44* alleles. The findings were statistically significant (*PDCD1* rs742861 *p* = 0.045;

*CD44* rs187115 *p*-0.009; Chi Square Exact test). While interesting, much larger numbers of cases with tumour histology would need to be systematically studied to attribute robust biological relevance to these observations.

### 3.3. Berne and Milan NAFLD and NAFLD-HCC Cohorts

Focusing on the predisposition to HCC development in patients with NAFLD, we went on to explore the fat and immunoregulatory SNPs in two additional European cohorts of control NAFLD and NAFLD-HCC cases, from Berne in Switzerland and Milan in Italy. The demographic and NAFLD relevant clinical data for the controls versus patients with HCC are presented in Table 4. In these smaller cohorts, not all the variables were significantly different. Although the trends were similar to the larger Newcastle cohort, there were some differences to note. Compared to Newcastle, the Berne control cases were similar, but the Berne NAFLD-HCC cases were younger, fewer were women and fewer had T2DM. The Milan control cohort was older, compared to Newcastle, with a lower BMI and a much greater prevalence of cirrhosis. The NAFLD-HCC Milan cohort was similar to that of Berne, with younger patients, with a lower BMI, less T2DM and a higher prevalence of cirrhosis.

**Table 4.** Demographic characteristics of Berne and Milan NAFLD and NAFLD-HCC cohorts.

| Phenotype | | Newcastle Cohort | | Berne Cohort | | | Milan Cohort * | | |
|---|---|---|---|---|---|---|---|---|---|
| | | NAFLD (n = 416) | NAFLD-HCC (n = 198) | NAFLD (n = 76) | NAFLD-HCC (n = 84) | p Value | NAFLD (n = 102) | NAFLD-HCC (n = 109) | p Value |
| Age (Mean ± SD) | | 53.0 ± 0.6 | 72.2 ± 0.7 | 53.7 ± 1.3 | 66.7 ± 0.9 | <0.0001 | 63.7 ± 1.1 | 66.9 ± 0.8 | 0.074 |
| Gender | male | 232 (55.8) | 157 (79.3) | 46 (60.5) | 80 (95.2) | <0.0001 | 47 (52.2) | 88 (81.5) | <0.0001 |
| (%) | female | 184 (42.8) | 41 (20.7) | 30 (39.5) | 4 (4.8) | | 43 (47.8) | 20 (18.5) | |
| BMI (Mean ± SD) | | 35.0 ± 0.3 | 32.0 ± 0.4 | 32.7 ± 0.6 | 29.6 ± 0.6 | <0.0001 | 30.0 ± 0.5 | 29.4 ± 0.5 | 0.445 |
| Diabetes | no | 198 (48.1) | 60 (30.3) | 35 (46.1) | 47 (56.0) | 0.211 | 34 (40.5) | 45 (44.6) | 0.577 |
| (%) | yes | 214 (51.9) | 138 (69.7) | 41 (53.9) | 37 (44.0) | | 50 (59.5) | 56 (55.4) | |
| Cirrhosis | no | 353 (84.9) | 77 (38.9) | 55 (72.4) | 15 (17.9) | <0.0001 | 6 (7.0) | 16 (15.5) | 0.068 |
| (%) | yes | 63 (15.1) | 121 (61.1) | 21 (27.6) | 69 (82.1) | | 80 (93.0) | 87 (84.5) | |

*p*-values estimated by Mann–Whitney or Chi Square tests for continuous or categorical data sets, respectively. * some categorical data unavailable.

Compared to Newcastle, the differences in the Milan cohort were at least in part due to differences in selection criteria, as these were patients recruited for a case–control study, as previously described [43]. The patients with NAFLD-HCC were representative of cases seen in that hospital between 2010 and 2016, while the NAFLD controls were not—these were recruited based on the presence of advanced fibrosis or cirrhosis in 2018–2020. The older age and lower BMI in the Milan controls may also be reflective of that selection bias. The Berne patients have been recruited from 2010, with both cases and controls being reflective of patients referred to that hospital. The differences in the cancer patients between the three institutions may reflect country differences. However, the largest cohort in Newcastle was recruited at a tertiary centre where all the patients regardless of age, fitness or stage of cancer were referred, from a large catchment area in the northeast of England. The cohorts from Berne and Milan may not have been as inclusive, potentially influenced by local referral practices (younger, fitter, with less diabetes and less obesity). As the Berne and Milan cohorts were relatively small compared to Newcastle, the differences were noted rather than interpreted further.

In the Berne and Milan cohorts, however, none of the allele frequencies—for either fat regulatory or immunoregulatory SNPs—was significantly different comparing NAFLD controls and NAFLD-HCC cases. The genotype frequencies are shown in Supplementary Table S12, with statistical analysis in Supplementary Table S13.

### 3.4. Newcastle, Berne and Milan NAFLD and NAFLD-HCC Cohorts Combined

The combined cohort data is summarised in Table 5. The meta-analysis for the *PNPLA3*, *TM6SF2* and the *PDCD-1* SNPs, also conditioned on *PNPLA3* and *TM6SF2* is shown in Table 6.

**Table 5.** The demographic characteristics of the combined European cohort.

| Combined Cohorts | Group | NAFLD (*n* = 594) | NAFLD-HCC (*n* = 391) | *p* Value |
|---|---|---|---|---|
| Age (Mean ± SD) | | 54.74 ± 0.50 | 69.52 ± 0.47 | <0.0001 |
| Gender | male (%) | 325 (55.8) | 325(83.1) | <0.0001 |
| | female (%) | 257 (44.2) | 66 (16.9) | |
| BMI (Mean ± SD) | | 33.99 ± 0.25 | 30.78 ± 0.30 | <0.0001 |
| Diabetes | no (%) | 267 (46.7) | 152 (39.6) | <0.030 |
| | yes (%) | 305 (53.3) | 232 (60.4) | |
| Cirrhosis | no (%) | 414 (71.6) | 108 (28.0) | <0.0001 |
| | yes (%) | 164 (28.4) | 278 (72.0) | |

**Table 6.** META analyses in the combined European cohort.

| Gene | rs Identity | *p* Value | OR | Condition on PNPLA3+TM6SF2 | |
|---|---|---|---|---|---|
| PNPLA3 | rs738409 | 0.043946 | 1.20 (1.00–1.43) | NA | NA |
| TM6SF2 | rs58542926 | 0.018524 | 1.37 (1.05–1.77) | NA | NA |
| PDCD1 | rs7421861 | 0.026279 | 0.79 (0.65–0.97) | 0.044663 | 0.81 |
| PDCD1 | rs10204525 | 0.123195 | 1.30 (0.93–1.83) | 0.137854 | 1.30 |
| | | Age, Gender, Cirrhosis, T2DM | | | |
| PNPLA3 | rs738409 | 0.231807 | 1.18 (0.90–1.55) | NA | NA |
| TM6SF2 | rs58542926 | 0.655423 | 0.91 (0.61–1.36) | NA | NA |
| PDCD1 | rs7421861 | 0.181639 | 0.82 (0.61–1.10) | 0.172542 | 0.81 (0.61–1.10) |
| PDCD1 | rs10204525 | 0.024180 | 1.90 (1.09–3.30) | 0.009843 | 2.13 (1.20–3.80) |

By univariate analysis, the SNPs in *PNPLA3*, *TM6SF2* and *PDCD1* rs7421861 were significantly different in NAFLD-HCC versus controls, but none retained significance after regression analysis including age, sex and cirrhosis. The *PDCD1* rs10204525 C > T variant again acquired significance in the regression analysis, including after conditioning on *PNPLA3* and *TM6SF2*.

### 3.5. Exploration of Functional Roles for PDCD1 rs7421861 and PDCD1 rs10204525

We have evaluated global effects of these genetic variants on *PD-1* expression by mining publicly available tissues data in the Ensembl Genome (https://www.ensembl.org/index.html (accessed on 10 March 2021)). We used multi-tissue expression quantitative trait locus (eQTL) mapping as a means of exploring whether the SNPs associate with altered gene expression in a given tissue [46–49]. The potential impact of the SNPs on transcript splicing was also evaluated by mining data from the GTEx portal (https://www.gtexportal.org/home/ (accessed on 10 March 2021) version 8), another publicly available resource. Splicing quantitative trait loci (sQTLs) are assigned for SNPs that are associated with altered or alternative splicing (AS), identified by Leafcutter [50] in the GTEx consortium project.

3.5.1. Elevated PD-1 Expression in the Presence of the PDCD1 rs7421861 "A" Major Allele

In Ensembl, there were 2765 data entries showing gene expression correlations with this *PDCD1* rs7421861, in a variety of tissues. We have assessed the effect of the variant "G" allele on the eQTL for *PD-1* expression, reported relative to "A" as the reference allele. Positive values indicate upregulation of the transcript in presence of the minor allele, while negative values indicate downregulation [51]. There were 357 changes significant at a

*p* value < 0.05. Restricting analysis to 84 data entries considering expression in immune cells, blood or spleen, *PD-1* was significantly downregulated in the presence of the minor allele (Table 7). Expression in the presence of the wild type *PDCD1* rs7421861 "A" allele, the one we have identified as associated with cancer risk, was therefore higher, with the minor allele associated with significant suppression. Note that besides *PD-1* expression, *PDCD1* rs7421861 was associated with altered expression of 5 other genes in locus vicinity (restricting analysis to effect size of $\geq 0.5$ or $\leq -0.5$, Supplementary Table S14).

**Table 7.** The eQTL effect size and significance for PDCD1 rs7421861 and PD-1 expression in indicated cell types, with the variant significantly associated with suppressed PD-1 expression.

| rs7421861 | | |
|---|---|---|
| ***p*-Value (-log10)** | **Effect Size** | **Tissue/Cell Type** |
| 2.028825 | −0.06357 | Whole_Blood |
| 2.737219 | −0.10859 | blood |
| 1.413284 | −0.21721 | macrophage_naive |
| 1.481194 | −0.22847 | monocyte_IAV |
| 2.160692 | −0.38254 | monocyte_Pam3CSK4 |
| 2.072197 | −0.39053 | monocyte_LPS |
| 2.878066 | −0.45225 | macrophage_Listeria |
| 4.672717 | −0.52111 | monocyte_R848 |

### 3.5.2. Elevated PD-1 Expression in the Presence of the PDCD1 rs10204525 Minor Allele

Similar to *PDCD1* rs7421861, there were 2652 gene expression data entries enabling the study of correlations with the *PDCD1* rs10204525 variant. Again, restricting the analysis to those focused on expression in immune cells, blood or spleen, with an effect size of $\geq 0.5$ or $\leq -0.5$; $p < 0.05$, there were 66 data entries. The SNP reportedly acts as a potential cis-eQTL for 11 genes within the locus (Supplementary Table S15). Overall, gene expression was upregulated in the presence of the minor allele *PDCD1* rs10204525 "T" allele, the one we have identified as associated with NAFLD-HCC risk, most notably in activated T cells and monocytes (Table 8).

**Table 8.** The eQTL effect size and significance for PDCD1 rs10204525 and PD-1 expression in indicated cell types, with the variant significantly associated with increased PD-1 expression.

| rs10204525 | | |
|---|---|---|
| ***p*-Value (-log10)** | **Effect Size** | **Tissue/Cell Type** |
| 8.853481 | 0.964331 | monocyte_Pam3CSK4 |
| 8.408515 | 0.96187 | monocyte_LPS |
| 9.49804 | 0.944919 | monocyte_naive |
| 2.34563 | 0.916861 | monocyte_CD16_naive |
| 5.266115 | 0.738328 | CD4_T-cell_anti-CD3-CD28 |
| 7.349027 | 0.735695 | monocyte_R848 |
| 4.761555 | 0.591227 | CD8_T-cell_anti-CD3-CD28 |
| 6.339877 | 0.581108 | monocyte_IAV |
| 1.772399 | −0.08286 | Whole_Blood |
| 5.520671 | −0.25549 | blood |

### 3.5.3. sQTLs Associated with PDCD1 rs7421861 and PDCD1 rs10204525

In the GTEx portal, *PDCD1* rs7421861 and rs10204525 were also identified as having a significant association with altered splicing, measured by the sQTL in different tissues. In particular, *PDCD1* rs7421861 was significantly associated with an altered intron excision ratio for long non-coding RNA LINC01238 in splenic tissue (NES = −0.78, *p* value = $3.3 \times 10^{-7}$), while *PDCD1* rs10204525 was significantly associated with an altered intron excision ratio for LINC01237 in whole blood (NES = 0.47, *p* value = $1.1 \times 10^{-7}$)

(Figure 2). Using the LINCs as candidate biomarkers of altered splicing in the region of the *PDCD1* locus, the data point towards the SNPs altering splicing factor binding sites.
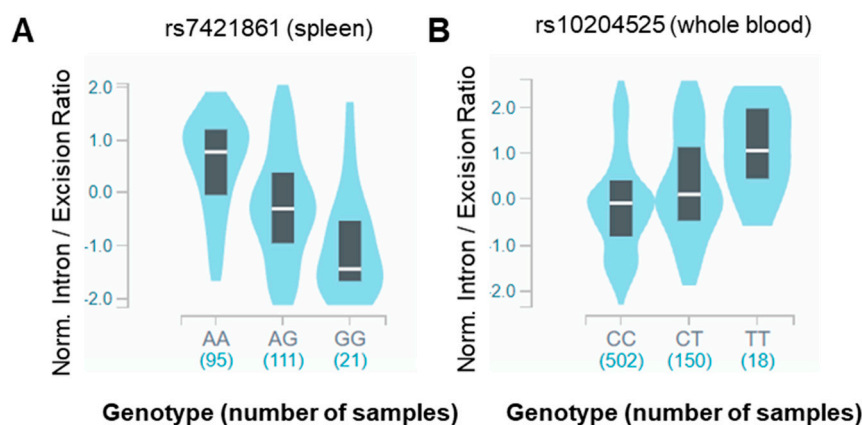


**Figure 2.** The figure shows sQTL violin plots for *PDCD1* rs7421861 on LINC01238 (intron location Chr2:241971462-241971680, hg38) in spleen (**A**) and *PDCD1* rs10204525 for LINC01237 (intron location Chr2: 241882635-241887744, hg38) in whole blood (**B**).

## 4. Discussion

As yet, there has been no GWAS study defining the genetic determinants of HCC arising in NAFLD, in either the presence or absence of cirrhosis. Of the NAFLD-HCC-associated candidate genes studied, the most robustly validated thus far have been those identified in GWAS studies looking for genes associated with NAFLD development. It is not a surprise, therefore, that these are variants in genes—like *PNPLA3* and *TM6SF2*—where the minor alleles alter protein function and promote steatosis, associated with both qualitative and quantitative alterations in hepatic fat content [52]. Although their impact on HCC predisposition is directly correlated with the promotion of NAFLD [18], these causal variants involved in fatty liver disease predisposition may also have pleiotropic effects, whereby they directly promote hepatic carcinogenesis by activating specific pathways. However, GWAS designed to identify genes elevating the risk of NAFLD would not necessarily capture those specifically promoting NAFLD-HCC. The candidate immunoregulatory genes studied here were purposefully selected.

A GWAS considering HCC risk in HCV previously identified a candidate immunoregulatory SNP in Major Histocompatibility Complex class I polypeptide related sequence A (MICA) [19]. MICA is a stress-induced cell surface antigen presenting peptide that binds and activates the NKG2D stimulatory receptor on the surface of NK cells and T cells, mediating their cytolytic immune surveillance. It has been reported that *MICA* rs2596542 C > T alters the expression of MICA, decreasing its membrane-bound form and increasing its soluble concentrations [53], favouring HCV progression to HCC in a Japanese population [19]. In our European NAFLD patients studied here, there was no significant difference between NAFLD control and NAFLD-HCC patients. This possibly supports *MICA* variation influencing the immune response and progression to fibrosis in HCV, as recently reported, rather than hepatocarcinogenesis [54].

Cluster of differentiation 44 (CD44) is a multifunctional transmembrane receptor expressed on various cell types, with ligand binding activating intracellular signalling pathways to stimulate cytokine release and lymphocyte activation [20]. The *CD44*rs187115 T > C SNP is located within intron 1, with the C allele reportedly associated with a higher risk of HBV-HCC development in Taiwan and China [21,55]. The mechanisms are yet to be determined, but a role in the p53 stress response has been proposed [22]. Our own interest in *CD44* arose having identified its expression in hepatic macrophages, in association with NC SH-HCC in murine models of NAFLD-HCC [56]. Although there was no significant differences in allele frequencies between NAFLD controls and NAFLD-HCC cases, the C

allele was noted to be significantly more common in SH-HCC arising in NC NAFLD. This was interesting, but the numbers of cases studied are presently small.

*PDCD1* encodes PD-1, a co-inhibitory receptor expressed most notably on activated T cells, with binding to the widely expressed ligands PD-L1 and PD-L2 suppressing activity and limiting potential damage to the host [24]. T cell receptor (TCR) activation stimulates the expression of nuclear factor of activated T cells (NFAT), a transcription factor that promotes IL-2-mediated T cell proliferation, but also binds the *PDCD1* promoter inducing expression of PD-1 [23]. Ongoing T cell activation with persistent PD-1 expression is associated with functional T cell exhaustion, with loss of T cell function and enhanced apoptosis. Tumour overexpression of PD-L1 is recognised as a means of supporting an immunosuppressive tolerant microenvironment and tumour immune evasion [57]. Sustained PD-1 expression, its decreased degradation, or expression of PD-L1 may also reportedly elevate individuals susceptible to HCC development [24,25,58,59].

A number of SNPs have been identified that may alter the expression and function of PD-1, with reports linking these to autoimmune disease [59], but also immunosuppressive conditions. *PDCD1* rs10204525 C > T is located at the 3′UTR, increasing PD-1 expression and reportedly promoting dysfunctional T cell responses and persistence in HBV infection [26–31,33,60]. The promotion of an immunosuppressive phenotype by *PDCD1* rs10204525 C > T may also contribute to tumour development but has not as yet been characterised. Our in silico analysis an influence of *PDCD1* rs10204525 on *PD-1* expression, being higher in the presence of the *PDCD1* rs10204525 "T" allele, with evidence in addition for a regulatory role in splicing. In our study, after adjusting for other cancer risks including gender, the variant "T" allele was significantly associated with NAFLD-HCC risk, regardless of the presence or absence of cirrhosis. As the T allele has low prevalence in European populations (~10%), the significance became much more noticeable in our cases after adjusting for gender—as over 35% of women with NAFLD-HCC carried the variant. Notably, the T allele is in fact the major allele in Asian populations, being present in 60–70% of people (NCBI dbSNP).

The rs7421861 A > G, the second *PDCD1* candidate SNP studied, is located in intron 1, where there are numerous known splicing and regulatory elements [24]. Although the function of this SNP has not been characterised, our in silico analysis indicates higher *PD-1* expression in association with the *PDCD1* rs7421861 A allele and also supports aberrant *PD-1* splicing as a means by which the G allele protects against cancers. An alteration in the prevalence of *PD-1* transcripts, with the G allele suppressing PD-1 mediated immune exhaustion and protecting against the development of cancers—including viral hepatitis HCC—has been previously suggested [24,30–33]. The data from the Newcastle cohort were in keeping with this, with the wild type "A" allele being significantly more common in NAFLD patients with cancers. This was not so apparent in the smaller cohorts from Berne and Milan, with fewer non-cirrhotic cases and fewer women.

Possibly the most important take home message from this paper, is that while the genetic variations in the genes influencing the environment in which HCC develops may differ in populations, there are likely to be pathways that are shared. Although we did not confirm these specific *PDCD1* SNP associations independently in the smaller replication cohorts from Berne and Milan, our study was limited to two SNPs and does not rule out a key role for dysregulation of PD-1 in these or other populations, at different levels as yet undefined. While the NAFLD predisposing genes promoting progression of fatty liver disease have been demonstrated to play a key role in NAFLD-HCC development, mainly mediated through the induction of steatohepatitis and fibrogenesis, genes regulating immunoregulatory pathways may have a role at least equally important—more so in those cases without significant fibrosis. This study highlights a potential role for PD-1 in susceptibility to HCC in NAFLD which is worthy of further study—especially as PD-1 is a druggable target. Furthermore, for patients with NAFLD-HCC, histological subtyping may carry information that will be important to define as we enter an era of immuno-oncology therapies. Going forward, particularly as emerging data suggest that those with NAFLD-

HCC may be less likely to respond to agents inhibiting PD-1/PD-L1 [61], understanding the immune environment is essential, as is identifying clinically useful tissues and blood-based biomarkers. Immunoregulatory SNPs have in the past directed liver disease-related therapies [62]. Those SNPs characterised here, or others, may yet have roles in stratifying surveillance or therapies for NAFLD-HCC. Their major contribution, however, may well be their highlighting key immunoregulatory mechanisms for further study.

## 5. Conclusions

We report that the genetic predisposition to HCC development in NAFLD may be influenced by immunoregulatory genes—most notably in this study, *PDCD1*—in addition to genes predisposing to increased fat accumulation and progression to advanced fibrosis and cirrhosis.

## References

1.  European Association for the Study of the Liver. Easl clinical practice guidelines: Management of hepatocellular carcinoma. *J. Hepatol.* **2018**, *69*, 182–236. [CrossRef]
2.  Chien, Y.C.; Jan, C.F.; Kuo, H.S.; Chen, C.J. Nationwide hepatitis b vaccination program in taiwan: Effectiveness in the 20 years after it was launched. *Epidemiol. Rev.* **2006**, *28*, 126–135. [CrossRef]
3.  Kudo, M.; Izumi, N.; Kubo, S.; Kokudo, N.; Sakamoto, M.; Shiina, S.; Tateishi, R.; Nakashima, O.; Murakami, T.; Matsuyama, Y.; et al. Report of the 20th nationwide follow-up survey of primary liver cancer in japan. *Hepatol. Res.* **2019**, *50*, 15–46. [CrossRef]
4.  European Association for the Study of the Liver. Easl 2017 clinical practice guidelines on the management of hepatitis b virus infection. *J. Hepatol.* **2017**, *67*, 370–398. [CrossRef] [PubMed]
5.  Pawlotsky, J.M.; Negro, F.; Aghemo, A.; Berenguer, M.; Dalgard, O.; Dusheiko, G.; European Association for the Study of the Liver. Easl recommendations on treatment of hepatitis c: Final update of the series. *J. Hepatol.* **2020**, *73*, 1170–1218. [CrossRef] [PubMed]
6.  Cancer Research UK. Cancer Mortality Statistics. Available online: https://www.cancerresearchuk.org/health-professional/cancer-statistics/mortality (accessed on 28 August 2020).
7.  Anstee, Q.M.; Reeves, H.L.; Kotsiliti, E.; Govaere, O.; Heikenwalder, M. From nash to hcc: Current concepts and future challenges. *Nat. Rev. Gastroenterol. Hepatol.* **2019**, *16*, 411–428. [CrossRef] [PubMed]
8.  Cheng, Y.J.; Imperatore, G.; Geiss, L.S.; Saydah, S.H.; Albright, A.L.; Ali, M.K.; Gregg, E.W. Trends and disparities in cardiovascular mortality among US adults with and without self-reported diabetes, 1988–2015. *Diabetes Care* **2018**, *41*, 2306–2315. [CrossRef] [PubMed]
9.  Frazer, K.; Callinan, J.E.; McHugh, J.; van Baarsel, S.; Clarke, A.; Doherty, K.; Kelleher, C. Legislative smoking bans for reducing harms from secondhand smoke exposure, smoking prevalence and tobacco consumption. *Cochrane Database Syst. Rev.* **2016**, *2*, CD005992. [CrossRef]
10. Younossi, Z.; Anstee, Q.M.; Marietti, M.; Hardy, T.; Henry, L.; Eslam, M.; George, J.; Bugianesi, E. Global burden of nafld and nash: Trends, predictions, risk factors and prevention. *Nat. Rev. Gastroenterol. Hepatol.* **2018**, *15*, 11–20. [CrossRef]
11. Geh, D.; Manas, D.M.; Reeves, H.L. Hepatocellular carcinoma in non-alcoholic fatty liver disease—A review of an emerging challenge facing clinicians. *Hepatobiliary Surg. Nutr.* **2021**, *10*, 59–75. [CrossRef] [PubMed]
12. Starley, B.Q.; Calcagno, C.J.; Harrison, S.A. Nonalcoholic fatty liver disease and hepatocellular carcinoma: A weighty connection. *Hepatology* **2010**, *51*, 1820–1832. [CrossRef]
13. Liu, Y.L.; Patman, G.L.; Leathart, J.B.; Piguet, A.C.; Burt, A.D.; Dufour, J.F.; Day, C.P.; Daly, A.K.; Reeves, H.L.; Anstee, Q.M. Carriage of the pnpla3 rs738409 C > G polymorphism confers an increased risk of non-alcoholic fatty liver disease associated hepatocellular carcinoma. *J. Hepatol.* **2014**, *61*, 75–81. [CrossRef]
14. Anstee, Q.M.; Targher, G.; Day, C.P. Progression of nafld to diabetes mellitus, cardiovascular disease or cirrhosis. *Nat. Rev. Gastroenterol. Hepatol.* **2013**, *10*, 330–344. [CrossRef]
15. Liu, Y.-L.; Reeves, H.L.; Burt, A.D.; Tiniakos, D.; McPherson, S.; Leathart, J.B.S.; Allison, M.E.D.; Alexander, G.J.; Piguet, A.-C.; Anty, R.; et al. Tm6sf2 rs58542926 influences hepatic fibrosis progression in patients with non-alcoholic fatty liver disease. *Nat. Commun.* **2014**, *5*, 4309. [CrossRef]
16. Tang, S.; Zhang, J.; Mei, T.-T.; Guo, H.-Q.; Wei, X.-H.; Zhang, W.-Y.; Liu, Y.-L.; Liang, S.; Fan, Z.-P.; Ma, L.-X.; et al. Association of tm6sf2 rs58542926 T/C gene polymorphism with hepatocellular carcinoma: A meta-analysis. *BMC Cancer* **2019**, *19*, 1128. [CrossRef]
17. Zhang, X.; Liu, S.; Dong, Q.; Xin, Y.; Xuan, S. The genetics of clinical liver diseases: Insight into the tm6sf2 e167k variant. *J. Clin. Transl. Hepatol.* **2018**, *6*, 326–331. [CrossRef]
18. Bianco, C.; Jamialahmadi, O.; Pelusi, S.; Baselli, G.; Dongiovanni, P.; Zanoni, I.; Santoro, L.; Maier, S.; Liguori, A.; Meroni, M.; et al. Non-invasive stratification of hepatocellular carcinoma risk in non-alcoholic fatty liver using polygenic risk scores. *J. Hepatol.* **2020**. [CrossRef]

19.　Kumar, V.; Kato, N.; Urabe, Y.; Takahashi, A.; Muroyama, R.; Hosono, N.; Otsuka, M.; Tateishi, R.; Omata, M.; Nakagawa, H.; et al. Genome-wide association study identifies a susceptibility locus for hcv-induced hepatocellular carcinoma. *Nat. Genet.* **2011**, *43*, 455–458. [CrossRef] [PubMed]

20.　Chen, C.; Zhao, S.; Karnad, A.; Freeman, J.W. The biology and role of cd44 in cancer progression: Therapeutic implications. *J. Hematol. Oncol.* **2018**, *11*, 64. [CrossRef]

21.　Chou, Y.-E.; Hsieh, M.-J.; Chiou, H.-L.; Lee, H.-L.; Yang, S.-F.; Chen, T.-Y. Cd44 gene polymorphisms on hepatocellular carcinoma susceptibility and clinicopathologic features. *Biomed. Res. Int.* **2014**, *2014*, 231474. [CrossRef]

22.　Endo, K.; Terada, T. Protein expression of cd44 (standard and variant isoforms) in hepatocellular carcinoma: Relationships with tumor grade, clinicopathologic parameters, p53 expression, and patient survival. *J. Hepatol.* **2000**, *32*, 78–84. [CrossRef]

23.　Bally, A.P.R.; Austin, J.W.; Boss, J.M. Genetic and epigenetic regulation of pd-1 expression. *J. Immunol.* **2016**, *196*, 2431–2437. [CrossRef]

24.　Salmaninejad, A.; Khoramshahi, V.; Azani, A.; Soltaninejad, E.; Aslani, S.; Zamani, M.R.; Zal, M.; Nesaei, A.; Hosseini, S.M. Pd-1 and cancer: Molecular mechanisms and polymorphisms. *Immunogenetics* **2018**, *70*, 73–86. [CrossRef] [PubMed]

25.　Tan, K.W.; Chacko, A.-M.; Chew, V. PD-1 expression and its significance in tumour microenvironment of hepatocellular carcinoma. *Transl. Gastroenterol. Hepatol.* **2019**, *4*, 51. [CrossRef]

26.　Ghorbani, P.; Mollaei, H.; Arabzadeh, S.; Zahedi, M. Upregulation of single nucleotide polymorphism of PD-1 gene (rs10204525) in chronic hepatitis B patients. *Int. Arch. Med. Microbiol.* **2019**, 2. [CrossRef]

27.　Chihab, H.; Jadid, F.Z.; Foka, P.; Zaidane, I.; El Fihry, R.; Georgopoulou, U.; Marchio, A.; Elhabazi, A.; Chair, M.; Pineau, P.; et al. Programmed cell death-1 3'-untranslated region polymorphism is associated with spontaneous clearance of hepatitis B virus infection. *J. Med. Virol.* **2018**, *90*, 1730–1738. [CrossRef] [PubMed]

28.　Huang, C.; Ge, T.; Xia, C.; Zhu, W.; Xu, L.; Wang, Y.; Wu, F.; Liu, F.; Zheng, M.; Chen, Z. Association of rs10204525 genotype gg and rs2227982 CC combination in programmed cell death 1 with hepatitis B virus infection risk. *Medicine* **2019**, *98*, e16972. [CrossRef]

29.　Zhang, G.; Li, N.; Zhang, P.; Li, F.; Yang, C.; Zhu, Q.; Han, Q.; Lv, Y.; Zhou, Z.; Liu, Z. PD-1 mrna expression is associated with clinical and viral profile and PD1 3'-untranslated region polymorphism in patients with chronic hbv infection. *Immunol. Lett.* **2014**, *162*, 212–216. [CrossRef]

30.　Zang, B.; Chen, C.; Zhao, J.-Q. PD-1 gene rs10204525 and rs7421861 polymorphisms are associated with increased risk and clinical features of esophageal cancer in a chinese han population. *Aging* **2020**, *12*, 3771–3790. [CrossRef]

31.　Dong, W.; Gong, M.; Shi, Z.; Xiao, J.; Zhang, J.; Peng, J. Programmed cell death-1 polymorphisms decrease the cancer risk: A meta-analysis involving twelve case-control studies. *PLoS ONE* **2016**, *11*, e0152448. [CrossRef]

32.　Huang, K.; Hu, E.; Li, W.; Lv, J.; He, Y.; Deng, G.; Xiao, J.; Yang, C.; Zhao, X.; Chen, L.; et al. Association of PD-1 polymorphisms with the risk and prognosis of lung adenocarcinoma in the northeastern chinese han population. *BMC Med. Genet.* **2019**, *20*, 177. [CrossRef]

33.　Hashemi, M.; Karami, S.; Sarabandi, S.; Moazeni-Roodi, A.; Małecki, A.; Ghavami, S.; Wiechec, E. Association between PD-1 and PD-l1 polymorphisms and the risk of cancer: A meta-analysis of case-control studies. *Cancers* **2019**, *11*, 1150. [CrossRef] [PubMed]

34.　Hardy, T.; Wonders, K.; Younes, R.; Aithal, G.P.; Aller, R.; Allison, M.; Bedossa, P.; Betsou, F.; Boursier, J.; Brosnan, M.J.; et al. The european nafld registry: A real-world longitudinal cohort study of nonalcoholic fatty liver disease. *Contemp. Clin. Trials* **2020**, *98*, 106175. [CrossRef]

35.　Anstee, Q.M.; Darlay, R.; Cockell, S.; Meroni, M.; Govaere, O.; Tiniakos, D.; Burt, A.D.; Bedossa, P.; Palmer, J.; Liu, Y.L.; et al. Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort. *J. Hepatol.* **2020**, *73*, 505–515. [CrossRef] [PubMed]

36.　Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.; Daly, M.J.; et al. Plink: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef]

37.　Liu, J.Z.; Tozzi, F.; Waterworth, D.M.; Pillai, S.G.; Muglia, P.; Middleton, L.; Berrettini, W.; Knouff, C.W.; Yuan, X.; Waeber, G.; et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat. Genet.* **2010**, *42*, 436–440. [CrossRef]

38.　Dudbridge, F. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum. Hered.* **2008**, *66*, 87–98. [CrossRef]

39.　Duan, X.Y.; Zhang, L.; Fan, J.G.; Qiao, L. Nafld leads to liver cancer: Do we have sufficient evidence? *Cancer Lett.* **2014**, *345*, 230–234. [CrossRef]

40.　Sanyal, A.J.; Harrison, S.A.; Ratziu, V.; Abdelmalek, M.F.; Diehl, A.M.; Caldwell, S.; Shiffman, M.L.; Aguilar Schall, R.; Jia, C.; McColgan, B.; et al. The natural history of advanced fibrosis due to nonalcoholic steatohepatitis: Data from the simtuzumab trials. *Hepatology* **2019**, *70*, 1913–1927. [CrossRef]

41.　Harrison, S.A.; Wong, V.W.; Okanoue, T.; Bzowej, N.; Vuppalanchi, R.; Younes, Z.; Kohli, A.; Sarin, S.; Caldwell, S.H.; Alkhouri, N.; et al. Selonsertib for patients with bridging fibrosis or compensated cirrhosis due to nash: Results from randomized phase iii stellar trials. *J. Hepatol.* **2020**, *73*, 26–39. [CrossRef]

42.　Dyson, J.; Jaques, B.; Chattopadyhay, D.; Lochan, R.; Graham, J.; Das, D.; Aslam, T.; Patanwala, I.; Gaggar, S.; Cole, M.; et al. Hepatocellular cancer: The impact of obesity, type 2 diabetes and a multidisciplinary team. *J. Hepatol.* **2014**, *60*, 110–117. [CrossRef] [PubMed]

43.  Pelusi, S.; Baselli, G.; Pietrelli, A.; Dongiovanni, P.; Donati, B.; McCain, M.V.; Meroni, M.; Fracanzani, A.L.; Romagnoli, R.; Petta, S.; et al. Rare pathogenic variants predispose to hepatocellular carcinoma in nonalcoholic fatty liver disease. *Sci. Rep.* **2019**, *9*, 3682. [CrossRef] [PubMed]

44.  Board WCoTE. Tumour of the liver and intrahepatic bile ducts. In *Digestive System Tumours*, 5th ed.; International Agency for Research on Cancer: Lyon, France, 2019; pp. 215–263.

45.  Falleti, E.; Cussigh, A.; Cmet, S.; Fabris, C.; Toniutto, P. Pnpla3 rs738409 and tm6sf2 rs58542926 variants increase the risk of hepatocellular carcinoma in alcoholic cirrhosis. *Dig. Liver Dis.* **2016**, *48*, 69–75. [CrossRef]

46.  Schadt, E.E.; Monks, S.A.; Friend, S.H. A new paradigm for drug discovery: Integrating clinical, genetic, genomic and molecular phenotype data to identify drug targets. *Biochem. Soc. Trans.* **2003**, *31*, 437–443. [CrossRef]

47.  Lappalainen, T.; Sammeth, M.; Friedlander, M.R.; Ac't Hoen, P.A.; Monlong, J.; Rivas, M.A.; Gonzalez-Porta, M.; Kurbatova, N.; Griebel, T.; Ferreira, P.G.; et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **2013**, *501*, 506–511. [CrossRef] [PubMed]

48.  Carithers, L.J.; Moore, H.M. The genotype-tissue expression (gtex) project. *Biopreserv. Biobank.* **2015**, *13*, 307–308. [CrossRef]

49.  Bahcall, O.G. Human genetics: Gtex pilot quantifies eqtl variation across tissues and individuals. *Nat. Rev. Genet.* **2015**, *16*, 375. [CrossRef]

50.  Li, Y.I.; Knowles, D.A.; Humphrey, J.; Barbeira, A.N.; Dickinson, S.P.; Im, H.K.; Pritchard, J.K. Annotation-free quantification of rna splicing using leafcutter. *Nat. Genet.* **2018**, *50*, 151–158. [CrossRef]

51.  Mohammadi, P.; Castel, S.E.; Brown, A.A.; Lappalainen, T. Quantifying the regulatory effect size of cis-acting genetic variation using allelic fold change. *Genome. Res.* **2017**, *27*, 1872–1884. [CrossRef]

52.  Romeo, S.; Sanyal, A.; Valenti, L. Leveraging human genetics to identify potential new treatments for fatty liver disease. *Cell Metab.* **2020**, *31*, 35–45. [CrossRef]

53.  Mohamed, A.A.; Elsaid, O.M.; Amer, E.A.; Elosaily, H.H.; Sleem, M.I.; Gerges, S.S.; Saleh, M.A.; El Shimy, A.; El Abd, Y.S. Clinical significance of snp (rs2596542) in histocompatibility complex class i-related gene a promoter region among hepatitis c virus related hepatocellular carcinoma cases. *J. Adv. Res.* **2017**, *8*, 343–349. [CrossRef]

54.  Sharkawy, R.E.; Bayoumi, A.; Metwally, M.; Mangia, A.; Berg, T.; Romero-Gomez, M.; Abate, M.L.; Irving, W.L.; Sheridan, D.; Dore, G.J.; et al. A variant in the mica gene is associated with liver fibrosis progression in chronic hepatitis c through tgf-beta1 dependent mechanisms. *Sci. Rep.* **2019**, *9*, 1439. [CrossRef] [PubMed]

55.  Pan, X.; Li, M.; Huang, L.; Mo, D.; Liang, Y.; Huang, Z.; Zhu, B.; Fang, M. Cd44, il-33, and st2 gene polymorphisms on hepatocellular carcinoma susceptibility in the chinese population. *Biomed. Res. Int.* **2020**, *2020*, 2918517. [CrossRef] [PubMed]

56.  Zaki, M.Y.W.; McCain, M.; Mhadi, A.K.; Mauricio, J.; Wilson, C.; Patman, G.L.; Whitehead, A.; Lunec, J.; Shukla, R.; Anstee, Q.M.; et al. OP-01: Genome Wide RNA Expression Analysis Identifies cd44 Positive Macrophages as Promoters of Hepatocyte Proliferation and the Development of Nafld-HCC. In *Proceedings of the EASL HCC Summit, Portugal, Lisbon, 14–16 February 2019*; Abstract Book; EASL: Portugal, Lisbon, 2019; p. 29.

57.  Gao, Q.; Wang, X.-Y.; Qiu, S.-J.; Yamato, I.; Sho, M.; Nakajima, Y.; Zhou, J.; Li, B.-Z.; Shi, Y.-H.; Xiao, Y.-S.; et al. Overexpression of pd-l1 significantly associates with tumor aggressiveness and postoperative recurrence in human hepatocellular carcinoma. *Clin. Cancer Res.* **2009**, *15*, 971. [CrossRef] [PubMed]

58.  Staron, M.M.; Gray, S.M.; Marshall, H.D.; Parish, I.A.; Chen, J.H.; Perry, C.J.; Cui, G.; Li, M.O.; Kaech, S.M. The transcription factor foxo1 sustains expression of the inhibitory receptor pd-1 and survival of antiviral CD8(+) t cells during chronic infection. *Immunity* **2014**, *41*, 802–814. [CrossRef]

59.  Boussiotis, V.A. Molecular and biochemical aspects of the PD-1 checkpoint pathway. *N. Engl. J. Med.* **2016**, *375*, 1767–1778. [CrossRef]

60.  Li, Z.; Li, N.; Zhu, Q.; Zhang, G.; Han, Q.; Zhang, P.; Xun, M.; Wang, Y.; Zeng, X.; Yang, C.; et al. Genetic variations of PD1 and tim3 are differentially and interactively associated with the development of cirrhosis and hcc in patients with chronic hbv infection. *Infect. Genet. Evol.* **2013**, *14*, 240–246. [CrossRef]

61.  Pinter, M.; Scheiner, B.; Peck-Radosavljevic, M. Immunotherapy for advanced hepatocellular carcinoma: A focus on special subgroups. *Gut* **2021**, *70*, 204–214. [CrossRef]

62.  Suppiah, V.; Moldovan, M.; Ahlenstiel, G.; Berg, T.; Weltman, M.; Abate, M.L.; Bassendine, M.; Spengler, U.; Dore, G.J.; Powell, E.; et al. Il28b is associated with response to chronic hepatitis c interferon-alpha and ribavirin therapy. *Nat. Genet.* **2009**, *41*, 1100–1104. [CrossRef] [PubMed]