

RESEARCH ARTICLE

Accumulation of Deleterious Passenger Mutations Is Associated with the Progression of Hepatocellular Carcinoma

Magdalena A. Budzinska^{1,2,3}, Thomas Tu^{1,2}, William M. H. d'Avigdor^{1,2}, Geoffrey W. McCaughan^{1,2,4}, Fabio Luciani³, Nicholas A. Shackel^{1,2,4*}

1 Centenary Institute, University of Sydney, Sydney, NSW, Australia, **2** Sydney Medical School, University of Sydney, Sydney, NSW, Australia, **3** School of Medical Sciences, The University of New South Wales, Sydney, NSW, Australia, **4** A.W. Morrow Gastroenterology and Liver Centre, Royal Prince Alfred Hospital, Sydney, NSW, Australia

☞ These authors contributed equally to this work.

* n.shackel@centenary.usyd.edu.au



OPEN ACCESS

Citation: Budzinska MA, Tu T, d'Avigdor WMH, McCaughan GW, Luciani F, Shackel NA (2016) Accumulation of Deleterious Passenger Mutations Is Associated with the Progression of Hepatocellular Carcinoma. PLoS ONE 11(9): e0162586. doi:10.1371/journal.pone.0162586

Editor: Diego Calvisi, Universitätsmedizin Greifswald, GERMANY

Received: April 28, 2016

Accepted: August 25, 2016

Published: September 15, 2016

Copyright: © 2016 Budzinska et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Sequence data from WES 1 has been deposited at the European Genome-phenome Archive (EGA), under study accession number PRJEB9907.

Funding: This project, T.T. and M.A.B. were funded by Sydney Foundation for Medical Research (Grant number 53030) NAS, <http://nfMRI.org.au/about-us/>.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

In hepatocellular carcinoma (HCC), somatic genome-wide DNA mutations are numerous, universal and heterogeneous. Some of these somatic mutations are drivers of the malignant process but the vast majority are passenger mutations. These passenger mutations can be deleterious to individual protein function but are tolerated by the cell or are offset by a survival advantage conferred by driver mutations. It is unknown if these somatic deleterious passenger mutations (DPMs) develop in the precancerous state of cirrhosis or if it is confined to HCC. Therefore, we studied four whole-exome sequencing datasets, including patients with non-cirrhotic liver (n = 12), cirrhosis without HCC (n = 6) and paired HCC with surrounding non-HCC liver (n = 74 paired samples), to identify DPMs. After filtering out putative germline mutations, we identified 187±22 DPMs per non-diseased tissue. DPMs number was associated with liver disease progressing to HCC, independent of the number of exonic mutations. Tumours contained significantly more DPMs compared to paired non-tumour tissue (258–293 per HCC exome). Cirrhosis- and HCC-associated DPMs do not occur predominantly in specific genes, chromosomes or biological pathways and the effect on tumour biology is presently unknown. Importantly, for the first time we have shown a significant increase in DPMs with HCC.

Introduction

Hepatocellular carcinoma (HCC) is a common cancer with 500,000–1,000,000 new cases annually, leading to ~600,000 deaths each year [1–3]. While surgical treatments are effective with early detection (70% 5-year survival), HCC diagnosis typically occurs in the late stages when no curative therapies exist [4–6], leading to a poor (<20%) 5-year survival rate [7, 8]. HCC typically occurs after decades of progressive chronic liver injury, caused by 3 main risk factors: (1)

chronic hepatitis B and C virus (HBV and HCV) infection; (2) chronic alcohol consumption; and (3) exposure to the food-borne mycotoxin aflatoxin B1 (AFB1) [9, 10].

As in other cancers, HCC is associated with the accumulation of genetic alterations in cancer driver genes. However, whole exome sequencing (WES) and whole genome sequencing (WGS) studies searching for genes responsible for tumour initiation have shown that HCC is a heterogeneous disease, and no driver mutation is necessary or sufficient for carcinogenesis [11–17]. For example, while mutations are commonly found in *hTERT*, *β-catenin*, and *p53*-dependent pathways [18–20], these mutations are also found in surrounding non-tumour tissue [21–24].

Much focus has been committed to identifying genetic variants common in different tumours or in HCC subtypes [25]. This approach ignores the majority of somatic variants unique to each patient, known as passenger mutations [26]. These stochastic mutations are more likely to be either neutral or deleterious than advantageous [27]. Passenger mutations observed in cancer biology are generally assumed to be neutral and to not play a role in cancer evolution. Deleterious passenger mutations (DPMs, defined as non-driver mutations that cause a deleterious effect on protein function) that confer a profound survival disadvantage would see the clone eliminated and thus are not easily detected. However, DPMs with only moderate effect may lead to changes in protein function that are tolerated due to a previously acquired survival advantage (provided, for example, by a driver mutation).

DPM accumulation has been observed in cancer mutations curated by Catalogue of Somatic Mutations in Cancer (COSMIC) and The Cancer Genome Atlas (TCGA), revealing that DPMs with moderate effect can evade deletion through selection and accumulate during the neoplastic progression [28]. While these studies have focused on patients in whom cancer has already occurred, we and others have shown that significant clonal expansion of histologically normal cells occurs prior to carcinogenesis in patients with procarcinogenic diseases, including chronic HBV infection, a major risk factor for HCC [29, 30]. Therefore, DPMs could also accumulate in precancerous liver tissues.

We hypothesise that DPMs progressively accumulate in the liver during injury progression to HCC. Further, the presence and frequency of DPMs may be a potential marker that can help estimate risk of HCC or help understand the pathobiology of the premalignant state. Here, we have analysed WES datasets of tumour and matched non-tumour adjacent liver tissue controls of HCC patients with differing aetiologies [11, 12, 17]. Further, we have generated a WES dataset of liver tissue from patients without overt liver injury and cirrhotic patients without HCC. Our results are consistent with the hypothesis that DPMs frequency increases with progression towards HCC and therefore may help identify individuals at risk of HCC.

Materials and Methods

Ethics Statement

Human tissue samples were obtained from Royal Prince Alfred Hospital, Sydney, Australia with approval of Human Research Ethics Committee of the Royal Prince Alfred Hospital (Protocol number X10-0072). Informed written consent was obtained from all participants.

Whole exome sequencing (WES) datasets

The WES 1 dataset included liver tissue from 12 patients with limited levels of liver injury and 6 HCV-positive patients with liver cirrhosis (Patient characteristics are shown in Table 1). Briefly, snap-frozen liver wedge biopsies of donor (HCV-negative) and recipient (HCV-positive) liver tissue were taken during liver transplants at the Royal Prince Alfred Hospital (RPAH), Sydney. Total DNA was extracted as previously described [31]. DNA was prepared

Table 1. Clinical characteristics of patients analysed in WES 1: non-HCC liver injury samples.

	Sample	Sex	Age	HCV infection	HCV Genotype	METAVIR score
Non-cirrhotic patients (n = 12)	NC1	M	28	No	NA	0–1
	NC2	M	13	No	NA	0–1
	NC3	M	27	No	NA	0–1
	NC4	M	17	No	NA	0–1
	NC5	M	18	No	NA	0–1
	NC6	M	18	No	NA	0–1
	NC7	M	60	No	NA	0–1
	NC8	F	53	No	NA	0–1
	NC9	F	22	No	NA	0–1
	NC10	F	55	No	NA	0–1
	NC11	M	18	No	NA	0–1
	NC12	M	68	No	NA	0–1
Cirrhotic HCV-positive patients without HCC (n = 6)	C1	M	54	Yes	3a	4
	C2	M	57	Yes	3a	4
	C3	F	61	Yes	3a	4
	C4	M	45	Yes	3b	4
	C5	M	53	Yes	1b	4
	C6	M	59	Yes	1a	4

doi:10.1371/journal.pone.0162586.t001

for WES using the Agilent SureSelect Human All Exon 51M enrichment kit by BGI Hong Kong. Sequence data has been deposited at the European Genome-phenome Archive (EGA), accession number PRJEB9907. Data for WES 2–5 were taken from previously published studies [11, 12, 17, 32]. Further information is available in [S1 Supplementary Methods](#).

Bioinformatics analysis pipeline

Details on alignment and variant filtering are shown in [Fig 1](#) and described in greater depth in [S1 Supplementary Methods](#). All variants were annotated using ANNOVAR [33] with UCSC Known Gene annotation to determine the amino acid changes. Probable germline mutations

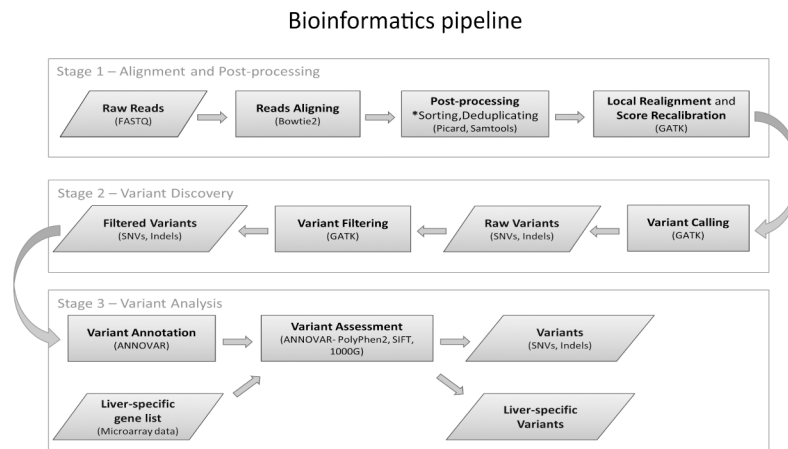


Fig 1. Bioinformatics analysis pipeline. Each resultant data file is indicated by a sloped rectangle and each process represented by a square rectangle. Our pipeline contains 3 stages: alignment and calibration; variant calling and filtering; and variants annotation and filtration of putative germline mutations.

doi:10.1371/journal.pone.0162586.g001

were excluded by filtering out variants present in 1000 Genomes Project database [34] (v1000g2014oct). The allelic frequency of each SNV was estimated by dividing the number of reads carrying the specific SNV by the number of total reads at that position.

Variants were classified as DPMs if they met one of the following criteria:

1. missense SNVs judged as “probably damaging” or “possibly damaging” by the PolyPhen-2 algorithm [35] (PolyPhen-2 score ≥ 0.453) or by the SIFT algorithm [36] (SIFT score ≤ 0.05);
2. stop-gain or stop-loss mutations
3. frameshift indels.

Missense variants that lay outside of these criteria were classed as benign. Due to their unknown effect, non-frameshift indels were classed neither as DPMs nor benign mutations and excluded from further analysis.

Analysis of liver-expressed genes

A list of genes expressed in the liver was generated from analysis of microarray gene expression data generated in our laboratory from total RNA extracts of non-diseased liver tissue of 6 donors (S2 Table). Specific details on analysis are given in [S1 Supplementary Methods](#).

Statistical analysis

Statistical analyses were carried out using PRISM 6 software (GraphPad, La Jolla, USA). The Wilcoxon matched-pairs signed-rank test was used to assess the differences between each set of paired samples (tumour vs. non-tumour) and the Mann-Whitney U test was used for unpaired samples and comparison of datasets. The association of DPMs relative to the occurrence of putative driver mutations was analysed by Spearman rank correlation coefficient test.

Pathway and functional enrichment analysis

The Ingenuity Pathway Analysis (IPA, Ingenuity Systems, Mountain View, CA; <http://www.ingenuity.com>) was used to identify the pathways and biological functions of genes affected by DPMs. The significance was set at a p-value of 0.01 by the right-tailed Fisher Exact Test.

Results

Normalisation and identification of somatic DPMs

The bioinformatics pipeline, outlined in [Fig 1](#), was used to analyse the number of exonic variants in datasets derived from liver tissue DNA (WES 1–4) and serum DNA (1000G and WES 5). Overviews of the datasets are provided in [Table 1](#) (WES 1), [Table 2](#) (WES 2–4) and [S1 Table](#) (1000G). Expectedly, the number of detectable variants differed with each dataset ([Fig 2A and 2B](#)), reflecting factors such as different enrichment kits, sequencing platforms and sequencing depth. Therefore, we normalised values to the total exonic mutations for each tissue sample to reduce inter-dataset and inter-patient variation in subsequent analyses.

We excluded potential germline mutations using the 1000 Genomes Project data. While peripheral blood mononuclear cells (PBMC) of the same patient are often used as a control for germline mutations, a number of important confounders are evident with this approach. Firstly, we had found that the mutational profile of PBMCs differs with liver injury, likely due to clonal expansion of circulating immune cells during inflammation associated with liver disease ([S2 Fig](#)). Further, somatic mutations in PBMCs acquired with age are not accounted for

Table 2. Summary data of publicly-available WES datasets used in this study.

WES dataset	Ref.	Aetiology	n	% Male	Mean depth	Mean read length	Mean read count per sample (millions)	Platform	Enrichment kit	Location
WES 2	[11]	Alcohol (50%) HBV (4%) HCV (17%) NASH (8%) Other (29%)	24	83	x 73	Paired-end 75 bp	132.7	Illumina HiSeq2000	SureSelect Human All Exon Kit v2 (44Mb)	France
WES 3*	[12]	HBV (43%) HCV (21%)	72* 10 5	77	x 59 x 4.8 x 5.8	Paired-end 100 bp 76 bp 76 bp	62.6*	Illumina HiSeq2000 Illumina GAllx Illumina GAllx	SureSelect Human All Exon v4(51Mb) All exon v1 (38 Mb) NimbleGen Human Exome v1 (2.1Mb)	USA/ Canada
WES 4*	[17]	NR	30*	NR	NR	Paired-end 75 bp	184.0*	Illumina HiSeq	NR	USA

*25 paired samples were used in the analysis from each of these studies to allow dataset comparisons.

NR = Not reported

doi:10.1371/journal.pone.0162586.t002

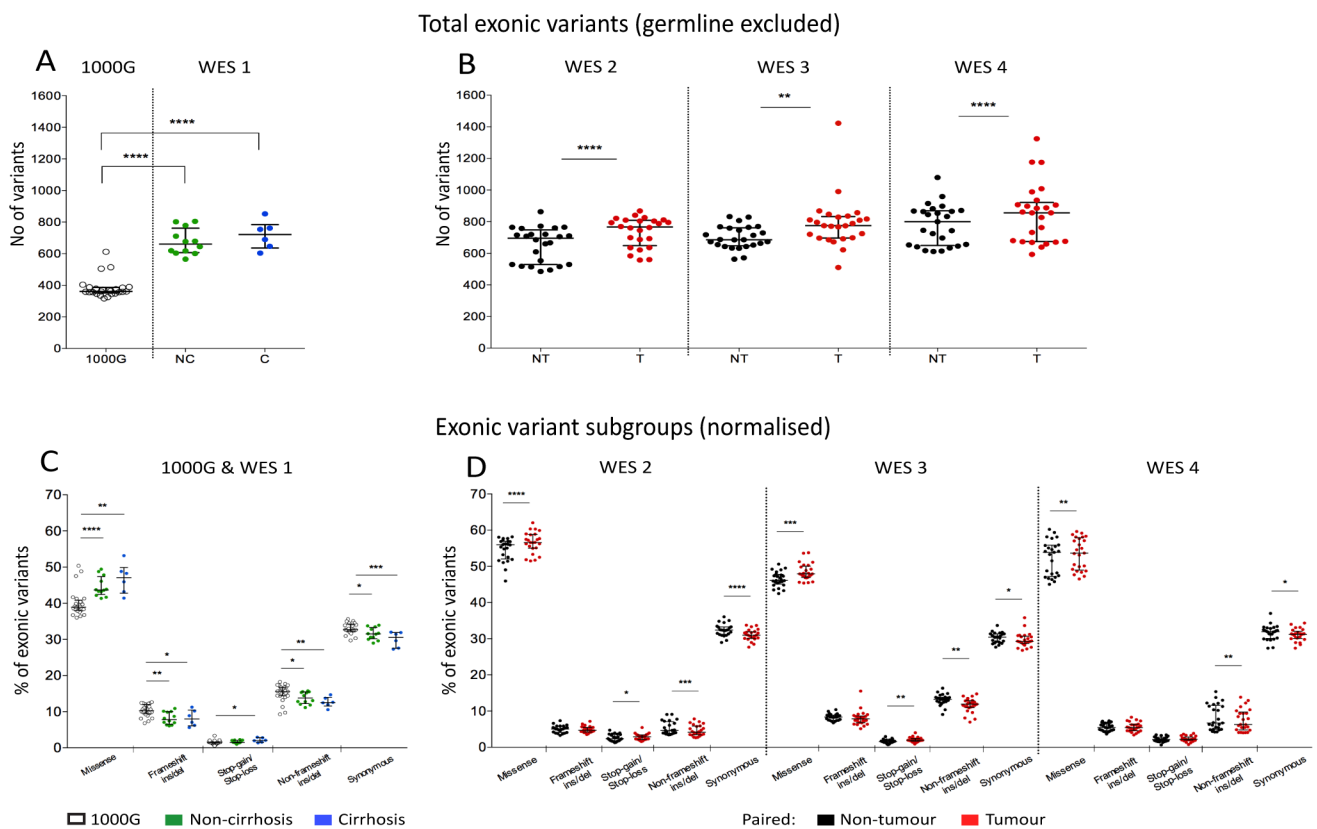


Fig 2. Absolute number of exonic variants and mutation subtypes in 1000G, liver injury, cirrhosis and HCC. The exonic variants in each of the 5 datasets were enumerated (A and B) and then subdivided into 5 groups (missense, frameshift ins/del, stop-gain/-loss and non-frameshift ins/del) (C and D, expressed as a percentage of all somatic exonic mutations). 1000G and WES 1 (A and C) contain unpaired samples, while WES 2–4 (B and D) are composed of paired tumour and non-tumour samples taken from the same individual. Data are expressed as median (interquartile range). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and **** $p < 0.0001$, Mann-Whitney U test (1000G and WES 1) or Wilcoxon matched-pairs signed-rank test (WES 2–4). NC-non-cirrhosis; C-cirrhosis; NT-non-tumour; T-tumour.

doi:10.1371/journal.pone.0162586.g002

and may be incorrectly assumed to be germline. Finally, previously described tissue differences in somatic mutation rates and profiles may be missed [37, 38]. Thus, germline mutations were imputed using the 1000 Genomes Project data and all samples were filtered identically. The number of excluded variants was not significantly different between tumour and non-tumour samples ($p > 0.05$, Wilcoxon signed-rank test). After filtration, only few of the variants (2.4%, 4.7%, 5.1%, 5.1%, 2.1%, 3.7% for WES 1–5, and 1000G respectively) occurred at an allelic frequency of 1.0 (S3 Fig), suggesting that the majority of homozygous germline variants have been excluded and the remainders were likely to be somatic variants.

While the numbers of somatic exonic variants (either single nucleotide variants (SNVs) or small indels) between non-cirrhotic and cirrhotic patients were not significantly different (Fig 2A, $p > 0.05$, Mann-Whitney U test), greater numbers of variants were detected in tumour compared to non-tumour tissue (Fig 2B, $p < 0.0001$, $p < 0.01$, and $p < 0.0001$ for WES 2–4 respectively, Wilcoxon signed-rank test). However, the absolute number of mutations did not consistently separate tumour and non-tumour tissue.

Description of variants and deleterious passenger mutations (DPMs) in liver injury and HCC

Exonic variants were then classified based on their effect on open reading frames (i.e. missense mutations, stop-gain/-loss mutations, and indels with or without a frame-shift). An increase in missense and synonymous mutations was observed with liver disease progression (S4 Fig). After normalisation of each sample to the number of exonic variants (Fig 2C and 2D), we found a consistent increase of missense mutations (a mean relative increase of 3.5%, 4.3%, and 2.2% from non-tumour to tumour in WES 2–4).

To test the hypothesis that DPMs accumulate in the development of HCC, we examined the percentage of benign SNVs and DPMs in 1000G, liver injury, cirrhosis and paired non-tumour and tumour samples (Fig 3). Our classification of benign SNVs and DPMs is shown in Fig 3A. Briefly, exonic mutations predicted to affect protein function (including stop-gain/-loss, frame-shift mutations and those judged to be damaging by PolyPhen2 or SIFT algorithms) were classified as DPMs. Benign missense SNVs were also classified by PolyPhen2 or SIFT algorithms. In the majority of patients (91.7%, 72% and 88% for WES 2, WES 3 and WES 4 respectively), more DPMs were observed in tumours compared to surrounding non-tumour tissue (Fig 3B–3E, $p < 0.01$, Wilcoxon signed-rank test). The mean relative increase in DPMs was 7.1%, 7.8% and 4.4% from non-tumour to tumour samples for WES 2–4, respectively. However, no significant differences were observed in benign missense SNVs, suggesting that the observed accumulation occurred specifically in DPMs. Similar results were seen when the SIFT algorithm was used (S6 Fig).

We estimated the allelic frequency of benign missense SNVs and DPMs in each patient by the ratio of wild type to mutated reads. The allelic frequency distributions of variants were similar between benign missense SNVs and DPMs for any given patient or disease stage including HCC (S3 Fig). Further, using available clinical data, we showed that DPM accumulation did not significantly correlate with patient age, cause of liver disease or tumor size (WES 2: $R^2 = 0.09$ – 0.12 and $p = 0.24$ – 0.3). Therefore, DPMs appear to accumulate from the non-tumour to tumour progression of HCC irrespective of a range of clinical features, and so may represent a general phenomenon in hepatocarcinogenesis.

DPM accumulation was observed even when the analysis was restricted to genes expressed in the liver (S5 Fig and S2 Table). As genes containing some DPMs may not be expressed (and so do not alter cell phenotype), we excluded mutations within genes not expressed in liver tissue (S2 Table). After filtration, significantly more DPMs (but not benign missense SNVs) were still

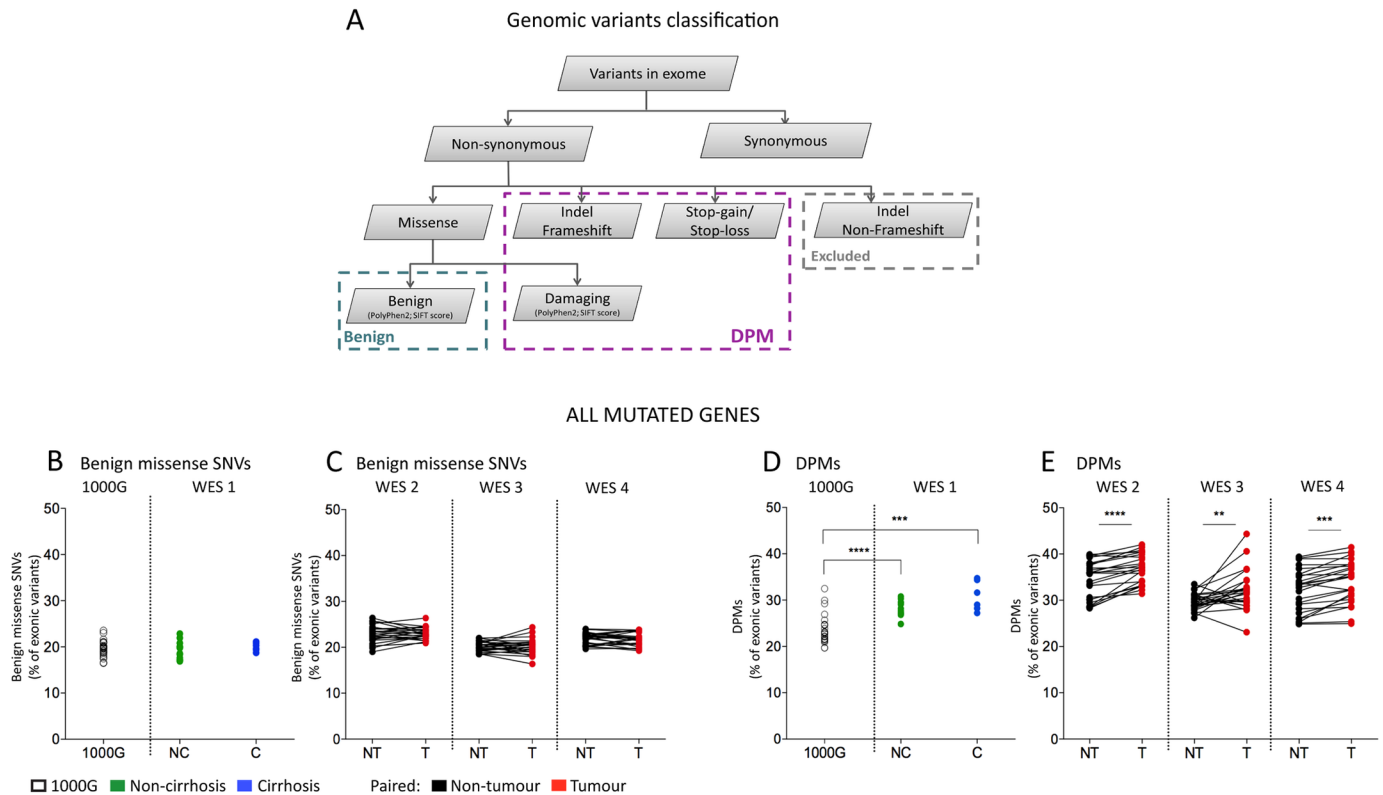


Fig 3. DPMs in HCC and surrounding non-tumour tissue. Variants were classified based on the predicted effect on the amino acid sequence (A). Total benign missense variants (B and D) and DPMs (C and E) in the datasets 1000G and WES 1–4 are shown as a percentage of all somatic exonic mutations. Significantly more DPMs (but not benign missense SNVs) were detected in tumour compared to paired non-tumour tissue (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and **** $p < 0.0001$, Wilcoxon matched-pairs signed-rank test). Lines link matched non-tumour and tumour tissues samples. NC-non-cirrhosis; C-cirrhosis; NT-non-tumour; T-tumour.

doi:10.1371/journal.pone.0162586.g003

observed in tumours compared to surrounding non-tumour tissue (S5D Fig, $p < 0.01$, Wilcoxon signed-rank test). Further, DPM accumulation in patients without HCC (WES 1) was significantly lower compared to both tumour (for WES 2 and 4) and non-tumour samples (for WES 2) in HCC patients (Table 3). In summary, the accumulated DPMs potentially generate a novel phenotype within the liver cells containing them due to alterations in encoded protein function.

The accumulation of DPMs in non-tumour tissue and its relationship with putative driver mutations

We tested the possibility that increasing DPMs were associated with accumulation of putative driver mutations and the development of HCC. We did not see a consistent association after analysis of putative HCC driver mutations in HCC patient datasets (Fig 4). Putative HCC driver genes were defined in this case as the 20 most frequently mutated genes in HCC tissues as retrieved from the COSMIC database (listed in S3 Table). The least frequent driver mutation in this list occurs at ~2%, and thus would not be expected to occur in our dataset more than once. These putative driver mutations occurred between 0 to 2 times per tissue, consistent with previous studies showing that drivers are relatively rare and that passenger mutations outnumber them by up to 2 orders of magnitude [39, 40].

Putative driver mutations were seen in both tumour and non-tumour tissue (S3 Table). Although, we observed both damaging and benign mutations in putative HCC driver genes

Table 3. Summary statistics for normalised DPMs between datasets.

		Total			
		1000G	WES 1		
			NC	C	
		Mean (±SD)	23.60 (±3.18)	28.17 (±1.71)	30.85 (±3.22)
1000G		23.60 (±3.18)	NA	p<0.0001	p<0.001
WES 2	NT	34.34 (±4.07)	p<0.0001	p<0.0001	p<0.05
	T	36.78 (±3.19)	p<0.0001	p<0.0001	p<0.01
WES 3	NT	29.76 (±1.8)	p<0.0001	p<0.05	p = 0.57
	T	32.07 (±4.27)	p<0.0001	p<0.001	p = 0.49
WES 4	NT	32.11 (±4.7)	p<0.0001	p<0.05	p = 0.55
	T	33.54 (±5.23)	p<0.0001	p<0.01	p = 0.12
		Liver-specific			
		1000G	WES 1		
			NC	C	
		Mean (±SD)	8.41 (±1.87)	14.56 (±1.48)	16.11 (±2.18)
1000G		8.41 (±1.87)	NA	p<0.0001	p<0.0001
WES 2	NT	19.35 (±2.54)	p<0.0001	p<0.0001	p<0.01
	T	20.80 (±2.34)	p<0.0001	p<0.0001	p<0.001
WES 3	NT	14.92 (±1.5)	p<0.0001	p = 0.44	p = 0.25
	T	16.44 (±2.66)	p<0.0001	p<0.05	p = 0.98
WES 4	NT	17.60 (±3.08)	p<0.0001	p<0.01	p = 0.19
	T	18.54 (±2.98)	p<0.0001	p<0.001	p<0.05

doi:10.1371/journal.pone.0162586.t003

repeated in HCC tissue, mutations in the majority of these genes (except for *CTNNB1* in dataset WES 2 and *TP53* in datasets WES 2, 3 and 4) were also observed at similar frequencies in the surrounding non-tumour tissue (S3 Table). Further, the average allelic frequency (as estimated by the ratio of wild type to mutated reads) of the mutations in the putative driver genes did not appear to differ between tumour and non-tumour samples (data not shown).

In the non-tumour tissue of WES 2 (but not WES 3 or 4), we observed a significantly greater proportion of DPMs with an increasing number of damaging mutations in driver genes (Fig 4B, p<0.0001, p = 0.095, and p>0.1 respectively, Spearman rank correlation coefficient test). We repeated this analysis on the tumour tissue and observed no significant association between detected driver mutations and either benign missense SNVs or DPMs (Fig 4C and 4D, p>0.1, Spearman rank correlation coefficient test). This was expected, as all tumours presumably have gained sufficient driver mutations (though not observable using the NGS data) to have proceeded to HCC. As a control, we performed the same analysis (n = 10), but using 20 randomly selected genes containing DPMs instead of known driver genes and observed no significant DPM increase in any datasets (data not shown). Together, these findings are consistent with the hypothesis that the surrounding non-tumour tissue is not necessarily normal and can contain precancerous changes.

The majority of DPMs are likely to be true passenger mutations

The majority of DPMs seen were not shared between patients (Fig 5). Pooled DPMs from all samples for each dataset were analysed to determine if the accumulated DPMs represented potential novel driver genes. We found that the majority (>70% for each dataset) of genes with DPMs were not recurrent and instead DPMs occurred in unique locations for each patient (Fig 5), consistent with random accumulation. Further, the chromosomal distribution

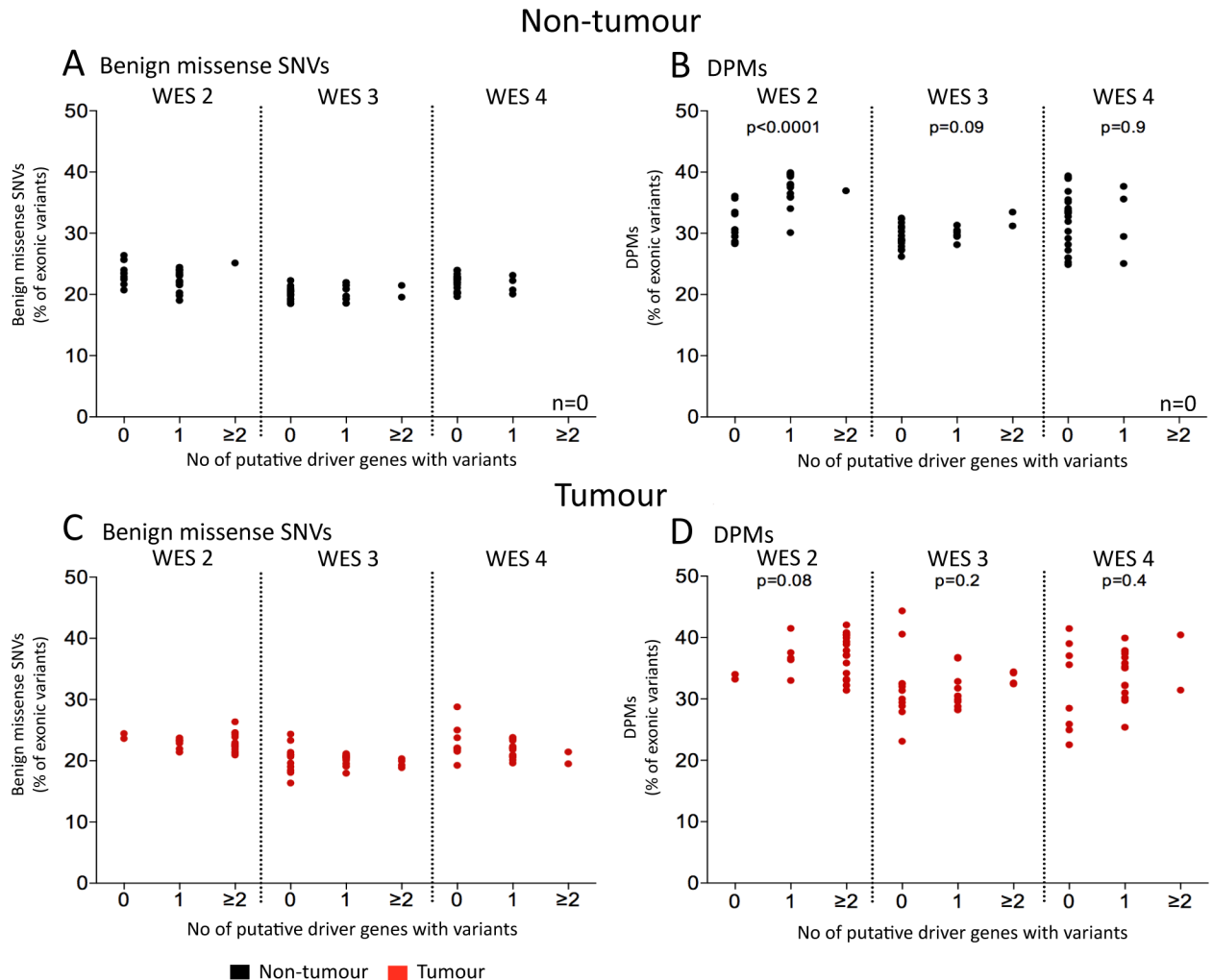


Fig 4. Driver mutations in non-tumour tissue. Patient samples were separated based on the number of mutations in putative driver genes (x-axis, defined as the 20 top recurrently mutated genes in HCC according to COSMIC database, listed in [S3 Table](#)) and analysed the number of benign missense SNVs (A and C) and DPMs (B and D). Significant correlation between DPMs and putative driver mutations ($p < 0.0001$, Spearman rank correlation test) was observed in non-tumour tissue of WES 2. No significant correlation was seen in HCC tissues ($p > 0.05$, Spearman rank correlation test).

doi:10.1371/journal.pone.0162586.g004

of DPMs ([S7 Fig](#)) showed broad occurrence throughout the genome, without any obvious hotspots.

Pathway enrichment analysis showed that there was significant enrichment ($p < 0.01$, right-tailed Fisher Exact Test) of DPMs in some functional biological pathways in both tumour and non-tumour samples ([S4 Table](#)). However, only a minority of DPMs contributed to these pathways: 0%, 0.61%, 1.7%, 0.36%, 0.44% in non-tumour tissues in WES 1, 2, 3, and 4, respectively; and 2.8%, 2.4%, 0.49%, in tumour tissues of WES 2, 3, and 4, respectively. Even if these DPMs in these functional pathways all represented novel driver mutations, this is still insufficient to explain the increase in DPMs associated with liver disease progression, which had a mean relative increase in DPMs of 7.1%, 7.8% and 4.4% from non-tumour to tumour tissues in WES 2, 3, and 4, respectively. In summary, these results suggest that the increased frequency of DPMs in tumour compared to non-tumour is due to stochastic accumulation of passenger mutations.

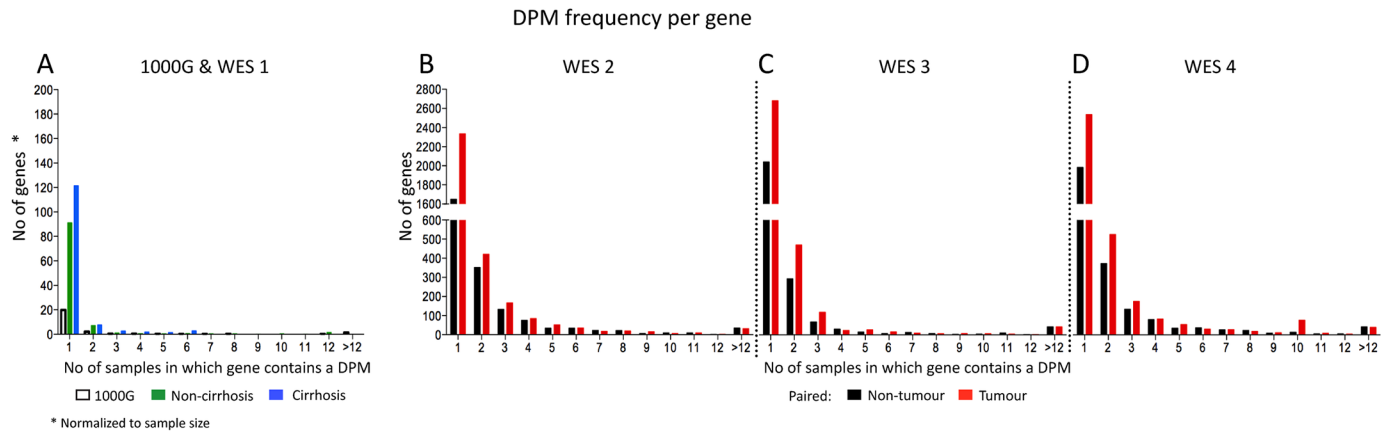


Fig 5. Frequency distribution of DPMs. A frequency distribution of the genes containing DPMs in 1000G and WES 1 (A), WES 2 (B), WES 3 (C), and WES 4 (D) shows that most are unique to a given patient. Each gene containing a DPM was grouped based on the number of patients in which that gene contained a DPM (x-axis).

doi:10.1371/journal.pone.0162586.g005

Further, the difference in DPM load between tumour and non-tumour samples does not likely represent a gain in novel driver mutations.

Discussion

This is the first NGS study to our knowledge to DNA sequence normal liver tissue and recognise that there are exome-wide DNA alterations in liver tissues prior to carcinogenesis. Our focus was on DPMs (defined as randomly-acquired somatic mutations that altered protein function), which composed of approximately a third of all somatic variants. Our key finding shows that an increase in DPMs is associated with progressively worse liver disease leading up to HCC. This was also observed even when genes not expressed in liver tissue (measured by microarray analysis) were excluded.

DPMs could be promoting tumour development in these tissues, but we could not find evidence of this occurring. The majority of DPMs were not found to occur predominantly in any specific genes, chromosomes or biological pathways. While this may be explained in part by the poor recognition and understanding of such pathways, given the rarity of tumour suppressors and the observed overall progressive accumulation of DPMs, our data would suggest that DPMs are randomly acquired and true passenger mutations rather than uncharacterised drivers of HCC.

The pattern of observed DPMs in HCC is consistent over multiple algorithms for scoring deleterious effect, in multiple aetiologies of HCC, and in multiple datasets with different ethnic compositions. Further, the observed DPMs are not at a low frequency (S3 Fig), which would be seen in sporadic occurrences, as they have the same overall allelic frequency as benign missense mutations. This suggests that DPM accumulation is a general mechanism accompanying tumour evolution and agrees with the theory that DPMs accumulate during the evolution of preneoplastic HCC subclones [28]. Our comparison with DNA extracted from normal tissue suggests that genetic changes have occurred in non-tumour hepatocytes in patients with HCC. These results are consistent with mathematical models suggesting that >50% of exonic mutations occur prior to carcinogenesis [41] and observations of *TERT* promoter mutations in preneoplastic nodules [42]. Here, we extend these studies showing that many driver mutations are observed in histologically-normal non-tumour tissue. Mutations that occur prior to tumourigenesis should not be ignored as they may contribute to the carcinogenic process.

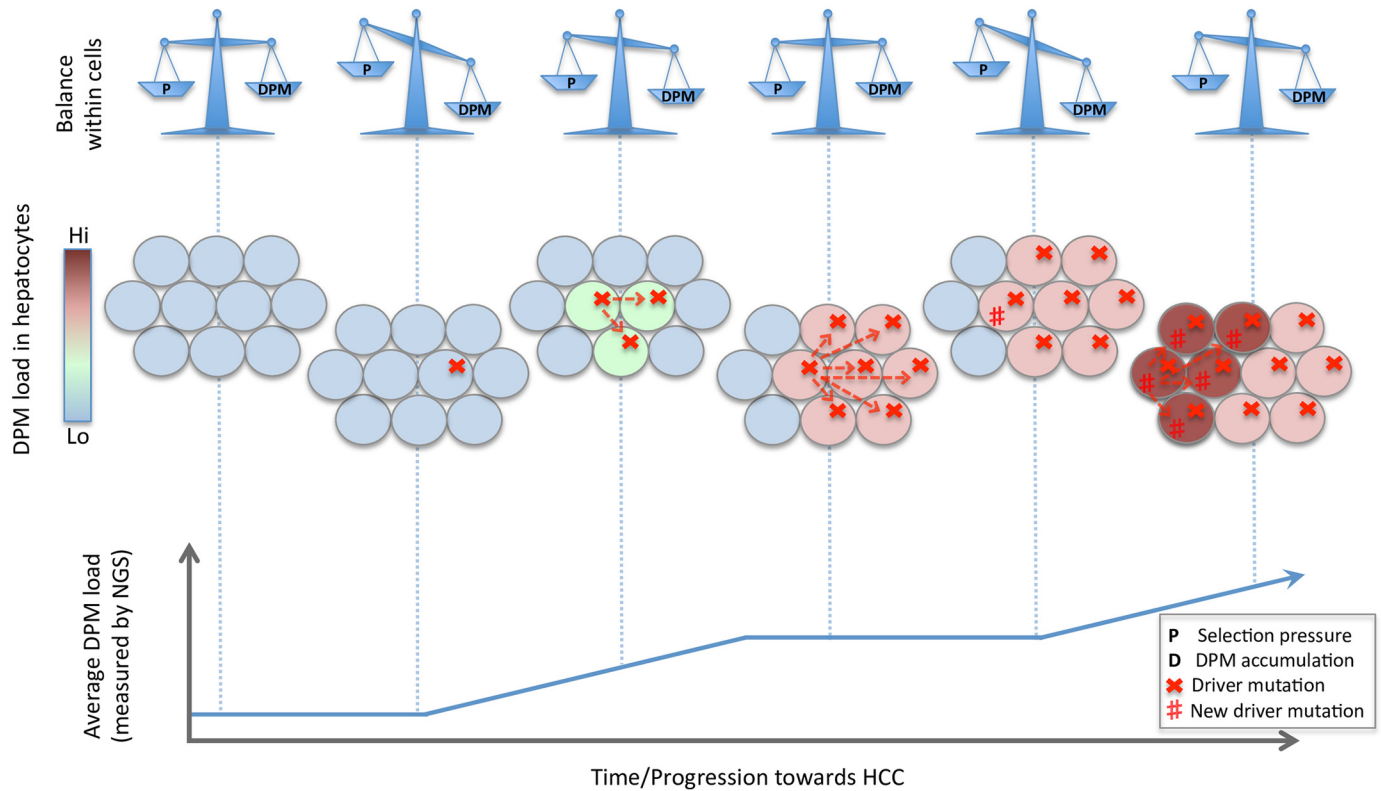


Fig 6. Hypothetical model of HCC progression. HCC progression is presented here as multiple waves of driver sweeps within hepatocyte subclones. The equilibrium between DPM accumulation and negative selection on the hepatocyte subclones are shown in the top row. A schematic model of the liver (with each circle representing a hepatocyte and the colour gradient representing the DPM load within each hepatocyte) is shown in the centre row. The average DPM load for the tissue is depicted in the bottom row.

doi:10.1371/journal.pone.0162586.g006

Based on our findings we propose the following model of HCC development (Fig 6):

1. Hepatocyte subclones acquire driver mutations through random mutation, giving them a survival advantage.
2. This disrupts the selection equilibrium in favour of DPM acquisition.
3. Equilibrium is restored when selection against the accumulated DPMs evens out the survival advantage.
4. The hepatocyte subclone population plateaus until the next driver mutation.
5. Steps 1–4 are repeated, eventually culminating into HCC through acquiring sufficient driver mutations.

This model suggests that DPMs could form the basis of a genetic biomarker, though our results suggest that interpatient variability is considerable and it may be of limited use as a measure of HCC risk. However, the data raises the intriguing possibility that cirrhosis progression with increasing DPM accumulation may be a risk factor or signature for HCC development. Further, it is unclear if certain subsets of DPMs may predict aspects of tumour biology and/or behaviour. These possibilities are difficult to investigate, as they require serial sampling in humans over months to many years. Animal models with their lack of cirrhosis associated with

HCC as well the use of agents that globally damage DNA such as diethyl-nitrosamine (DEN) are poor surrogates to answer these questions. In future studies (especially as more sequencing data becomes publically available), larger patient cohorts, serial samples and a better understanding of deleterious effects of DNA mutations on liver cell phenotype will allow better tests for this hypothetical model.

In summary, we have shown that progressive liver injury and HCC are accompanied by accumulation of DPMs. We also have provided evidence that surrounding non-tumour tissue is not genetically “normal”. While the true effect of accumulated DPMs on tumour biology is still unknown, given their frequency and functional implications, they cannot be ignored

Supporting Information

S1 Fig. Summary statistics of WES 1 reads. (A) Mean depth of reads for each sample, (B) Fraction of target covered in caption region (4-fold, 10-fold and 20-fold coverage) per exome. (TIFF)

S2 Fig. Comparison of benign missense variants and DPMs in 1000G and WES 5 datasets. To determine if circulating leukocytes could be used as a control to account for germline mutations in individuals, we compared DNA from peripheral blood mononuclear cells (PBMCs) in 1000G and patients with hepatitis B virus (HBV) infection (WES 5) [32]. We found significantly more benign mutations (A) and DPMs (B) in HBV-exposed patients compared to healthy people from the 1000G dataset (* $p < 0.05$, **** $p < 0.0001$, Mann-Whitney test), suggesting the DNA genome of PBMCs are altered as a result of HBV infection. This may be due many factors dependent on HBV-associated inflammation, including: DNA mutations introduced during high levels of PBMC mitosis; or DPMs being accumulating as a result of clonal expansion of PBMCs. Greater immune activation associated with HBV infection would be expected to increase clonal expansion, and therefore DPMs according to our model. Crucially, this result suggests that liver disease causes changes in the DNA within the blood (not just the liver) and so using PBMC-derived DNA sequences to exclude germline variants would introduce bias. This therefore justifies our approach of using only the 1000 Genomes Project database to exclude probable germline mutations. (TIFF)

S3 Fig. Estimated allelic frequency distribution of benign missense variants and DPMs. The allelic frequency of each benign missense variant (left) and DPMs (right) was estimated by the number of reads containing the variant divided by the number of the total reads at that particular base (x-axis). This was expressed as a cumulative plot with each patient as different colours for all benign missense variants and DPMs for 1000G and WES 1–4 (top, middle and bottom respectively). For WES 2–4, paired tumour (solid line) and non-tumour (dashed line) for the same patient are coloured the same colour. (TIFF)

S4 Fig. Absolute number of variants. Non-synonymous mutations for all datasets were subdivided into 4 groups: missense, non-frameshift ins/del, frameshift ins/del and stop-gain/-loss. Samples in 1000G and WES 1 are unpaired, while samples in WES 2–4 paired. After excluding probable germline mutations, absolute numbers of variants (A and B) are shown for each sample. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and **** $p < 0.0001$, Mann-Whitney U test (1000G and WES 1) or Wilcoxon matched-pairs signed-rank test (WES 2–4). (TIFF)

S5 Fig. DPMs in HCC and surrounding non-tumour tissue in genes expressed in non-diseased liver. We analysed benign missense variants (A and B) and DPMs (C and D) in genes expressed in non-diseased liver tissue (measured by microarray analysis). The significant increase in DPMs in tumour tissue compared to paired non-tumour tissue was maintained (** $p < 0.01$, *** $p < 0.001$ and **** $p < 0.0001$, Wilcoxon matched-pairs signed-rank test). No significant differences in benign missense variants or DPMs were detected between non-cirrhotic and cirrhotic patients ($p > 0.05$, Mann-Whitney U test).
(TIFF)

S6 Fig. Comparison of benign missense variants and DPMs using SIFT algorithm. We compared total benign missense variants (A and B) and DPMs (C and D) in the datasets 1000G and WES 1–4. 1000G and WES 1 are unpaired and WES 2–4 paired. Significantly more DPMs (but not benign missense SNVs) were detected in tumour compared to paired non-tumour tissue (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ and **** $p < 0.0001$, Wilcoxon matched-pairs signed-rank test). We also analysed benign missense variants (E and F) and DPMs (G and H) in genes expressed in non-diseased liver tissue. The significant increase in tumour tissue was maintained. No significant differences in benign missense variants or DPMs were detected between non-cirrhotic and cirrhotic patients ($p > 0.05$, Mann-Whitney U test). All variants were normalised to the total exonic variants after exclusion of probable germline mutations. Lines show linkages between matched paired non-tumour and tumour tissues samples.
(TIFF)

S7 Fig. DPMs distribution throughout the exome. The genomic distribution of benign missense mutations (radial lines) are shown in Circos plot for (A) 1000G (grey) and WES 1 for non-cirrhotic (green) and cirrhotic (blue) patients, and for WES 2 (B), WES 3 (C), and WES 4 (D) for non-tumour (black) and tumour (red) tissues. The distribution of DPMs is also shown for 1000G and WES 1 (E), WES 2 (F), WES 3 (G) and WES 4 (H). The outer grey circle represents the exons location (USCS). Even distribution throughout the genome was observed given the exon distribution and the coverage of the reference sequence hg19.
(TIFF)

S1 Supplementary Methods.

(PDF)

S1 Table. 1000 Genome samples summary.

(DOCX)

S2 Table. List of genes expressed above background levels detected in liver tissue of donor patients.

(DOCX)

S3 Table. Frequency of driver mutations in analysed samples.

(DOCX)

S4 Table. Significantly enriched ($p \leq 10^{-3}$) canonical pathways in DPM-affected genes.

(DOCX)

Acknowledgments

The authors would like to thank Prof Jessica Zucman-Rossi for providing raw data of WES 2 dataset and Dr. Frederick Sierrro for constructive discussion of the manuscript.

Author Contributions

Conceptualization: TT NAS.

Data curation: MAB.

Formal analysis: MAB.

Funding acquisition: NAS GWM.

Investigation: TT MAB WMHdA.

Methodology: MAB.

Project administration: NAS.

Resources: NAS GWM.

Software: MAB FL.

Supervision: TT NAS.

Validation: MAB TT FL.

Visualization: MAB TT.

Writing – original draft: TT MAB NAS.

Writing – review & editing: TT MAB FL NAS.

References

1. Bosch FX, Ribes J, Diaz M, Cleries R. Primary liver cancer: worldwide incidence and trends. *Gastroenterology*. 2004; 127(5 Suppl 1):S5–S16. Epub 2004/10/28. S0016508504015902 [pii]. PMID: [15508102](#).
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin*. 2011; 61(2):69–90. Epub 2011/02/08. doi: [10.3322/caac.20107](#) caac.20107 [pii]. PMID: [21296855](#).
3. Parkin DM, Pisani P, Ferlay J. Estimates of the worldwide incidence of 25 major cancers in 1990. *Int J Cancer*. 1999; 80(6):827–41. Epub 1999/03/13. doi: [10.1002/\(SICI\)1097-0215\(19990315\)80:6<827::AID-IJC6>3.0.CO;2-P](#) [pii]. PMID: [10074914](#).
4. Kumar R, Saraswat MK, Sharma BC, Sakhuja P, Sarin SK. Characteristics of hepatocellular carcinoma in India: a retrospective analysis of 191 cases. *QJM*. 2008; 101(6):479–85. Epub 2008/04/29. doi: [10.1093/qjmed/hcn033](#) hcn033 [pii]. PMID: [18440958](#).
5. Schoniger-Hekele M, Muller C, Kutilek M, Oesterreicher C, Ferenci P, Gangl A. Hepatocellular carcinoma in Austria: aetiological and clinical characteristics at presentation. *Eur J Gastroenterol Hepatol*. 2000; 12(8):941–8. Epub 2000/08/25. PMID: [10958223](#).
6. Wong PY, Xia V, Imagawa DK, Hoefs J, Hu KQ. Clinical presentation of hepatocellular carcinoma (HCC) in Asian-Americans versus non-Asian-Americans. *J Immigr Minor Health*. 2011; 13(5):842–8. Epub 2010/10/05. doi: [10.1007/s10903-010-9395-8](#) PMID: [20890660](#); PubMed Central PMCID: PMC3155747.
7. Lim KC, Chow PK, Allen JC, Siddiqui FJ, Chan ES, Tan SB. Systematic review of outcomes of liver resection for early hepatocellular carcinoma within the Milan criteria. *Br J Surg*. 2012; 99(12):1622–9. Epub 2012/10/02. doi: [10.1002/bjs.8915](#) PMID: [23023956](#).
8. Morris-Stiff G, Gomez D, de Liguori Carino N, Prasad KR. Surgical management of hepatocellular carcinoma: is the jury still out? *Surg Oncol*. 2009; 18(4):298–321. Epub 2008/12/09. doi: [10.1016/j.suronc.2008.08.003](#) S0960-7404(08)00082-0 [pii]. PMID: [19062271](#).
9. Omer RE, Kuijsten A, Kadaru AM, Kok FJ, Idris MO, El Khidir IM, et al. Population-attributable risk of dietary aflatoxins and hepatitis B virus infection with respect to hepatocellular carcinoma. *Nutr Cancer*. 2004; 48(1):15–21. Epub 2004/06/19. doi: [10.1207/s15327914nc4801_3](#) PMID: [15203373](#).
10. Yeh FS, Yu MC, Mo CC, Luo S, Tong MJ, Henderson BE. Hepatitis B virus, aflatoxins, and hepatocellular carcinoma in southern Guangxi, China. *Cancer Res*. 1989; 49(9):2506–9. Epub 1989/05/01. PMID: [2539905](#).

11. Cleary SP, Jeck WR, Zhao X, Chen K, Selitsky SR, Savich GL, et al. Identification of driver genes in hepatocellular carcinoma by exome sequencing. *Hepatology*. 2013; 58(5):1693–702. Epub 2013/06/04. doi: [10.1002/hep.26540](https://doi.org/10.1002/hep.26540) PMID: [23728943](https://pubmed.ncbi.nlm.nih.gov/23728943/); PubMed Central PMCID: PMC3830584.
12. Guichard C, Amaddeo G, Imbeaud S, Ladeiro Y, Pelletier L, Maad IB, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat Genet*. 2012; 44(6):694–8. Epub 2012/05/09. doi: [10.1038/ng.2256](https://doi.org/10.1038/ng.2256) ng.2256 [pii]. PMID: [22561517](https://pubmed.ncbi.nlm.nih.gov/22561517/); PubMed Central PMCID: PMC3819251.
13. Huang J, Deng Q, Wang Q, Li KY, Dai JH, Li N, et al. Exome sequencing of hepatitis B virus-associated hepatocellular carcinoma. *Nat Genet*. 2012; 44(10):1117–21. Epub 2012/08/28. doi: [10.1038/ng.2391](https://doi.org/10.1038/ng.2391) ng.2391 [pii]. PMID: [22922871](https://pubmed.ncbi.nlm.nih.gov/22922871/).
14. Jiang S, Yang Z, Li W, Li X, Wang Y, Zhang J, et al. Re-evaluation of the Carcinogenic Significance of Hepatitis B Virus Integration in Hepatocarcinogenesis. *PLoS ONE*. 2012; 7(9):e40363. Epub 2012/09/11. doi: [10.1371/journal.pone.0040363](https://doi.org/10.1371/journal.pone.0040363) PMID: [22962577](https://pubmed.ncbi.nlm.nih.gov/22962577/); PubMed Central PMCID: PMC3433482.
15. Jiang Z, Jhunjunwala S, Liu J, Haverly PM, Kennemer MI, Guan Y, et al. The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. *Genome research*. 2012; 22(4):593–601. Epub 2012/01/24. doi: [10.1101/gr.133926.111](https://doi.org/10.1101/gr.133926.111) PMID: [22267523](https://pubmed.ncbi.nlm.nih.gov/22267523/); PubMed Central PMCID: PMC3317142.
16. Schulze K, Imbeaud S, Letouze E, Alexandrov LB, Calderaro J, Rebouissou S, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet*. 2015; 47(5):505–11. Epub 2015/03/31. doi: [10.1038/ng.3252](https://doi.org/10.1038/ng.3252) ng.3252 [pii]. PMID: [25822088](https://pubmed.ncbi.nlm.nih.gov/25822088/).
17. Jhunjunwala S, Jiang Z, Stawiski EW, Gnad F, Liu J, Mayba O, et al. Diverse modes of genomic alteration in hepatocellular carcinoma. *Genome Biol*. 2014; 15(8):436. Epub 2014/08/28. doi: [10.1186/s13059-014-0436-9](https://doi.org/10.1186/s13059-014-0436-9) s13059-014-0436-9 [pii]. PMID: [25159915](https://pubmed.ncbi.nlm.nih.gov/25159915/); PubMed Central PMCID: PMC4189592.
18. Azechi H, Nishida N, Fukuda Y, Nishimura T, Minata M, Katsuma H, et al. Disruption of the p16/cyclin D1/retinoblastoma protein pathway in the majority of human hepatocellular carcinomas. *Oncology*. 2001; 60(4):346–54. Epub 2001/06/16. 58531 [pii] 58531. PMID: [11408803](https://pubmed.ncbi.nlm.nih.gov/11408803/).
19. Edamoto Y, Hara A, Biernat W, Terracciano L, Cathomas G, Riehle HM, et al. Alterations of RB1, p53 and Wnt pathways in hepatocellular carcinomas associated with hepatitis C, hepatitis B and alcoholic liver cirrhosis. *Int J Cancer*. 2003; 106(3):334–41. Epub 2003/07/08. doi: [10.1002/ijc.11254](https://doi.org/10.1002/ijc.11254) PMID: [12845670](https://pubmed.ncbi.nlm.nih.gov/12845670/).
20. Ito T, Nishida N, Fukuda Y, Nishimura T, Komeda T, Nakao K. Alteration of the p14(ARF) gene and p53 status in human hepatocellular carcinomas. *J Gastroenterol*. 2004; 39(4):355–61. Epub 2004/05/29. doi: [10.1007/s00535-003-1302-9](https://doi.org/10.1007/s00535-003-1302-9) PMID: [15168247](https://pubmed.ncbi.nlm.nih.gov/15168247/).
21. Su H, Zhao J, Xiong Y, Xu T, Zhou F, Yuan Y, et al. Large-scale analysis of the genetic and epigenetic alterations in hepatocellular carcinoma from Southeast China. *Mutat Res*. 2008; 641(1–2):27–35. PMID: [18358501](https://pubmed.ncbi.nlm.nih.gov/18358501/). doi: [10.1016/j.mrfmmm.2008.02.005](https://doi.org/10.1016/j.mrfmmm.2008.02.005)
22. Luo D, Liu QF, Gove C, Naomov N, Su JJ, Williams R. Analysis of N-ras gene mutation and p53 gene expression in human hepatocellular carcinomas. *World J Gastroenterol*. 1998; 4(2):97–9. Epub 2002/01/31. PMID: [11819246](https://pubmed.ncbi.nlm.nih.gov/11819246/).
23. Iwamoto KS, Mizuno T, Tokuoka S, Mabuchi K, Seyama T. Frequency of p53 mutations in hepatocellular carcinomas from atomic bomb survivors. *J Natl Cancer Inst*. 1998; 90(15):1167–8. Epub 1998/08/13. PMID: [9701367](https://pubmed.ncbi.nlm.nih.gov/9701367/).
24. Tu T, Budzinska MA, Shackel NA, Jilbert AR. Conceptual models for the initiation of hepatitis B virus-associated hepatocellular carcinoma. *Liver Int*. 2015; 35(7):1786–800. Epub 2015/02/03. doi: [10.1111/liv.12773](https://doi.org/10.1111/liv.12773) PMID: [25640596](https://pubmed.ncbi.nlm.nih.gov/25640596/).
25. Woo HG, Park ES, Thorgeirsson SS, Kim YJ. Exploring genomic profiles of hepatocellular carcinoma. *Mol Carcinog*. 2011; 50(4):235–43. Epub 2011/04/06. doi: [10.1002/mc.20691](https://doi.org/10.1002/mc.20691) PMID: [21465573](https://pubmed.ncbi.nlm.nih.gov/21465573/); PubMed Central PMCID: PMC3496176.
26. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009; 458(7239):719–24. Epub 2009/04/11. doi: [10.1038/nature07943](https://doi.org/10.1038/nature07943) nature07943 [pii]. PMID: [19360079](https://pubmed.ncbi.nlm.nih.gov/19360079/); PubMed Central PMCID: PMC2821689.
27. Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet*. 2007; 80(4):727–39. Epub 2007/03/16. S0002-9297(07)61104-5 [pii] doi: [10.1086/513473](https://doi.org/10.1086/513473) PMID: [17357078](https://pubmed.ncbi.nlm.nih.gov/17357078/); PubMed Central PMCID: PMC1852724.
28. McFarland CD, Korolev KS, Kryukov GV, Sunyaev SR, Mirny LA. Impact of deleterious passenger mutations on cancer progression. *Proc Natl Acad Sci U S A*. 2013; 110(8):2910–5. Epub 2013/02/08.

- doi: [10.1073/pnas.1213968110](https://doi.org/10.1073/pnas.1213968110) 1213968110 [pii]. PMID: [23388632](https://pubmed.ncbi.nlm.nih.gov/23388632/); PubMed Central PMCID: PMC3581883.
29. Mason WS, Liu C, Aldrich CE, Litwin S, Yeh MM. Clonal expansion of normal-appearing human hepatocytes during chronic hepatitis B virus infection. *J Virol.* 2010; 84(16):8308–15. Epub 2010/06/04. doi: [10.1128/JVI.00833-10](https://doi.org/10.1128/JVI.00833-10) JVI.00833-10 [pii]. PMID: [20519397](https://pubmed.ncbi.nlm.nih.gov/20519397/); PubMed Central PMCID: PMC2916518.
 30. Merlo LM, Shah NA, Li X, Blount PL, Vaughan TL, Reid BJ, et al. A comprehensive survey of clonal diversity measures in Barrett's esophagus as biomarkers of progression to esophageal adenocarcinoma. *Cancer Prev Res (Phila).* 2010; 3(11):1388–97. Epub 2010/10/16. doi: [10.1158/1940-6207.CAPR-10-0108](https://doi.org/10.1158/1940-6207.CAPR-10-0108) 1940-6207.CAPR-10-0108 [pii]. PMID: [20947487](https://pubmed.ncbi.nlm.nih.gov/20947487/); PubMed Central PMCID: PMC3004782.
 31. Tu T, Mason WS, Clouston AD, Shackel NA, McCaughan GW, Yeh MM, et al. Clonal expansion of hepatocytes with a selective advantage occurs during all stages of chronic hepatitis B virus infection. *J Viral Hepat.* 2015; 22(9):737–53. Epub 2015/01/27. doi: [10.1111/jvh.12380](https://doi.org/10.1111/jvh.12380) PMID: [25619231](https://pubmed.ncbi.nlm.nih.gov/25619231/).
 32. Peng L, Zhao Q, Li Q, Li M, Li C, Xu T, et al. The p.Ser267Phe variant in SLC10A1 is associated with resistance to chronic hepatitis B. *Hepatology.* 2015; 61(4):1251–60. doi: [10.1002/hep.27608](https://doi.org/10.1002/hep.27608) PMID: [25418280](https://pubmed.ncbi.nlm.nih.gov/25418280/).
 33. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010; 38(16):e164. Epub 2010/07/06. doi: [10.1093/nar/gkq603](https://doi.org/10.1093/nar/gkq603) gkq603 [pii]. PMID: [20601685](https://pubmed.ncbi.nlm.nih.gov/20601685/); PubMed Central PMCID: PMC2938201.
 34. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012; 491(7422):56–65. Epub 2012/11/07. doi: [10.1038/nature11632](https://doi.org/10.1038/nature11632) nature11632 [pii]. PMID: [23128226](https://pubmed.ncbi.nlm.nih.gov/23128226/); PubMed Central PMCID: PMC3498066.
 35. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010; 7(4):248–9. Epub 2010/04/01. doi: [10.1038/nmeth0410-248](https://doi.org/10.1038/nmeth0410-248) nmeth0410-248 [pii]. PMID: [20354512](https://pubmed.ncbi.nlm.nih.gov/20354512/); PubMed Central PMCID: PMC2855889.
 36. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 2003; 31(13):3812–4. Epub 2003/06/26. PMID: [12824425](https://pubmed.ncbi.nlm.nih.gov/12824425/); PubMed Central PMCID: PMC168916.
 37. Vijg J, Dolle ME. Large genome rearrangements as a primary cause of aging. *Mech Ageing Dev.* 2002; 123(8):907–15. PMID: [12044939](https://pubmed.ncbi.nlm.nih.gov/12044939/).
 38. Dolle ME, Snyder WK, Gossen JA, Lohman PH, Vijg J. Distinct spectra of somatic mutations accumulated with age in mouse heart and small intestine. *Proc Natl Acad Sci U S A.* 2000; 97(15):8403–8. PMID: [10900004](https://pubmed.ncbi.nlm.nih.gov/10900004/); PubMed Central PMCID: PMC26960.
 39. Bozic I, Reiter JG, Allen B, Antal T, Chatterjee K, Shah P, et al. Evolutionary dynamics of cancer in response to targeted combination therapy. *Elife.* 2013; 2:e00747. Epub 2013/06/28. doi: [10.7554/eLife.00747](https://doi.org/10.7554/eLife.00747) 00747 [pii]. PMID: [23805382](https://pubmed.ncbi.nlm.nih.gov/23805382/); PubMed Central PMCID: PMC3691570.
 40. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. *Nature.* 2007; 446(7132):153–8. Epub 2007/03/09. nature05610 [pii] doi: [10.1038/nature05610](https://doi.org/10.1038/nature05610) PMID: [17344846](https://pubmed.ncbi.nlm.nih.gov/17344846/); PubMed Central PMCID: PMC2712719.
 41. Tomasetti C, Vogelstein B, Parmigiani G. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proceedings of the National Academy of Sciences of the United States of America.* 2013; 110(6):1999–2004. Epub 2013/01/25. doi: [10.1073/pnas.1221068110](https://doi.org/10.1073/pnas.1221068110) 1221068110 [pii]. PMID: [23345422](https://pubmed.ncbi.nlm.nih.gov/23345422/); PubMed Central PMCID: PMC3568331.
 42. Nault JC, Calderaro J, Tommaso LD, Balabaud C, Zafrani ES, Bioulac-Sage P, et al. TERT promoter mutation is an early somatic genetic alteration in the transformation of premalignant nodules in hepatocellular carcinoma on cirrhosis. *Hepatology.* 2014. Epub 2014/08/16.