OPEN ACCESS

ORIGINAL ARTICLE

# Integration of tumour and viral genomic characterisations in HBV-related hepatocellular carcinomas

Giuliana Amaddeo,[1,2] Qian Cao,[1,2] Yannick Ladeiro,[1,2] Sandrine Imbeaud,[1,2] Jean-Charles Nault,[1,2] Daphne Jaoui,[3] Yann Gaston Mathe,[3] Christophe Laurent,[4] Alexis Laurent,[5,6] Paulette Bioulac-Sage,[7,8] Julien Calderaro,[1,2,9] Jessica Zucman-Rossi[1,2,10]

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2013-306228).

For numbered affiliations see end of article.

**Correspondence to**
Professor Jessica Zucman-Rossi, Inserm U674, Génomique fonctionnelle des tumeurs solides, 27 rue Juliette Dodu, Paris 75010, France; jessica.zucman-rossi@inserm.fr

GA and QC contributed equally.

Received 10 October 2013
Revised 13 May 2014
Accepted 14 May 2014
Published Online First 9 June 2014

## ABSTRACT

**Background and aim** Hepatocellular carcinoma (HCC) is the most common liver cancer. We characterised HCC associated with infection compared with non-HBV-related HCC to understand interactions between viral and hepatocyte genomic alterations and their relationships with clinical features.

**Methods** Frozen HBV (n=86) or non-HBV-related (n=90) HCC were collected in two French surgical departments. Viral characterisation was performed by sequencing *HBS* and *HBX* genes and quantifying HBV DNA and cccDNA. Nine genes were screened for somatic mutations and expression profiling of 37 genes involved in hepatocarcinogenesis was studied.

**Results** *HBX* revealed frequent non-sense, frameshift and deletions in tumours, suggesting an *HBX* inactivation selected in HCC. The number of viral copies was frequently lower in tumour than in non-tumour tissues (p=0.0005) and patients with low HBV copies in the non-tumour liver tissues presented additional risk factor (HCV, alcohol or non-alcoholic steato-hepatitis, p=0.006). P53 was the most frequently altered pathway in HBV-related HCC (47%, p=0.001). Furthermore, *TP53* mutations were associated with shorter survival only in HBV-related HCC (p=0.02) whereas R249S mutations were identified exclusively in migrants. Compared with other aetiologies, HBV-HCC were more frequently classified in tumours subgroups with upregulation of genes involved in cell-cycle regulation and a progenitor phenotype. Finally, in HBV-related HCC, transcriptomic profiles were associated with specific gene mutations (*HBX*, *TP53*, *IRF2*, *AXIN1* and *CTNNB1*).

**Conclusions** Integrated genomic characterisation of HBV and non-HBV-related HCC emphasised the immense molecular diversity of HCC closely related to aetiologies that could impact clinical care of HCC patients.

## Significance of this study

### What is already known on this subject?

- Hepatocellular carcinoma (HCC) is the fifth most common cancer and hepatitis B is the most common underlying cause of HCC worldwide.
- HCC is a group of heterogeneous tumours with several molecular subclasses.
- HBV infection synergises with exposure to aflatoxin B1 to promote HCC with R249S *TP53* mutation in subtropical countries.

### What are the new findings?By analysing resected HCC in France, we identified specific molecular features related to HBV infection:

- *HBX* inactivating mutations are selected in HCC tissues suggesting specific pressure of selection during hepatocarcinogenesis.
- In patients with a low number of HBV DNA copies per cell in the liver, we identified additional risk factor (HCV infection, alcohol intake or non-alcoholic steato-hepatitis) suggesting a cooperative effect for HBV-induced carcinogenesis.
- *TP53* mutations demonstrate different prognostic consequences according to the viral status; relation with poor prognosis was restricted to HBV-infected resected patients.
- HBV-related tumours demonstrate more frequent progenitor phenotype than non-HBV HCC with an upregulation of genes involved in cell-cycle regulation and encoding oncofetal/progenitor proteins.

### How might it impact on clinical practice in the foreseeable future?

- HBV and non-HBV-related HCC are molecularly different as a result of a different selection of genomic and transcriptomic alterations. Consequently, identification of prognostic markers and therapeutic targets in HCC should be validated in HBV and non-HBV cohorts of patients.
- A molecular classification of HCC is required to design a more personalised care of the patients taking into account the molecular diversity of the tumours.

Open Access
Scan to access more free content

CrossMark

**To cite:** Amaddeo G, Cao Q, Ladeiro Y, *et al*. *Gut* 2015;**64**:820–829.

## INTRODUCTION

Hepatitis B is a potentially life-threatening liver disease affecting approximately 5% of world population (350–400 million people).[1] Persistent HBV infection causes chronic liver disease and is thought to be responsible for approximately 50% of all hepatocellular carcinoma (HCC) cases and virtually all HCC in childhood. Overall, HCC occur

predominantly in male patients and is the fifth most common cancer worldwide.[2 3]

HBV is a small enveloped DNA virus and contains four overlapping reading frames: S (encoding the viral surface proteins, *HBS*), P (encoding viral polymerase), X (encoding the regulatory X protein, *HBX*, which is capable of transactivating the expression of numerous cellular and viral genes) and, finally, pre C (encoding the antigens 'e' and 'c').

HBV is thought to play a key role in HCC development by increasing chromosomal instability and promoting cell proliferation.[4] In infected liver tissues, several variants of HBV proteins and genes have been described, including truncated *HBS*,[5 6] mutated or deleted *HBX*,[7 8] and a novel spliced transcript of HBV, referred to as the hepatitis B spliced protein.[9 10] HBV integration into the host cell genome has also been reported to be responsible for gene mutations, insertions, deletions or rearrangements of the host genome.[11 12] Recently, using next generation sequencing technics, a large number of insertion sites were identified.[13–15] Two insertion sites of HBV were most frequent: in proximity to *TERT*, activating telomerase, and in the *MLL4* gene. These alterations are suspected to play a major role in HCC carcinogenesis.[13–15]

However, the underlying nature of viral–host interaction remains elusive. A synergistic effect of HBV infection and exposition to aflatoxin B1 (AFB1) has been reported specifically in subtropical countries in Asia and Africa. AFB1 is a mycotoxin produced by the fungus *Aspergillus flavus* growing on improperly stored grains; it is metabolised in the liver by cytochrome P450 and induces the formation of highly promutagenic DNA adducts. Patients with both HBV and AFB1 risk factors display a 5–10-fold higher risk of HCC development as compared with HBV or AFB1 alone.[16–18] A common site for adduction of AFB1 metabolites is the third base of codon 249 in the *TP53* tumour suppressor gene, leading to G to T transversions (AGG to AGT, R to S). This base substitution is identified in 50% of HCC cases from areas with AFB1 exposure, but not in HCC without AFB1 exposure.[19 20] Therefore, the R249S mutation is considered to be a signature of AFB1 exposure.

HCC is also a heterogeneous group of tumours with variable clinical outcome and molecular features. We formerly reported a robust classification of HCC that comprised six main subgroups (G1–G6) according to their transcriptomic profile.[21] These subgroups were closely associated with clinical and molecular features: G1–G2 subgroup demonstrated overexpression of fetal stage-associated genes and were controlled by parental imprinting; G3 tumours were characterised by *TP53* mutations and demonstrated adverse clinical outcome;[21 22] G4 was a heterogeneous subgroup of tumours; and, finally, G5–G6 subgroups were strongly related to β-catenin mutations, leading to Wnt pathway activation.[21] We observed that HBV infection was associated with G1–G2 tumours.[21]

However, the relationship between molecular classification and HBV infection remains to be elucidated. We therefore analysed two groups of tumours related (n=86) or unrelated (n=90) to HBV infection, and investigated the molecular diversity between the two cohorts by screening mutations in genes frequently mutated in HCC[14 23 24] and by profiling of gene expression. Finally, our data were correlated with clinical features and HBV characteristics.

## MATERIALS AND METHODS

### Patients and tissue samples

The study population consisted of 86 HBV-HCC cases (83 AgHBs-positive and three AgHBs-negative/AbHBc-positive and viral DNA-positive liver samples) and 90 cases of HBV-negative HCCs. Tissue samples were obtained after surgical resection from two French medical care centres in Bordeaux and Creteil. All tumour specimens and their matched adjacent non-tumour liver tissue samples were immediately cryopreserved at −80°C after surgical treatment.[25] Demographic features and clinical-pathological records of these patients were prospectively collected in dedicated databases. These data were blindly extracted and are summarised in table 1 and online supplementary table S1. Patients who presented a fatal outcome within 2 months after surgery were censored from the survival analysis. Follow-up was stopped at 60 months (mean 33±20). All patients had given their informed consent according to French law and the Saint Louis Hospital's ethic committee approved the study.

### HBS/HBX sequencing and genotyping

Viral DNA was amplified by PCR techniques using the Expand High Fidelity PCR System (Roche Diagnostics, Mannheim, Germany), according to manufacturer's instructions, and oligonucleotide primers specific for HBV DNA sequences flanking the S and X genomic regions. In several cases, HBV full-length genomes were amplified by a nested PCR approach using two different primer pairs (see online supplementary table S2) to obtain the amplification of *HBS* and *HBX* fragment.[21 26] Each purified PCR product was sequenced using Applied Biosystems BigDye terminator sequencing chemistry and run on an ABI Prism 3500Genetic Analyzer (Applied Biosystems, California, USA) according to manufacturer's instructions. The sequences were analysed with Sequencher V.4.7 software (Gene Codes Corporation, Ann Harbor, USA). HBV genotyping (A to H) was then performed according to the sequence of *HBS* by multiple sequence alignments using the Clustal Omega (http://www.clustal.org/).[27 28]

### Virus quantification

Absolute quantification of HBV genome was performed by qPCR in 86 HBV-related tumour samples and in 84 counterpart non-tumour tissues using a set of primers which target *HBS* gene (HBS4F: GTCCTGGTTATCGCTGGATG; HBS2R: CAAACGGGCAACATACCTTG)[29] via the SYBR Green (Applied Biosystems)-based expression analysis method. Results were reported to a gene (*ANKRD49*) that is located in a chromosomal region (11q21) usually not altered by copy number variation in HCC.[25] HBV cccDNA (copy number/ cell) was quantified by a specific real-time PCR, according to previous protocol.[30]

### Mutation screening

DNA was extracted from frozen tissue, in accordance with standard procedures.[21] For each tumour, nine genes were screened for somatic mutations: *TP53* (exons 2–11), *CTNNB1* (exons 2–4, 6–8), *PIK3CA* (exons 2, 9 and 20), *RPS6KA3* (exons 1–22), *ARID1A* (exons 1–20), *IRF2* (exons 2–9), *ARID2* (exons 1–20), *NFE2L2* (exons 1–5) and *AXIN1* (exons 1–10). DNA sequencing was performed by Sanger method. Primers and protocols are available on request. In all cases, the somatic origin of the mutation found in a tumour sample was verified by sequencing the corresponding adjacent, non-tumour liver sample.

### Quantitative RT-PCR

RNAs were extracted with RNeasy kit and RT-PCR was performed, as previously described, with the 2-delta delta CT method using the ribosomal 18S as calibrator gene and the mean level expression in normal liver tissues as calibrator

**Table 1** Clinical, histological and pathological data of 86 HBV- and 90 non-HBV-related HCC

| | HBV positive HCC | HBV negative HCC | p Value |
|---|---|---|---|
| Gender | | | |
| Male | 80% (69/86) | 81% (73/90) | NS |
| Female | 20% (17/86) | 19% (17/90) | |
| Age (year) | | | |
| Means±SD | 53.3±12.4 | 65±9.8 | <0.0001* |
| Geographic origins | | | |
| Africans | 45% (39/86) | 5% (4/90) | <0.0001** |
| Asians | 18% (15/86) | 0 (0/90) | |
| Europeans | 37% (36/86) | 95% (86/90) | |
| Cofactors/aetiologies | | | |
| HDV | 5% (4/86) | 0 (0/90) | NS |
| HCV | 9% (8/86) | 23% (21/90) | 0.01** |
| Alcohol | 10% (9/86) | 40% (36/90) | <0.0001** |
| Non-alcoholic steato-hepatitis | 2% (2/86) | 4% (4/90) | NS |
| Haemochromatosis | 0 (0/86) | 7% (6/90) | NS |
| Others | 0 (0/86) | 26% (23/90) | NS |
| α-Fetoprotein (ng/mL) | | | |
| Means±SD | 15 026±48 569 | 2455±9265 | 0.008* |
| Tumour size (mm) | | | |
| Means±SD | 72.7±49.6 | 77.1±52.6 | 0.03* |
| Edmondson grade | | | |
| Well differentiated (I–II) | 38% (33/86) | 53% (48/90) | 0.05** |
| Poorly differentiated (III–IV) | 58% (50/86) | 45% (40/90) | |
| Unknown | 4% (3/86) | 2% (2/90) | |
| Portal invasions | | | |
| Yes | 20% (17/86) | 16% (14/90) | |
| No | 79% (68/86) | 83% (75/90) | NS |
| Unknown | 1% (1/86) | 1% (1/90) | |
| Microvascular invasions | | | |
| Yes | 43% (37/86) | 56% (50/90) | |
| No | 55% (47/86) | 43% (39/90) | NS |
| Unknown | 2% (2/86) | 1% (1/90) | |
| Satellite nodules | | | |
| Yes | 48% (41/86) | 49% (44/90) | |
| No | 50% (43/86) | 51% (46/90) | NS |
| Unknown | 2% (2/86) | 0 (0/90) | |
| Early relapse (<24 month) | | | |
| No | 36% (31/86) | 37% (33/90) | |
| Yes | 43% (37/86) | 60% (54/90) | NS |
| Unknown | 21% (18/86) | 3% (3/90) | |

p Values obtained from Mann–Whitney (*) and $\chi^2$ (**) tests based on the given clinical variable are shown.
HCC, hepatocellular carcinoma; HDV, hepatitis D virus.

sample.[21] Expression levels of a set of 37 genes, known in the literature to be involved in tumourigenesis pathways,[21 31–35] were analysed using TaqMan predesigned gene expression assay (Applied Biosystems). This includes a combination of 16 genes used in the G1–G6 HCC transcriptome classification.[21] TaqMan gene expression assays are listed in online supplementary table S3.

## Statistical analysis

We used a hypercube algorithm technique to explore model parameter space. HyperCube is an alternative and original approach for the discovery of subgroups derived from least general generalisation algorithms (Torre, Globo). The key characteristic of HyperCube is that it uses non-parametric modelling without any metric and without making any assumption on the distribution of variables or the relationship between the outcome and the input variables.[36 37] Here we apply this methodology and consider 89 explanatory variables for the series of HBV-related HCC and 80 variables for the series of non-HBV-related HCC (the variables between these two series are similar except for those nine variables about the virus) to identify key factors underlying the risk of HCC of different aetiologies. Quantitative data were presented as mean±SD, and non-parametric Mann–Whitney tests were used to compare quantitative values. All differences were considered significant when the p value was less than 0.05. The interdependence between numerical variables was performed by the use of the Spearman rank correlation test. $\chi^2$ And Fisher exact tests were used for given genetic or clinical variable. We used log-rank test and Kaplan–Meier method to assess survival.

## RESULTS

### Studied cohort characteristics

In this study, we included 86 patients with HBV-HCC and 90 patients with non-HBV-HCC between 1999 and 2009 after tumour resection. In both cohorts, male patients were predominant (80%), and aggressiveness of the tumours were similar as shown by non-significant differences in the frequency of vascular invasion, satellite nodules and early relapse. In contrast, HBV-related patients were younger (p<0.0001), frequently migrants from Africa or Asia (63%, p<0.0001), presented higher α-fetoprotein (AFP) serum levels (p=0.008), had less differentiated and smaller tumours (p=0.05 and p=0.03) than non-HBV patients. Other risk factors of HCC were associated with HBV infection in almost 26% of the cases, whereas aetiologies in non-HBV-related HCC were mainly alcohol intake (40%), HCV infection (23%), haemochromatosis (7%) and non-alcoholic steato-hepatitis (NASH, 4%), with a non-negligible percentage of patients who had a cryptogenic aetiology (26%) (table 1 and see online supplementary table S1).

### Viral characterisation in tumour and non-tumour tissues

Sequencing the *HBS* gene, we identified HBV genotypes according to the Viral Genotype Tool of the NCBI databases in 82 HCC tumours and in 80 of their corresponding non-tumour tissues (four HCCs could not be determined for genotypes because of *HBS* gene deletion). Genotypes A, B, C, D and E were found in 40%, 12%, 10%, 23% and 15% of non-tumour tissues and in 43%, 10%, 16%, 23% and 8% of tumour samples, respectively (see online supplementary figure S1). In 13 cases (15%), two different genotypes were identified in the same sample (nine only in HCC, three in non-tumour tissue and one in both HCC and non-tumour counterpart). Genotypes were similar in the tumour when compared with the corresponding non-tumour liver tissues, except in eight patients (9% of the cases) with non-tumour to tumour genotypes E:A (n=3), A:C (n=3), B:C (n=1) and D:C (n=1). Genotypes identified in the non-tumour tissues were selected to search for correlations with clinical features. As described in the literature, genotypes A and D were common in patients from Africa and Europe, genotypes C and B in patients from Asia and genotype E in patients from Africa (p<0.0001) (see online supplementary table S4).[38 39] Genotype D was also more frequent in old patients (p=0.01) and genotype E was associated with hepatitis D virus co-infection in our limited series (p=0.01, n=3/4) (see online supplementary table S4).

Then, we sequenced *HBS* and *HBX* genes in 86 HCC and 84 non-tumour counterparts' samples and compared the spectrum of nucleotide variants in tumour and non-tumour samples (figure 1A, B). Missense mutations were defined as sequence variants modifying amino acid translation and found in less than 1% of the 7822 *HBS* sequences available in public databases (Hepatitis Virus Database). Inactivating mutations were defined as nucleotide variants leading to stop or frameshift; they also included truncated deletions. Among the tumour and non-tumour tissues, frequency of *HBS* mutations (88% and 88%, respectively) were similar and the ratio of missense/stop and frameshift mutants were 1.8 and 2.5 (48/27 and 53/21, respectively; figure 1A, B). In contrast, in *HBX* gene, inactivating alterations were significantly more frequent in tumours (71%) than in non-tumour tissues (33%, p<0.0001, figure 1A, B). The I127N/T/L, K130M/K/Q, V131I/L/T amino acid substitutions in *HBX* gene, which have been previously identified in HCC tumours,[40 41] were frequently detected in the present study both in non-tumour and tumour tissues at a similar frequency. These observations suggest that whereas mutations inactivating *HBX* are selected in tumours, substitution at residues I127, K130 and V131 are not.

Using quantitative PCR, we quantified the total number of viral copies by determining an absolute quantification of *HBS* DNA corresponding to the different types of HBV DNA in the tissues, that is, episomal, integrated or the covalently closed circular DNA (cccDNA). According to the median number of HBS copies/cell in the non-tumour liver tissues, we defined two groups of patients with high total *HBS* copies (>0.5 copy/cell) or with low total *HBS* copies (<0.5 copy/cell) in the non-tumour tissues (figure 1C). Overall, in 53 out of 84 (63%) of the paired samples, we found a lower amount of total *HBS* DNA copy in the tumour tissues compared with their corresponding non-tumour tissues (p=0.0005; figure 1D). In contrast, cccDNA number of copies is similarly distributed in both compartments (see online supplementary figure S2A,B). Consequently, the lower amount of HBV DNA observed in tumours is mainly related to a decrease of non-replicative viral genome. Thus, in tumours, the mutations observed in HBV sequences should be mainly related to non-replicative form of the virus.

Interestingly, we showed a correlation between viral copy numbers in non-tumour tissue, viral genotype in non-tumour tissue and the presence of other risk factors. Patients with low *HBS* DNA copy number presented more frequently an additional risk factor (HCV, alcohol or NASH) than patients with high viral quantification (38% vs 12%, p=0.006). Moreover, patients with a low viral quantification were more frequently infected by HBV genotypes D and E (p=0.01).

### Spectrum of somatic gene mutations correlates with HBV infection and survival

In the overall series of 176 HBV and non-HBV-related HCC, we searched for somatic mutations by screening the coding sequence of nine genes previously described as frequently altered in HCC[23] (table 2, figure 2 and see online supplementary tables S5 and S6). Comparing the spectrum of somatic mutations in HBV and non-HBV HCC, we identified significant differences. *NFE2L2*, coding NRF2 a key transcription factor involved in oxidative stress response, was only mutated in non-HBV HCC (7/90, 8%) and not in HBV-infected tumours (0/86, p=0.008; table 2). *CTNNB1* activating mutations were also less frequent in HBV-HCC (13/86, 15%) compared with non-HBV HCC (40/90, 44%, p<0.0001; table 2). These results suggested that the WNT/ß-catenin and the oxidative stress pathways are significantly less frequently activated by gene mutation in HBV-related carcinogenesis than in HCC associated with other risk factors.

In contrast, P53 pathway was more frequently altered in HBV-HCC and the *TP53* gene was the most frequently mutated gene in HBV-related HCC (41%), whereas it was rarely mutated in non-HBV tumours (16%, p=0.0002). R249S mutation in *TP53*, specific of the AFB1 exposure, was exclusively found in HBV-HCC (16%, p<0.0001; table 2) particularly with genotype B (p=0.005; table 3). According to the lack of AFB1 exposure in France, all the patients with R249S mutations were migrants coming from Asia or sub-Saharan Africa (p=0.02; table 3). Interestingly, we have observed a correlation between *R249S* mutations and poor tumour differentiation (p=0.05; table 3); tumours with R249S mutations were also more frequently developed in non-cirrhotic liver (p=0.02) in contrast to other HBV-infected HCC.[42] Overall, most of the *TP53* mutations were described to inactivate P53 according to the IARC

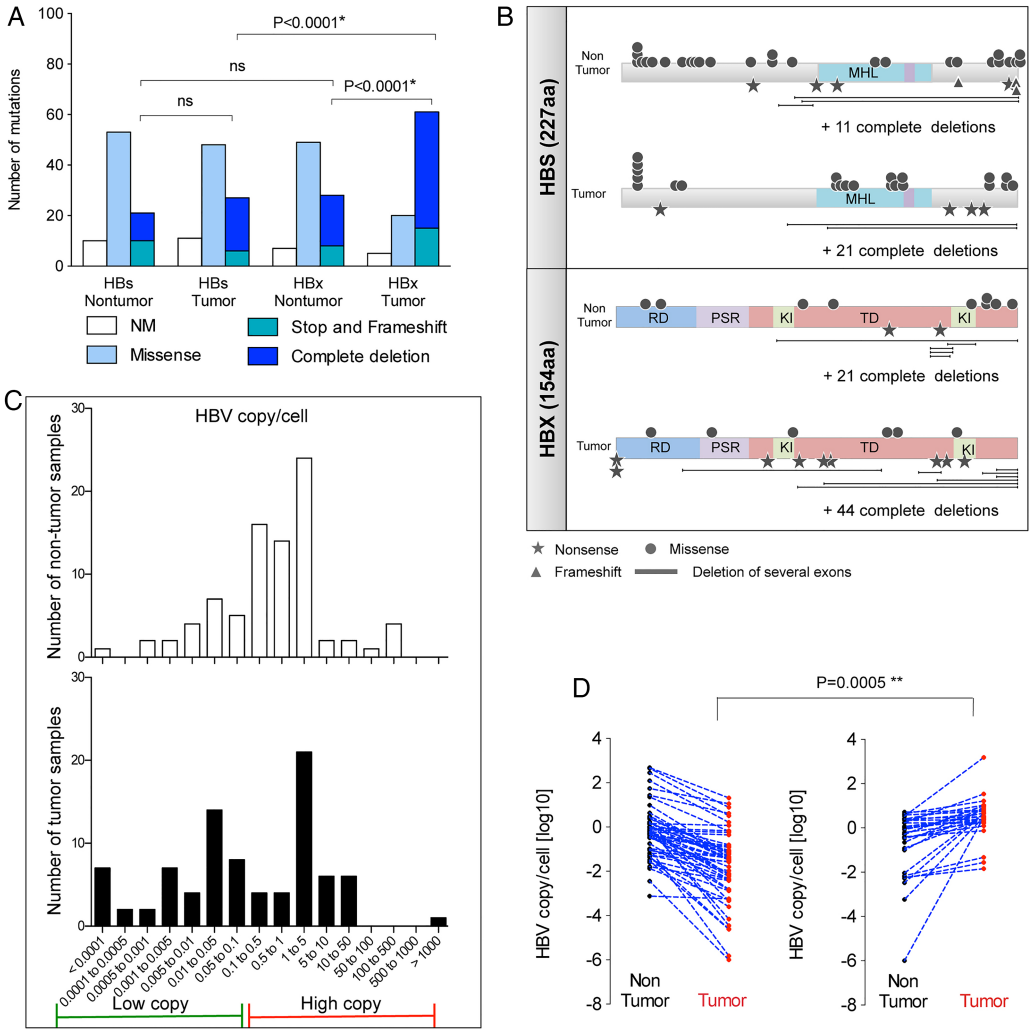



**Figure 1** HBV virus profile in tumour and non-tumour adjacent samples. (A and B) Spectra of mutation in *HBS* and *HBX* gene in tumour and paired non-tumour samples. (C) Distribution of tumour and non-tumour samples according to quantification of HBV copy/cell and determination of two groups characterised by low (<0.5 HBV copy/cell) and high (>0.5 HBV copy/cell) copy number. (D) Correlation of HBV quantification (log10 copy/cell) in tumour and paired non-tumour samples. p Values obtained from $\chi^2$ (*) and Willcoxon signed rank (**) tests are shown. MHL, major hydrophilic loop; RD, regulatory domain; PSR, proline serine rich hypervariable region; KI, Kunitz domain-like; TD, transactivation domain.

database (see online supplementary table S5). Alterations inactivating *IRF2*, a tumour suppressor gene controlling p53 protein activation, were also exclusively identified in HBV-HCC (7%, p=0.01; table 2). According to their function in the same pathway, *IRF2* and *TP53* mutations were mutually exclusive in tumours (p=0.002).

**Table 2** Gene mutations of 86 HBV- and 90 non-HBV-related HCC

| Gene | HBV+HCC | HBV−HCC | p Value |
|---|---|---|---|
| TP53 | 41% (35/86) | 16% (14/90) | 0.0002 |
| TP53R249S | 16% (14/86) | 0% (0/90) | <0.0001 |
| IRF2 | 7% (6/86) | 0% (0/90) | 0.01 |
| CTNNB1 | 15% (13/86) | 44% (40/90) | <0.0001 |
| AXIN1 | 15% (13/86) | 13% (12/90) | NS |
| ARID1A | 9% (8/86) | 16% (14/90) | NS |
| ARID2 | 5% (4/86) | 7% (6/90) | NS |
| PIK3CA | 1% (1/86) | 3% (3/90) | NS |
| RPS6KA3 | 6% (5/86) | 8% (7/90) | NS |
| NFE2L2 | 0% (0/86) | 8% (7/90) | 0.008 |

p Values obtained from $\chi^2$ test are shown.
HCC, hepatocellular carcinoma.

Strikingly, *TP53* mutations were significantly associated with shorter disease-specific survival (tumour-related death) in HBV-HCC (p=0.03; figure 3A) but not in non-HBV-HCC (figure 3B). In HBV-HCC, Edmondson grade III–IV (p=0.01), the presence of vascular invasion (portal invasion p<0.0001 and microvascular invasion p=0.009) and satellites nodules (p=0.02) were also significantly associated with a shorter disease-specific survival using univariate analysis (table 4). In multivariate analysis, *TP53* mutations ($p$=0.004), the presence of portal invasion (p=0.001) and of microvascular invasion (p=0.09) were independently associated with an increased risk of death (table 4).

Interestingly, among HBV-HCC patients, European origin was associated with good prognosis in univariate and multivariate analyses (p=0.02 and p=0.02, respectively) and R249S hot spot mutations were not associated with survival (see online supplementary figure S3). Finally, *TP53* mutations were also significantly associated with poor prognosis in the subgroups of HBV patients with small (<55 mm; p=0.05) or well-differentiated HCC (Edmondson grade I–II; p=0.04; see online supplementary figure S4).

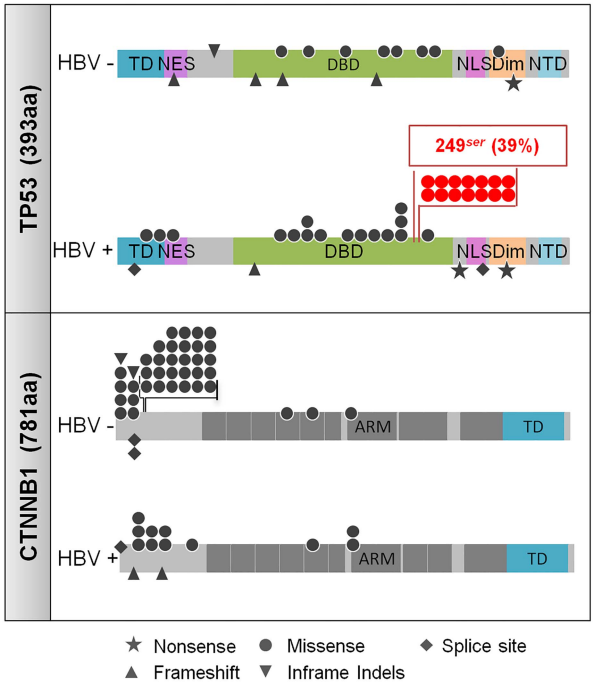



**Figure 2** Somatic mutation spectra of TP53 and CTNNB1 in 86 HBV-related and 90 non-HBV-related HCC. Functional domains are coloured boxes. ARM, armadillo repeat; DBD, DNA binding domain; Dim, dimerisation domain; NES, nuclear export signal; NLS, nuclear localisation signal; NTD, negative transactivation domain; TD, transactivation domain; ser, serine; HCC, hepatocellular carcinoma.

**Table 3** Clinical, histological and viral data of 86 HBV-related HCC, classified by the *TP53* R249S mutation or not

| | R249S | Non-R249S | p Value |
|---|---|---|---|
| Geographic origins | | | |
| Africans | 57% (8/14) | 42% (30/72) | 0.02* |
| Asians | 36% (5/14) | 15% (11/72) | |
| Europeans | 7% (1/14) | 43% (31/72) | |
| HBV genotype | | | |
| A | 36% (5/14) | 42% (30/72) | NS |
| B | 36% (5/14) | 5% (4/72) | 0.005* |
| C | 0% (0/14) | 17% (12/72) | NS |
| D | 14% (2/14) | 24% (17/72) | NS |
| E | 14% (2/14) | 7% (5/72) | NS |
| Undetermined | 0% (0/14) | 5% (4/72) | NS |
| Edmondson grade | | | |
| Well differentiated (I–II) | 14% (2/13) | 43% (31/70) | 0.05** |
| Poorly differentiated (III–IV) | 79% (11/13) | 54% (39/70) | |
| Liver histology | | | |
| Cirrhosis | 14% (2/14) | 50% (36/72) | 0.02* |
| Non-cirrhosis | 86% (12/14) | 50% (36/72) | |
| HBX sequence | | | |
| Amino acid substitutions | 7% (1/14) | 33% (24/72) | 0.06* |
| Inactivating mutations | 93% (13/14) | 67% (48/72) | |

p Values obtained from Fisher's exact (*) and $\chi^2$ (**) tests based on the given clinical or viral variables are shown.
HCC, hepatocellular carcinoma.

Altogether these data underlined that *TP53* mutations have a prognostic value in early HCC independent of classical clinical features in HBV-related HCC but not in other aetiologies.

### Integration of transcriptomic profiling in HBV-related HCC classification

We selected 37 genes differentially expressed among the different HCC subgroups previously defined by transcriptomic profiling and we classified the 176 HCC samples according to the six defined transcriptomic subgroups G1–G6.[21] HBV-HCC retains a genomic diversity and they were distributed in all G1–G6 subgroups. However, the genomic distribution according to the G1–G6 classification was significantly different between HBV and non-HBV-related HCC (p=0.001) with an enrichment of G2 and G3 HCC in HBV patients compared with non-HBV HCC (figure 4A). We further were able to demonstrate that the genes associated with progenitor features (*EpCAM*, *AFP*, *KRT19* and *CCNB1*) were globally significantly overexpressed in HBV-HCC compared with HCC related to other aetiologies (figure 4B). This result showed that HBV-HCC more frequently demonstrates a progenitor profile than the other tumours.

Within HBV-HCC, using transcriptome classification, we defined subgroups of patients in terms of clinical, pathological, viral and genetic features (figure 5). Partition in the two major groups (G1–G3/G4–G6) showed that G1–G3 subgroups included 57% of HBV-HCC and were enriched in large tumours (>55 mm; p=0.006; figure 5), characterised by frequent *HBX* gene inactivating mutations (p=0.001; figure 5) and frequent somatic *AXIN1* mutations (p=0.03; figure 5). Similar to HCC based on various aetiologies, HBV-HCC classified in G1–G3 was also characterised by the proliferative/stem cell features with an overexpression of a large number of genes implicated in cell-cycle control (*AURKA* p=0.03, *BIRC5* p<0.0001, *NEU1* p=0.005, *CCNB1* p<0.0001; figure 5) and a down-expression of genes involved in differentiation like *UGT2B7* (p<0.0001; figure 5). These associations were confirmed at the histological level, demonstrating that tumours categorised in G1–G3 are less differentiated (p=0.01; figure 5).

Among the G1–G3 group of HBV-HCC, in G1–G2 classes, we identified a specific subgroup of tumours enriched with *IRF2* inactivating alterations (p=0.006; figure 5) and characterised by a high overexpression of genes encoding oncofetal/progenitor proteins like epithelial cell adhesion molecule (*EPCAM*) (p=0.001), *AFP* (p<0.0001) and *KRT19* (p=0.002; figure 5). Moreover, tumours classified in G3 showed frequent vascular invasion (p=0.008) and *SPP1* (gene target of DNA replication, coding osteopontin) overexpression (p=0.008) underlining their aggressiveness (figure 5).[21 43] Next, in G2–G3, we confirmed the association with frequent *TP53* mutations (p=0.0009), all including R249S substitution (p=0.003) and a high rate of early recurrence (p=0.01; figure 5).

Interestingly, HBV-HCC classified in G4–G6, the second major transcriptomic class, was frequently associated with an additional risk factor (p=0.04) like HCV, alcohol intake or NASH, suggesting that HBV genome interaction would be less important for carcinogenesis in G4–G6 HCC than for those classified in G1–G3 subgroups (figure 5). Accordingly, HCC group G4 was characterised by a lower number of HBS DNA copy/cell in their corresponding non-tumour liver tissues (p=0.06) and by a good histological HCC differentiation (p=0.006; figure 5). G5–G6 groups consisted of a homogeneous group of HCC developed in older patients (≥60 years; p=0.0007; figure 5) with *CTNNB1* activating mutation related to WNT/β-catenin pathway activation (p<0.0001),

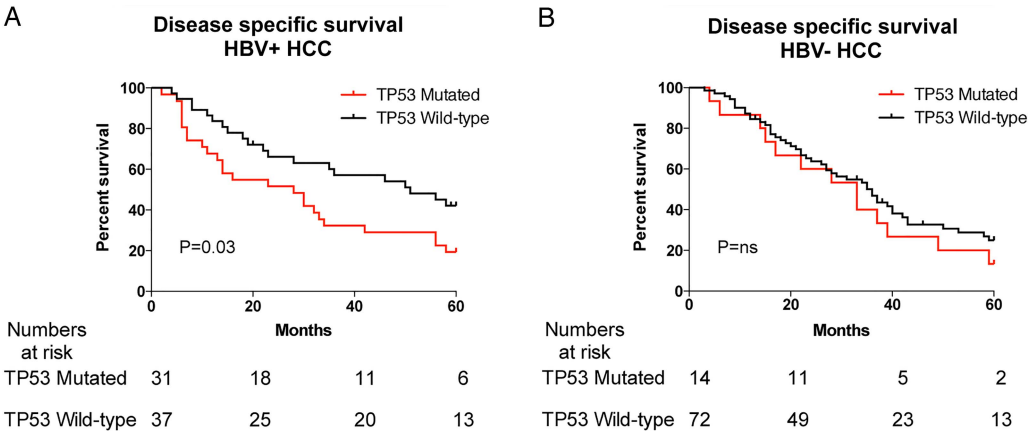



**Figure 3** Significant prognostic values of TP53 mutations. (A) Disease-specific survival in patients infected by HBV with HCC mutated or wild-type to TP53 and (B) in patients with HCC, mutated or wild-type to TP53, related to other aetiologies. HCC, hepatocellular carcinoma.

demonstrating an overexpression of *GLUL* (p=0.004), *TBX3* (p=0.02) and *RHBG* (p=0.02) and a down-expression of *HAL* (p=0.01), which are four well-known WNT/β-catenin target genes (figure 5). In this line, the low number of HCC classified in the G5–G6 subgroup in HBV patients is closely related to the low number of *CTNNB1* mutations observed in these tumours.

## DISCUSSION

In this study, we have undertaken a comprehensive and integrative approach to elucidate the molecular characteristics of HBV-HCC. Using transcriptome classification, gene sequencing, viral HBV characterisation and gene expression analysis, we showed that HBV-related HCC differs from non-HBV-related tumours. Moreover, among HBV-related HCC, we identified subgroups of tumours revealing specific viral, genetic and transcriptomic features.

We showed that genomic features were significantly different in the group of HBV-HCC compared with non-infected patients. These results suggest that HBV-related HCC use alternative mechanisms for tumourigenesis to some extent. A high frequency of p53 inactivation and stem cell genes overexpression provide a potential pathogenic link between impaired cell reprogramming and HBV infection. Notably, these observations have clinical implication since *TP53* mutations were associated with poor prognosis only in HBV-related tumours. These results could at least partly explain the discrepancies in the literature regarding the prognostic predictive value in HCC patients. Indeed, most frequently, *TP53* mutations were shown to be associated with prognosis in Asian populations infected by HBV,[44–46] whereas in a series of patients

**Table 4** Univariate and multivariate analysis for the selected clinical features of HBV- and non-HBV-related HCC

| | Univariate analysis | | | Multivariate analysis | | |
|---|---|---|---|---|---|---|
| ***HBV+HCC*** | **HR** | **95% CI** | **p Value*** | **HR** | **95% CI** | **p Value**** |
| **TP53 mutated** | **1.94** | **1.07; 3.5** | **0.03** | **2.27** | **1.10; 4.67** | **0.004** |
| AFP (>24 ng/mL) | 1.39 | 0.73; 2.63 | 0.3 | | | |
| **Edmondson III–IV** | **2.26** | **1.17; 4.34** | **0.01** | 1.32 | 0.60; 2.93 | 0.2 |
| Tumour size (>55 mm) | 1.49 | 0.81; 2.7 | 0.2 | | | |
| **Origin (European)** | **0.39** | **0.19; 0.77** | **0.02** | **0.51** | **0.24; 1.08** | **0.02** |
| **Portal invasion** | **5.17** | **2.55; 10.5** | **<0.0001** | **2.91** | **1.26; 6.69** | **0.001** |
| Tumour number (single) | 0.66 | 0.32; 1.34 | 0.25 | | | |
| **Satellites nodules** | **2.02** | **1.11; 3.66** | **0.02** | 1.44 | 0.68; 3.05 | 0.3 |
| **Microvascular invasion** | **2.2** | **1.22; 4.1** | **0.009** | **2.91** | **0.73; 3.44** | **0.09** |
| Alcohol intake | 0.9 | 0.37; 2.11 | 0.8 | | | |
| *HBV-HCC* | | | | | | |
| TP53 mutated | 0.80 | 0.43; 1.52 | 0.5 | | | |
| AFP (>24 ng/mL) | 1.16 | 0.68; 1.97 | 0.6 | | | |
| **Edmondson III–IV** | **1.72** | **1.03; 2.87** | **0.04** | 1.25 | 0.69; 2.24 | 0.06 |
| Tumour size (>55 mm) | 1.23 | 0.73; 2.04 | 0.4 | | | |
| Origin (European) | 1.33 | 0.47; 3.68 | 0.7 | | | |
| **Portal invasion** | **2.78** | **1.48; 5.23** | **0.001** | **2.37** | **1.17; 4.8** | **0.009** |
| Tumour number (single) | 0.88 | 0.44; 1.73 | 0.7 | | | |
| **Satellites nodules** | **1.8** | **1.08; 2.97** | **0.02** | 1.40 | 0.77; 2.55 | 0.3 |
| **Microvascular invasion** | **1.88** | **1.12; 3.17** | **0.02** | 1.11 | 0.57; 2.16 | 0.4 |
| **Alcohol intake** | **1.99** | **1.19; 3.31** | **0.008** | **2.06** | **1.22; 3.5** | **0.005** |

p Values obtained from univariate Cox proportional hazards model (*) and ANOVA (**) are shown.
AFP, α-fetoprotein; HCC, hepatocellular carcinoma.

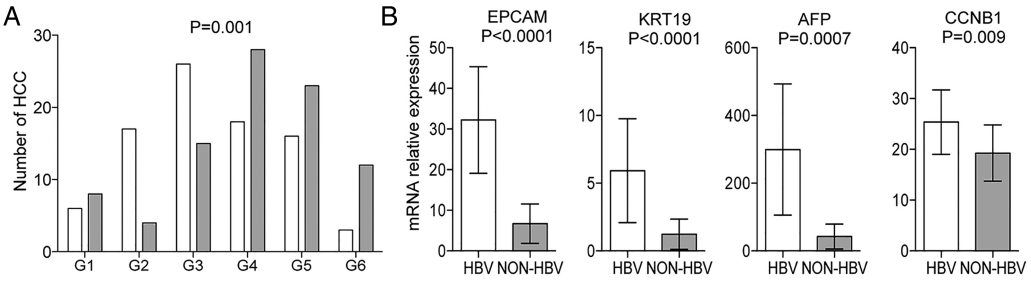



**Figure 4** Integration of transcriptomic profiling. (A) Transcriptomic groups (G1–G6) in HBV (*white*) and non-HBV-related (*grey*) HCC (mean with 95% CI). p Values obtained from $\chi^2$ test are shown. (B) Gene's expressions in HBV- and non-HBV-related HCC (mean with 95% CI). p Values obtained from Mann–Whitney test are shown. EPCAM, epithelial cell adhesion molecule; KRT19, keratin 19; AFP, α-fetoprotein; CCNB1, cyclin B1; HCC, hepatocellular carcinoma.

with mixed aetiologies, *TP53* mutations were usually not an independent prognostic marker.[43 47]

According to our previous study,[21] *CTNNB1* mutations defined a homogenous subtype of HCC classified in G5–G6 transcriptomic groups and overall G4–G6 groups of HCC were less frequently HBV infected. However, the tumours classified in G4–G6 subgroups also represent a small subset of tumours among the HBV-related HCC, characterised by frequently older patients, with other cofactors such as HCV, alcohol consumption or NASH. In the same line, we observed an association between low HBS DNA copy and presence of other risk factors. Altogether, these results suggest that among the group of HBV-infected patients, the association with an additional cofactor could interfere with the mechanism of carcinogenesis and the molecular features of this tumour type. We could hypothesise a different time frame of genetic alterations with a minor role of direct viral carcinogenesis, absence of AFB1 mutagenesis, and a predominant role of cofactor and stochastic accumulation of mutations during ageing. This carcinogenic pathway could promote the occurrence of *CTNNB1*-mutated HCC and HCC

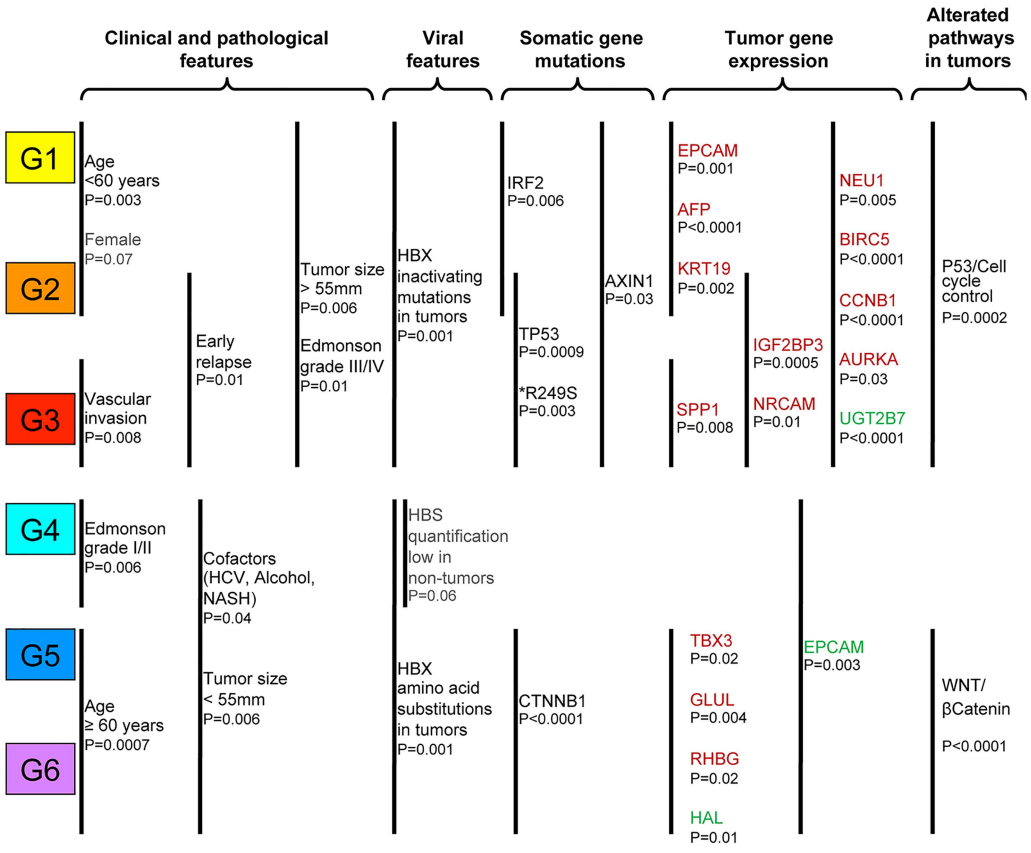



**Figure 5** Schematisation of the different HBV-HCC subgroups (G1–G6) defined by transcriptome analysis with their related clinical and genetic alteration. *Red* and *green* indicate overexpression and under-expression of gene, respectively, in the corresponding transcriptomic subgroup(s). p Values obtained from $\chi^2$ tests are shown. IRF2, interferon regulatory factor 2; EPCAM, epithelial cell adhesion molecule; KRT19, keratin 19; AFP, α-fetoprotein; SPP1, osteopontin; TBX3, T-box protein 3; GLUL, glutamate-ammonia ligase; RHBG, Rh family, B glycoprotein; HAL, histidine ammonia-lyase; IGF2BP3, insulin-like growth factor 2 MRNA binding protein 3; NRCAM, neuronal cell adhesion molecule; NEU1, sialidase 1; BIRC5, baculoviral IAP repeat containing 5; CCNB1, cyclin B1; AURKA, aurora kinase A; UGT2B7: UDP glucuronosyltransferase 2 family, polypeptide B7; HCC, hepatocellular carcinoma.

classified in the G4 subgroup that is otherwise unusual in HBV-related HCC.

Among the tumours related to HBV infection, genomic alterations can be associated with specific subgroups of tumours with particular clinical and epidemiological features. The R249S variant, the hot spot of mutation induced by AFB1 in *TP53*, is the result of the synergistic effects of AFB1 and HBV on tumour development.[16] In this regard, migrants coming from AFB1 exposed regions were treated and their tumours analysed similar to the non-migrant and non-infected patients. In our study, we were able to demonstrate that R249S mutation is associated with *HBX* inactivation and that this association is detected essentially, if not exclusively, in patients who developed HCC without clinical evidence of pre-existing or simultaneous liver cirrhosis. This observation supports the hypothesis that *R249S* and *HBX* could cooperate in promoting a pathway of hepatocarcinogenesis that does not require liver cirrhosis as an underlying, preneoplastic state.

Despite a large number of published studies, the function of HBX in liver carcinogenesis is still debated.[8 48 49] Here, by viral profiling of more than 80 cases, we identified frequent inactivating mutations in *HBX* gene in more than 70% of the tumours. We identified deletion of the COOH terminal part similar to that described by Iavarone and collaborators in their first report of six analysed tumours.[40] Interestingly, in the present work, by analysing a larger number of cases we also identified premature stop codon and large deletions leading to a complete inactivation of the *HBX* gene. It seems that wild type HBX or overexpression of truncated HBX proteins could demonstrate oncogenic function and promote tumourigenesis by abrogating cell-cycle arrest and apoptosis inhibition.[40 50–52] Our results suggest that mutations inactivating *HBX* are selected in the tumours in contrast to the non-tumour liver tissues. Consequently, we can hypothesise that HBX inactivation could participate in liver carcinogenesis. Surprisingly, we also identified a correlation between inactivating mutations of *HBX* gene and G1–G3 transcriptomic groups suggesting a close association between viral genome alterations and transcriptome expression during carcinogenesis that remains to be further investigated. Interestingly, transcriptomic classification revealed that these tumours belong to the group of patients with specific features: younger female patients frequently coming from Africa or Asia with tumours characterised by abnormal expression of oncofetal genes (*EPCAM*, *AFP* and *KRT19*).

In conclusion, our data collectively suggest that HBV infection leads to a specific type of HCC characterised by immature features with stem cell features when compared with HCC related to other risk factors. These results suggest that HBV-related HCC is associated with both HBX and p53 inactivation in a subset of tumours, proposing that these two factors could cooperate. Moreover, *TP53* mutation is associated with outcome for HBV-infected patients and it could have a notable clinical implication in risk stratification and treatment decision in this population of patients. Moreover, we identified in HBV-related HCC a high variability that probably reflects molecular heterogeneity and this aspect may also explain the high level of resistance again in a large number of therapeutic agents.

**Author affiliations**
[1]Inserm, UMR-674, Génomique fonctionnelle des tumeurs solides, IUH, Paris, France
[2]Université Paris Descartes, Labex Immuno-oncology, Sorbonne Paris Cité, Faculté de Médecine, Paris, France
[3]Hypercube research, France
[4]Department of Surgery, CHU de Bordeaux, Bordeaux, France
[5]Assistance Publique-Hôpitaux de Paris, digestive, hepatobiliary and liver transplantation, CHU Henri Mondor, Créteil, France
[6]IMRB—Inserm U955 Equipe n. 18 "Virologie moleculaire et immunologie – Physiopathologie et therapeutique des Hépatites virales chroniques", Créteil, France
[7]Inserm, UMR-1053; Université Victor Segalen Bordeaux 2, Bordeaux, France
[8]Department of Pathology, CHU de Bordeaux, Pellegrin Hospital, Bordeaux, France
[9]Department of Pathology, CHU Henri Mondor, Créteil, France
[10]Hopital Europeen Georges Pompidou, Paris, France

**Acknowledgements** We thank Kornelius Schulze for critical reading of this manuscript. We warmly thank all physicians who have participated in this work and particularly Jean Saric, Brigitte Le Bail, Anne Rullier, Antonio Sa Cunha, Daniel Cherqui and Jeanne Tran Van Nhieu. We thank Cecile Guichard and Ichrafe Ben Maad for the help in sequencing genes. We thank the tumour bank of CHU Henri Mondor (Creteil, Paris) and CHU Bordeaux for providing tumour samples. This work was supported by the ANRS, INCa with the ICGC project, the Réseau national CRB Foie, HEPTROMIC (FP7), BioIntelligence and INSERM. GA is supported by a fellowship from the ANRS and QC is supported by a fellowship from HEPTROMIC.

**Contributors** GA, QC and YL performed molecular characterisations. J-CN contributed to survival analysis. SI, DJ and YGM performed statistical analyses. CL and AL provided samples and clinical information. JC and PB-S provided samples and pathological reviewing. JZ-R designed and coordinated the overall study. All authors contributed to writing the manuscript.

**Funding** ANRS; INCa; FP7 (HEPTROMIC); OSEO (BIOINTELLIGENCE); INSERM.

**Competing interests** None.

**Patient consent** Obtained.

**Data sharing statement** The authors commit to assisting any investigators seeking to replicate or follow up this study with advice and sharing data.

**Ethics approval** Saint Louis Hospital's ethic committee (Paris, France).

**Provenance and peer review** Not commissioned; externally peer reviewed.

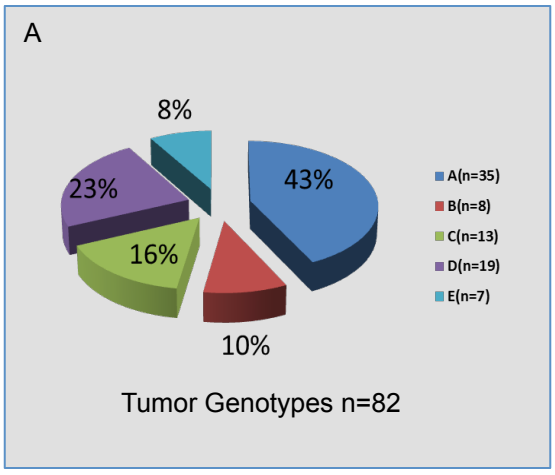


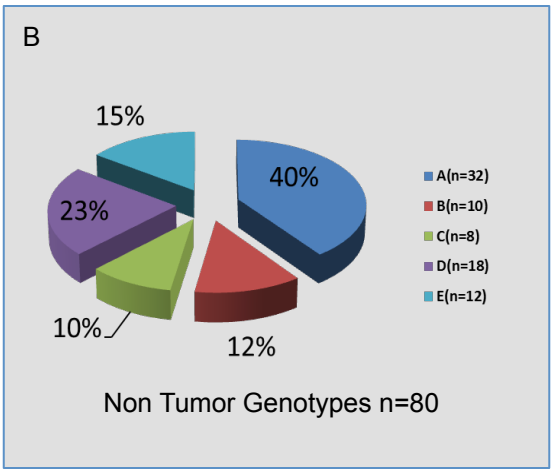
## REFERENCES

1 El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* 2007;132:2557–76.
2 El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* 2012;142:1264–73 e1.
3 El-Serag HB. Hepatocellular carcinoma. *N Engl J Med* 2011;365:1118–27.
4 Arzumanyan A, Reis HM, Feitelson MA. Pathogenic mechanisms in HBV- and HCV-associated hepatocellular carcinoma. *Nat Rev Cancer* 2013;13:123–35.
5 Liu S, Zhang H, Gu C, *et al*. Associations between hepatitis B virus mutations and the risk of hepatocellular carcinoma: a meta-analysis. *J Natl Cancer Inst* 2009;101:1066–82.
6 Hsieh YH, Su IJ, Wang HC, *et al*. Pre-S mutant surface antigens in chronic hepatitis B virus infection induce oxidative stress and DNA damage. *Carcinogenesis* 2004;25:2023–32.
7 Tu H, Bonura C, Giannini C, *et al*. Biological impact of natural COOH-terminal deletions of hepatitis B virus X protein in hepatocellular carcinoma tissues. *Cancer Res* 2001;61:7803–10.
8 Hsia CC, Yuwen H, Tabor E. Hot-spot mutations in hepatitis B virus X gene in hepatocellular carcinoma. *Lancet* 1996;348:625–6.
9 Soussan P, Garreau F, Zylberberg H, *et al*. In vivo expression of a new hepatitis B virus protein encoded by a spliced RNA. *J Clin Invest* 2000;105:55–60.
10 Soussan P, Tuveri R, Nalpas B, *et al*. The expression of hepatitis B spliced protein (HBSP) encoded by a spliced hepatitis B virus RNA is associated with viral replication and liver fibrosis. *J Hepatol* 2003;38:343–8.
11 Brechot C. Pathogenesis of hepatitis B virus-related hepatocellular carcinoma: old and new paradigms. *Gastroenterology* 2004;127:S56–61.
12 Bonilla Guerrero R, Roberts LR. The role of hepatitis B virus integrations in the pathogenesis of human hepatocellular carcinoma. *J Hepatol* 2005;42:760–77.
13 Sung WK, Zheng H, Li S, *et al*. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet* 2012;44:765–9.
14 Fujimoto A, Totoki Y, Abe T, *et al*. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet* 2012;44:760–4.
15 Jiang Z, Jhunjhunwala S, Liu J, *et al*. The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. *Genome Res* 2012;22:593–601.
16 Kew MC. Synergistic interaction between aflatoxin B1 and hepatitis B virus in hepatocarcinogenesis. *Liver Int* 2003;23:405–9.

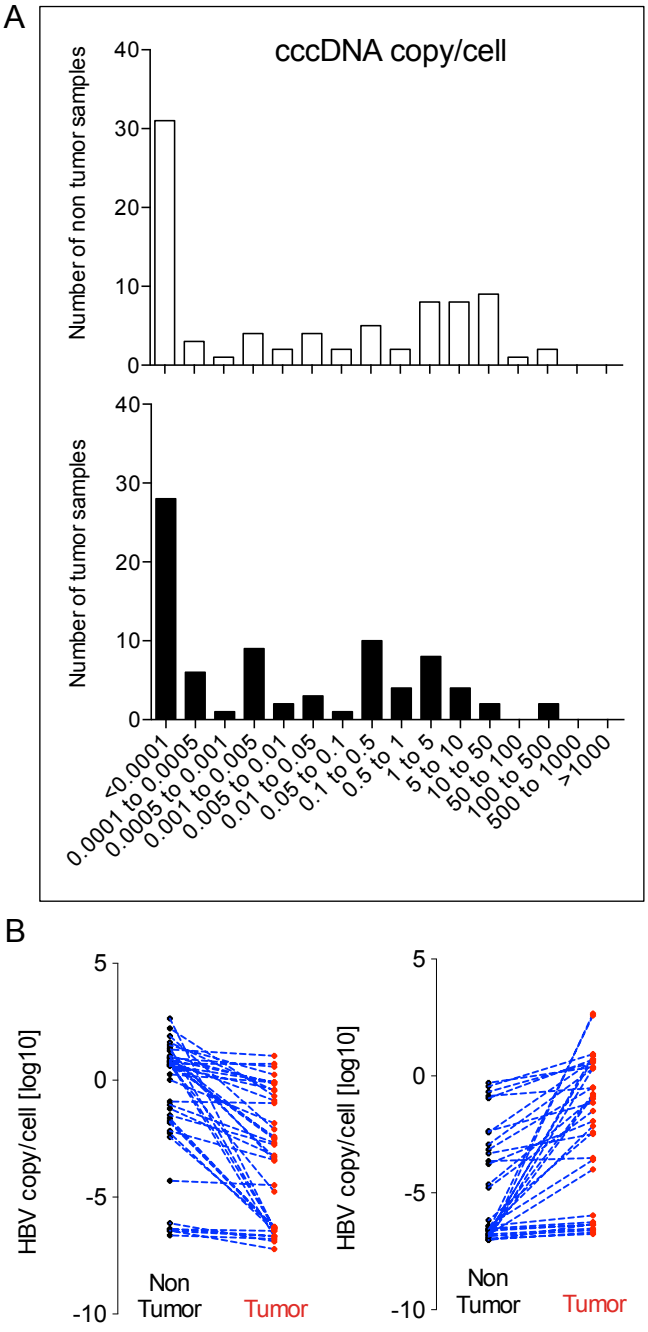
17 Lunn RM, Zhang YJ, Wang LY, *et al*. p53 mutations, chronic hepatitis B virus infection, and aflatoxin exposure in hepatocellular carcinoma in Taiwan. *Cancer Res* 1997;57:3471–7.
18 Ross RK, Yu MC, Henderson BE, *et al*. Aflatoxin biomarkers. *Lancet* 1992;340:119.
19 Hsu IC, Metcalf RA, Sun T, *et al*. Mutational hotspot in the p53 gene in human hepatocellular carcinomas. *Nature* 1991;350:427–8.
20 Bressac B, Kew M, Wands J, *et al*. Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. *Nature* 1991;350:429–31.
21 Boyault S, Rickman DS, de Reynies A, *et al*. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* 2007;45:42–52.
22 Villanueva A, Hoshida Y. Depicting the role of TP53 in hepatocellular carcinoma progression. *J Hepatol* 2011;55:724–5.
23 Guichard C, Amaddeo G, Imbeaud S, *et al*. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat Genet* 2012;44:694–8.
24 Sung CY, Lee NP, El-Nezami H. Regulation of T helper 17 by bacteria: an approach for the treatment of hepatocellular carcinoma. *Int J Hepatol* 2012;2012:439024.
25 Laurent-Puig P, Legoix P, Bluteau O, *et al*. Genetic alterations associated with hepatocellular carcinomas define distinct pathways of hepatocarcinogenesis. *Gastroenterology* 2001;120:1763–73.
26 Gunther S, Li BC, Miska S, *et al*. A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. *J Virol* 1995;69:5437–44.
27 Sievers F, Wilm A, Dineen D, *et al*. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;7:539.
28 McWilliam H, Li W, Uludag M, *et al*. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res* 2013;41:W597–600.
29 Lole KS, Arankalle VA. Quantitation of hepatitis B virus DNA by real-time PCR using internal amplification control and dual TaqMan MGB probes. *J Virol Methods* 2006;135:83–90.
30 Werle-Lapostolle B, Bowden S, Locarnini S, *et al*. Persistence of cccDNA during the natural history of chronic hepatitis B and decline during adefovir dipivoxil therapy. *Gastroenterology* 2004;126:1750–8.
31 Bioulac-Sage P, Rebouissou S, Thomas C, *et al*. Hepatocellular adenoma subtype classification using molecular markers and immunohistochemistry. *Hepatology* 2007;46:740–8.
32 Chan JY, Ong CW, Salto-Tellez M. Overexpression of neurone glial-related cell adhesion molecule is an independent predictor of poor prognosis in advanced colorectal cancer. *Cancer Sci* 2011;102:1855–61.
33 Kondoh N, Wakatsuki T, Ryo A, *et al*. Identification and characterization of genes associated with human hepatocellular carcinogenesis. *Cancer Res* 1999;59:4990–6.
34 Kurinna S, Stratton SA, Coban Z, *et al*. p53 regulates a mitotic transcription program and determines ploidy in normal mouse liver. *Hepatology* 2013;57:2004–13.
35 Di Tommaso L, Destro A, Seok JY, *et al*. The application of markers (HSP70 GPC3 and GS) in liver biopsies is useful for detection of hepatocellular carcinoma. *J Hepatol* 2009;50:746–54.
36 Loucoubar C, Paul R, Bar-Hen A, *et al*. An exhaustive, non-euclidean, non-parametric data mining tool for unraveling the complexity of biological systems—novel insights into malaria. *PLoS ONE* 2011;6:e24085.
37 Loucoubar C, Grange L, Paul R, *et al*. High number of previous Plasmodium falciparum clinical episodes increases risk of future episodes in a sub-group of individuals. *PLoS ONE* 2013;8:e55666.
38 Kidd-Ljunggren K, Miyakawa Y, Kidd AH. Genetic variability in hepatitis B viruses. *J Gen Virol* 2002;83:1267–80.
39 Pujol FH, Navas MC, Hainaut P, *et al*. Worldwide genetic diversity of HBV genotypes and risk of hepatocellular carcinoma. *Cancer Lett* 2009;286:80–8.
40 Iavarone M, Trabut JB, Delpuech O, *et al*. Characterisation of hepatitis B virus X protein mutants in tumour and non-tumour liver cells using laser capture microdissection. *J Hepatol* 2003;39:253–61.
41 Pollicino T, Raffa G, Costantino L, *et al*. Molecular and functional analysis of occult hepatitis B virus isolates from patients with hepatocellular carcinoma. *Hepatology* 2007;45:277–85.
42 Villar S, Ortiz-Cuaran S, Abedi-Ardekani B, *et al*. Aflatoxin-induced TP53 R249S mutation in hepatocellular carcinoma in Thailand: association with tumors developing in the absence of liver cirrhosis. *PLoS ONE* 2012;7:e37707.
43 Villanueva A, Hoshida Y, Battiston C, *et al*. Combining clinical, pathology, and gene expression data to predict recurrence of hepatocellular carcinoma. *Gastroenterology* 2011;140:1501–12 e2.
44 Woo HG, Wang XW, Budhu A, *et al*. Association of TP53 mutations with stem cell-like gene expression and survival of patients with hepatocellular carcinoma. *Gastroenterology* 2011;140:1063–70.
45 Park NH, Chung YH, Youn KH, *et al*. Close correlation of p53 mutation to microvascular invasion in hepatocellular carcinoma. *J Clin Gastroenterol* 2001;33:397–401.
46 Yuan RH, Jeng YM, Chen HL, *et al*. Stathmin overexpression cooperates with p53 mutation and osteopontin overexpression, and is associated with tumour progression, early recurrence, and poor prognosis in hepatocellular carcinoma. *J Pathol* 2006;209:549–58.
47 Nault JC, De Reynies A, Villanueva A, *et al*. A hepatocellular carcinoma 5-gene score associated with survival of patients after liver resection. *Gastroenterology* 2013;145:176–87.
48 Kew MC. Hepatitis B virus x protein in the pathogenesis of hepatitis B virus-induced hepatocellular carcinoma. *J Gastroenterol Hepatol* 2011;26(Suppl 1):144–52.
49 Bouchard MJ, Schneider RJ. The enigmatic X gene of hepatitis B virus. *J Virol* 2004;78:12725–34.
50 Kim CM, Koike K, Saito I, *et al*. HBx gene of hepatitis B virus induces liver cancer in transgenic mice. *Nature* 1991;351:317–20.
51 Clippinger AJ, Gearhart TL, Bouchard MJ. Hepatitis B virus X protein modulates apoptosis in primary rat hepatocytes by regulating both NF-kappaB and the mitochondrial permeability transition pore. *J Virol* 2009;83:4718–31.
52 Ma NF, Lau SH, Hu L, *et al*. COOH-terminal truncated HBV X protein plays key role in hepatocarcinogenesis. *Clin Cancer Res* 2008;14:5061–8.

**Supplementary Figure 1**

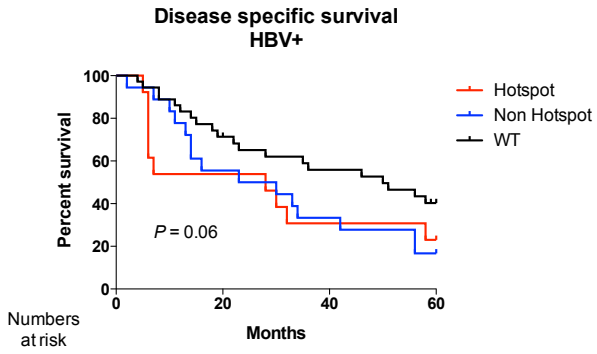


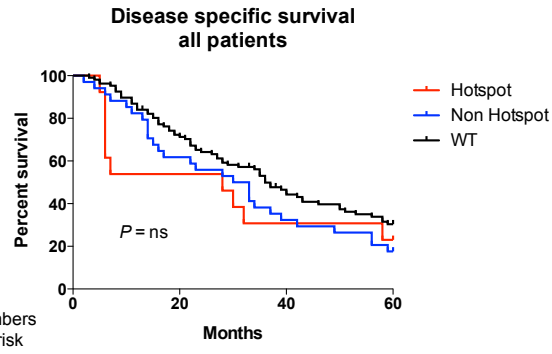
**Supplementary Figure 2**



# Supplementary Figure 3

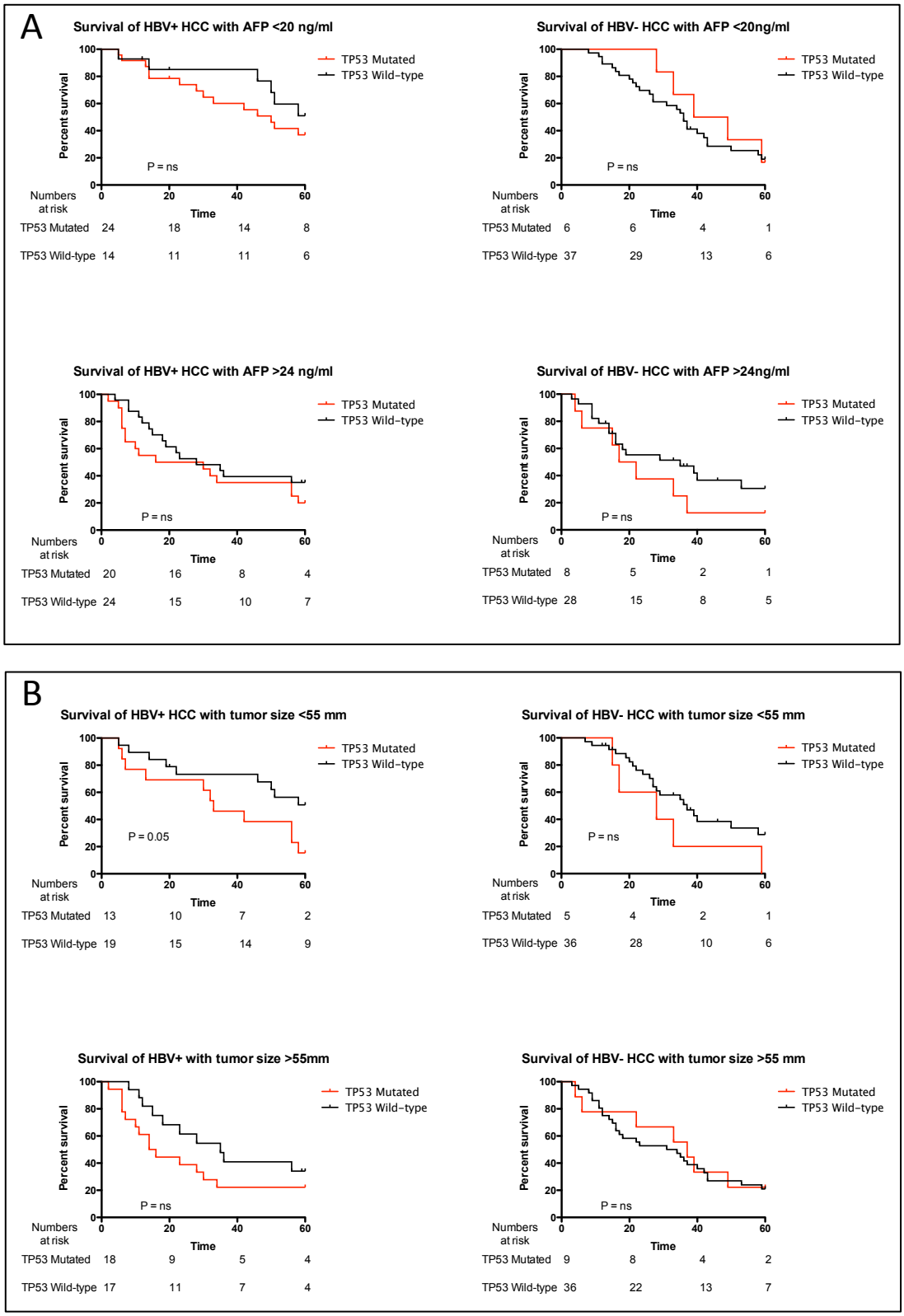



| Numbers at risk | | | | |
|---|---|---|---|---|
| Hotspot | 13 | 8 | 5 | 3 |
| Non hotspot | 18 | 11 | 7 | 3 |
| WT | 36 | 24 | 19 | 12 |

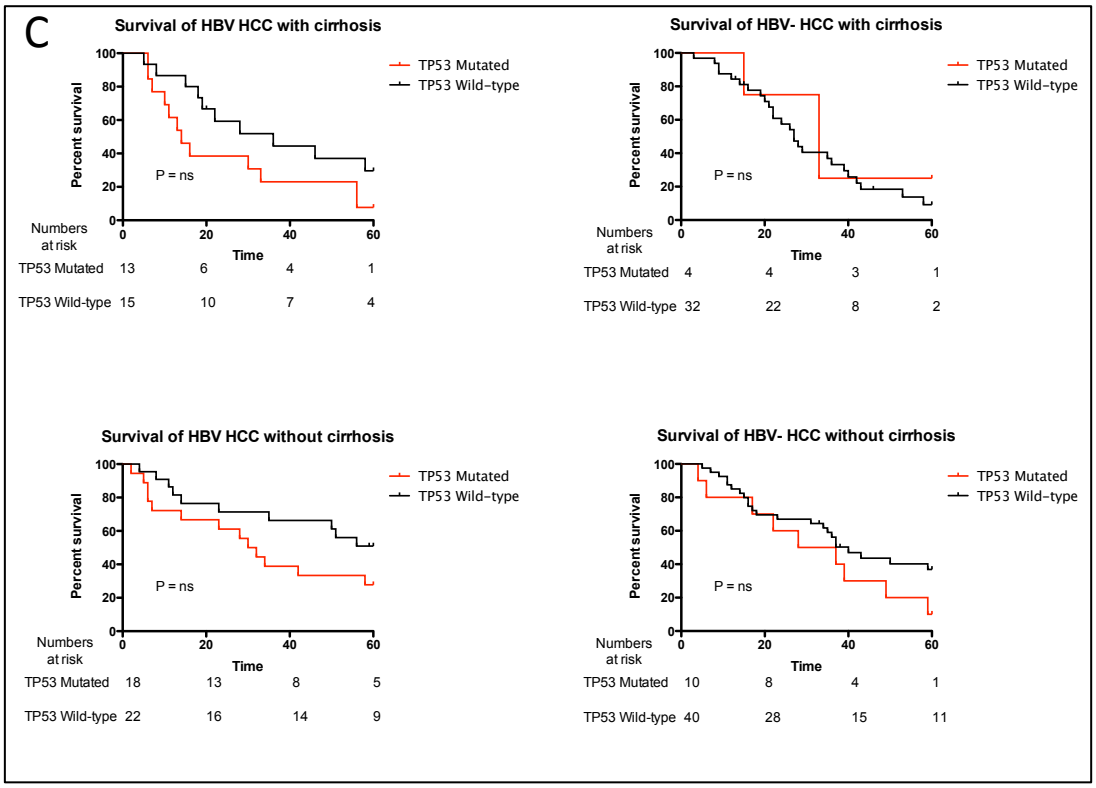


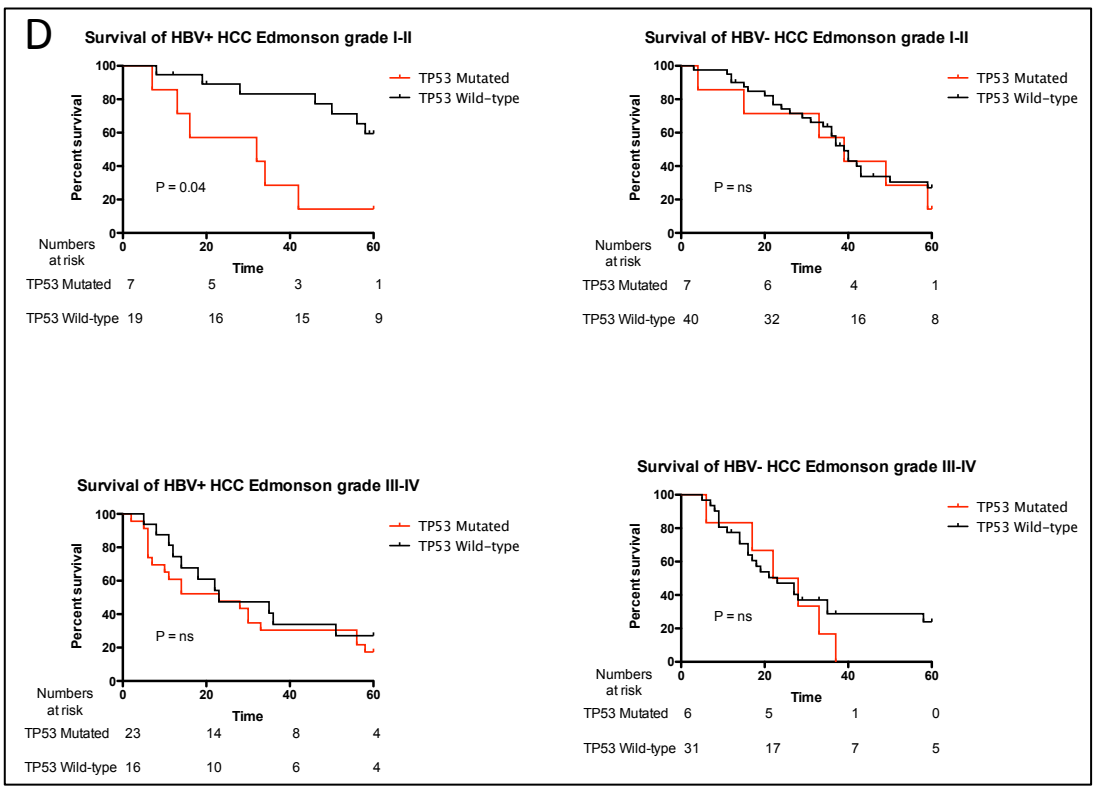

| Numbers at risk | | | | |
|---|---|---|---|---|
| Hotspot | 13 | 8 | 5 | 3 |
| Non hotspot | 34 | 22 | 12 | 6 |
| WT | 107 | 73 | 41 | 25 |

# Supplementary Figure 4



# Supplementary Figure 4



**Supplementary Table 1** Clinical, histological and pathological data of 86 HBV- and 90 non-HBV-related HCC**.** Shown are p Values obtained from Chi-square test based on the given clinical variable.

| | HBV-associated HCC | Non-HBV-associated HCC | p Value |
|---|---|---|---|
| **Age (Year)** | | | |
| <60 | 69% (59/86) | 22% (20/90) | <0.0001 |
| ≥60 | 31% (27/86) | 78% (70/90) | |
| **Alpha-fetoprotein [log10] (ng/mL)** | | | |
| ≤20 | 37% (28/75) | 57% (47/83) | 0.01 |
| >20 | 63% (47/75) | 43% (36/83) | |
| **Tumor size (mm)** | | | |
| ≤55 | 51% (43/85) | 47% (42/90) | ns |
| >55 | 49% (42/85) | 53% (48/90) | |
| **Number of tumor nodule(s)** | | | |
| Single | 76% (65/85) | 84% (76/90) | ns |
| Multiple | 24% (20/85) | 16% (14/90) | |

† All analyses were performed based on available data

**Supplementary Table 2** List of HBV primers

| Name | Sequence |
|---|---|
| P1 (1821 to 1841) | CCGGAAAGCTTGAGCTCTTCTTTTTCACCTCTGCCTAATCA |
| P2 (1823 to 1806) | CCGGAAAGCTTGAGCTCTTCAAAAAGTTGCATGGTGCTGG |
| P3 (1825 to 1841) | CTCGCTCGCCCAAATTTTTCACCTCTGCCTAATCA |
| P4 (1823 to 1806) | CTGGTTCGGCCCAAAAAGTTGCATGGTGCTGG |
| P5 (1928 to 1950) | GGAGCTACTGTGGAGTTACTCTC |
| P8 (67 to 90) | CTCCAGTTCAGGAACAGTAAACCC |
| P9 (634 to 656) | ATTCCTATGGGAGTGGGCCTCAG |
| P10 (1266 to 1286) | CCATACTGCGGAACTCCTAGC |
| P11 (2043 to 2021) | CAATGCTCAGGAGACTCTAAGGC |
| P12 (2400 to 2381) | CTTCGTCTGCGAGGCGAGGG |
| P15 (738 to 716) | ATAACTGAAAGCCAAACAGTGGG |
| P16 (1394 to 1372) | GCAGCACAGCCTAGCAGCCATGG |
| P17(2744 to2767) | CTATTTACACACTCTATGGAAGGC |
| P18 (3039 to 3061) | GGGGTGGAGCCCTCAGGCTCAGG |
| P19 (2510 to 2487) | AGGTACAGTAGAAGAATAAAGCCC |
| HBX3R (1751 to 1733) | GACCCTCCTCAACCCCCTC |
| HBX2R (1894 to 1875) | GGTTCGACACGGAACCCACC |
| HBS5R (840 to 820) | AGAAACCCATATGTAAATTTG |
| HBS3R (707 to 688) | CGGTAAACAAGTCACCAAGC |
| HBS2R (477 to 458) | GTTCCATACAACGGGCAAAC |
| HBS1F (127 to 147) | CTTCTCGAGGACTGGGGACCC |
| HBS2F (412 to 429) | CATCCTGCTATGCCT |
| HBS3F (640 to 656) | CCTATGGGAGTGGGCCT |
| HBX1F (1266 to 1285) | GATCCATACTGCGGAACTCC |
| HBX2F (1519 to 1537) | ACCACGGGGCGCACCTCTC |
| HBX3F (1548 to 1568) | CTCCCCGTCTGTGCCTTCTCA |
| HBX4F (1692 to 1713) | ACCGACCTTGAGGCATACTTCA |

**Supplementary Table 3** TaqMan® pre-designed gene expression assays

| Gene symbol | Name | Assay ID |
|---|---|---|
| AFP | Alpha-fetoprotein | Hs00173490_m1 |
| AMACR | Alpha-methylacyl-CoA racemase isoform 3 | Hs02786742_s1 |
| AURKA | Aurora kinase A | Hs01597773_mH |
| BIRC5 | Baculoviral IAP repeat-containing protein 5 isoform 2 | Hs00153353_m1 |
| CCNB1 | G2/mitotic-specific cyclin-B1 | Hs00259126_m1 |
| CDH2 | Cadherin 2, type 1, N-cadherin (neuronal) | Hs00169953_m1 |
| CRP | C-reactive protein | Hs00357041_m1 |
| CYP2C9 | Cytochrome P450, family 2, subfamily C, polypeptide 9 | Hs00426397_m1 |
| CYP3A7 | Cytochrome P450, family 3, subfamily A, polypeptide 7 | Hs00426361_m1 |
| EPCAM | Epithelial cell adhesion molecule | Hs00158980_m1 |
| EPHA1 | EPH receptor A1 | Hs00178313_m1 |
| FABP1 | Fatty acid binding protein 1, liver | Hs00155026_m1 |
| G0S2 | G0/G1 Switch Regulatory Protein 2 | Hs00274783_m1 |
| GLUL | Glutamine synthetase | Hs00374213_m1 |
| GPC3 | Glypican-3 isoform 1 | Hs00170471_m1 |
| HAL | Histidine ammonia-lyase isoform 2 | Hs00157887_m1 |
| HAMP | epcidin antimicrobial peptide | Hs00221783_m1 |
| HN1 | Hematological and neurological expressed 1 | Hs00602957_m1 |
| IGF2BP3 | Insulin-like growth factor 2 mRNA-binding protein 3 | Hs00365742_g1 |
| JUP;KRT19 | Keratin, type I cytoskeletal 19 | Hs01051611_gH |
| LAMA3 | Laminin, alpha 3 | Hs00165042_m1 |
| LGR5 | Leucine-rich repeat-containing G-protein coupled receptor 5 isoform 2 | Hs00173664_m1 |
| MERTK | C-mer proto-oncogene tyrosine kinase | Hs00179024_m1 |
| NEU1 | Sialidase-1 | Hs00166421_m1 |
| NRCAM | Neuronal cell adhesion molecule isoform A | Hs00170554_m1 |
| PAK2 | p21 protein (Cdc42/Rac)-activated kinase 2 | Hs00605586_m1 |
| PAP/REG3A | Regenerating Islet-Derived 3 Alpha | Hs00170171_m1 |
| PIR | Pirin (iron-binding nuclear protein) | Hs00186374_m1 |
| RAB1A | Member RAS Oncogene Family | Hs00366313_m1 |
| RAMP3 | Receptor (G protein-coupled) activity modifying protein 3 | Hs00389130_m1 |
| RHBG | Rh family, B glycoprotein | Hs00220735_m1 |
| SAA2 | Serum amyloid A-2 protein isoform b | Hs00754237_s1 |
| SAE1 | SUMO1 activating enzyme subunit 1 | Hs00271440_m1 |
| SPP1 | Secreted phosphoprotein 1/osteopontin | Hs00167093_m1 |
| TBX3 | T-box transcription factor TBX3 isoform 1 | Hs00255591_m1 |
| UGT2B7 | UDP glucuronosyltransferase 2 family, polypeptide B7 | Hs00426592_m1 |
| NRAS | Neuroblastoma RAS viral (v-ras) oncogene homolog | Hs00180035_m1 |

**Supplementary Table 4** Correlation between the HBV genotype in non-tumor tissue and the clinical data of HBV associated HCC. Shown are p Values obtained from Chi-square test based on the given clinical or viral variable.

| HBV Genotype | A | B | C | D | E | p Value |
|---|---|---|---|---|---|---|
| **Origin** | | | | | | |
| Africans | 19 | 2 | 0 | 6 | 10 | |
| Asians | 1 | 7 | 7 | 0 | 0 | <0.0001 |
| Europeans | 12 | 1 | 1 | 12 | 2 | |
| **Age (Years)** | | | | | | |
| < 60 | 27 | 6 | 6 | 7 | 10 | 0.01 |
| ≥ 60 | 5 | 4 | 2 | 11 | 2 | |
| **HDV co-infection** | | | | | | |
| No | 32 | 10 | 8 | 17 | 9 | 0.01 |
| Yes | 0 | 0 | 0 | 1 | 3 | |

**Supplementary Table 5** Mutations of TP53 gene, with p53 loss of- or gain of function found in HBV and non-HBV HCC.

| HCC-ID | Ensembl Gene ID | Gene symbol | Chromosome | Nucleotide (genomic hg19) | cDNA | Amino Acid | Loss of- and gain of-function* |
|---|---|---|---|---|---|---|---|
| CHC137T | ENSG00000141510 | TP53 | 17 | g.7574012C>A | c.1015G>T | p.Glu339X | Loss |
| CHC1210T | ENSG00000141510 | TP53 | 17 | g.7579586G>A | c.101C>T | p.Pro34Leu | Loss |
| CHC1987T | ENSG00000141510 | TP53 | 17 | g.7579548G>A | c.139C>T | p.Pro47Ser | Loss |
| CHC1205T | ENSG00000141510 | TP53 | 17 | g.7579373C>A | c.314G>T | p.Gly105Val | Loss |
| CHC1758T | ENSG00000141510 | TP53 | 17 | g.7578527del | c.403del | p.Cys135AlafsX35 | Loss |
| CHC996T | ENSG00000141510 | TP53 | 17 | g.7578457C>T | c.473G>A | p.Arg158His | Loss and gain |
| CHC1992T | ENSG00000141510 | TP53 | 17 | g.7578443A>T | c.487T>A | p.Tyr163Asn | Loss and gain |
| CHC1168T | ENSG00000141510 | TP53 | 17 | g.7578406C>T | c.524G>A | p.Arg175His | Loss |
| CHC080T | ENSG00000141510 | TP53 | 17 | g.7578406C>A | c.524G>T | p.Arg175Leu | Loss and gain |
| CHC008T | ENSG00000141510 | TP53 | 17 | g.7578272G>C | c.577C>G | p.His193Asp | Loss |
| CHC1987T | ENSG00000141510 | TP54 | 17 | g.7578203C>T | c.646G>A | p.Val216Met | Loss |
| CHC024T | ENSG00000141510 | TP53 | 17 | g.7578191A>T | c.658T>A | p.Tyr220Asn | Loss |
| CHC254T | ENSG00000141510 | TP53 | 17 | g.7577568C>A | c.713G>T | p.Cys238Phe | Loss |
| CHC245T | ENSG00000141510 | TP53 | 17 | g.7577550C>T | c.731G>A | p.Gly244Asp | Loss |
| CHC1194T | ENSG00000141510 | TP53 | 17 | g.7577545T>C | c.736A>G | p.Met246Val | Loss and gain |
| CHC1208T | ENSG00000141510 | TP53 | 17 | g.7577545T>C | c.736A>G | p.Met246Val | Loss and gain |
| CHC402T | ENSG00000141510 | TP53 | 17 | g.7577535C>T | c.746G>A | p.Arg249Lys | NA |
| CHC014T | ENSG00000141510 | TP53 | 17 | g.7577534C>A | c.747G>T | p.Arg249Ser | Loss and gain |
| CHC016T | ENSG00000141510 | TP53 | 17 | g.7577534C>A | c.747G>T | p.Arg249Ser | Loss and gain |
| CHC023T | ENSG00000141510 | TP53 | 17 | g.7577534C>A | c.747G>T | p.Arg249Ser | Loss and gain |
| CHC043T | ENSG00000141510 | TP53 | 17 | g.7577534C>A | c.747G>T | p.Arg249Ser | Loss and gain |
| CHC1035T | ENSG00000141510 | TP53 | 17 | g.7577534C>A | c.747G>T | p.Arg249Ser | Loss and gain |
| CHC1151T | ENSG00000141510 | TP53 | 17 | g.7577534C>A | c.747G>T | p.Arg249Ser | Loss and gain |
| CHC1704T | ENSG00000141510 | TP53 | 17 | g.7577534C>A | c.747G>T | p.Arg249Ser | Loss and gain |
| CHC1735T | ENSG00000141510 | TP53 | 17 | g.7577534C>A | c.747G>T | p.Arg249Ser | Loss and gain |
| CHC1754T | ENSG00000141510 | TP53 | 17 | g.7577534C>A | c.747G>T | p.Arg249Ser | Loss and gain |
| CHC198T | ENSG00000141510 | TP53 | 17 | g.7577534C>A | c.747G>T | p.Arg249Ser | Loss and gain |
| CHC1998T | ENSG00000141510 | TP53 | 17 | g.7577534C>A | c.747G>T | p.Arg249Ser | Loss and gain |
| CHC2000T | ENSG00000141510 | TP53 | 17 | g.7577534C>A | c.747G>T | p.Arg249Ser | Loss and gain |
| CHC226T | ENSG00000141510 | TP53 | 17 | g.7577534C>A | c.747G>T | p.Arg249Ser | Loss and gain |
| CHC1989T | ENSG00000141510 | TP53 | 17 | g.7577534C>A | c.747G>T | p.Arg249Ser | Loss and gain |
| CHC1756T | ENSG00000141510 | TP53 | 17 | g.7577111G>A | c.827C>T | p.Ala276Val | Loss |
| CHC1211T | ENSG00000141510 | TP53 | 17 | g.7577058C>A | c.880G>T | p.Glu294X | Loss |
| CHC1984T | ENSG00000141510 | TP53 | 17 | g.7579698A>C | c.96+2T>G | splicing mutation | Loss |
| CHC736T | ENSG00000141510 | TP53 | 17 | g.7576852C>G | c.993+1G>C | splicing mutation | NA |
| CHC1061T | ENSG00000141514 | TP53 | 17 | g.7574017C>A | c.1010G>T | p.Arg337Leu | Loss |
| CHC1055T | ENSG00000141513 | TP53 | 17 | g.7579502_7579518dup | c.169_185dup | p.Ala63ThrfsX66 | Loss |
| CHC304T | ENSG00000141522 | TP53 | 17 | g.7579398_7579439dup | c.248_289dup | p.Ala83_Ser96dup | Loss |

| HCC-ID | Ensembl Gene ID | Gene symbol | Chromosome | Nucleotide (genomic hg19) | cDNA | Amino Acid | Loss of- and gain of-function* |
|---|---|---|---|---|---|---|---|
| CHC1607T | ENSG00000141519 | TP53 | 17 | g.7578226del | c.263del | p.Ser90ProfsX33 | Loss |
| CHC037T | ENSG00000141512 | TP53 | 17 | g.7579420dup | c.267dup | p.Ser90LeufsX59 | NA |
| CHC1604T | ENSG00000141518 | TP53 | 17 | g.7578496A>T | c.434T>A | p.Leu145Gln | Loss |
| CHC129T | ENSG00000141516 | TP53 | 17 | g.7578445A>C | c.485T>G | p.Ile162Ser | Loss |
| CHC208T | ENSG00000141521 | TP53 | 17 | g.7578190T>G | c.659A>C | p.Tyr220Ser | Loss |
| CHC154T | ENSG00000141517 | TP53 | 17 | g.7577602del | c.679del | p.Ser227LeufsX20 | NA |
| CHC013T | ENSG00000141511 | TP53 | 17 | g.7577556C>A | c.725G>T | p.Cys242Phe | Loss |
| CHC1616T | ENSG00000141520 | TP53 | 17 | g.7577536T>C | c.745A>G | p.Arg249Gly | Loss |
| CHC126T | ENSG00000141515 | TP53 | 17 | g.7577085C>T | c.853G>A | p.Glu285Lys | Loss |
| CHC314T | ENSG00000141523 | TP53 | 17 | g.7578395G>A | c.535C>T | p.His179Tyr | Loss and gain |
| CHC793T | ENSG00000141524 | TP53 | 17 | g.7573982C>A | c.1045G>T | p.Glu349X | Loss |

*According to IARC TP53 database (*http://p53.iarc.fr*)

**Supplementary Table 6** Mutations of different genes found in HBV and non-HBV HCC.

| HCC-ID | Ensembl Gene ID | Gene symbol | Chromosome | Nucleotide (genomic hg19) | cDNA | Amino Acid |
|---|---|---|---|---|---|---|
| CHC012T | ENSG00000168310 | IRF2 | 4 | Homozygous deletion | Deletion | Deletion |
| CHC018T | ENSG00000168310 | IRF2 | 4 | Homozygous deletion | Deletion | Deletion |
| CHC034T | ENSG00000168310 | IRF2 | 4 | Homozygous deletion | Deletion | Deletion |
| CHC239T | ENSG00000168310 | IRF2 | 4 | Homozygous deletion | Deletion | Deletion |
| CHC339T | ENSG00000168310 | IRF2 | 4 | g.185329422_185329435del | c.412-6_419del | Splicing mutation |
| CHC398T | ENSG00000168310 | IRF2 | 4 | g.185339323T>G | c.409A>C | p.Lys137Gln |
| CHC1982T | ENSG00000168036 | CTNNB1 | 3 | g.41266112T>C | c.109T>C | p.Ser37Pro |
| CHC252T | ENSG00000168036 | CTNNB1 | 3 | g.41268766A>T | c.1004A>T | p.Lys335Ile |
| CHC1208T | ENSG00000168036 | CTNNB1 | 3 | g.41266106_41266117del | c.103_114del | p.Ile35_Gly38del |
| CHC609T | ENSG00000168036 | CTNNB1 | 3 | g.41266113C>G | c.110C>G | p.Ser37Cys |
| CHC1991T | ENSG00000168036 | CTNNB1 | 3 | g.41274911T>A | c.1161T>A | p.Asn387Lys |
| CHC1702T | ENSG00000168036 | CTNNB1 | 3 | g.41274911T>G | c.1161T>G | p.Asn387Lys |
| CHC1756T | ENSG00000168036 | CTNNB1 | 3 | g.41266125C>T | c.122C>T | p.Thr41Ile |
| CHC1982T | ENSG00000168036 | CTNNB1 | 3 | g.41266013A>G | c.14-4A>G | exon 3 deletion |
| CHC1142T | ENSG00000168036 | CTNNB1 | 3 | g.41266050_41266603del | c.47_400del | p.Pro16_Lys133del |
| CHC1760T | ENSG00000168036 | CTNNB1 | 3 | g.41266097G>C | c.94G>C | p.Asp32His |
| CHC1989T | ENSG00000168036 | CTNNB1 | 3 | g.41266097G>T | c.94G>T | p.Asp32Tyr |
| CHC1717T | ENSG00000168036 | CTNNB1 | 3 | g.41266098A>G | c.95A>G | p.Asp32Gly |
| CHC335T | ENSG00000168036 | CTNNB1 | 3 | g.41266100T>C | c.97T>C | p.Ser33Pro |
| CHC996T | ENSG00000168036 | CTNNB1 | 3 | g.41266101C>G | c.98C>G | p.Ser33Cys |
| CHC1603T | ENSG00000168036 | CTNNB1 | 3 | g.41268766A>T | c.1004A>T | p.Lys335Ile |
| CHC230T | ENSG00000168036 | CTNNB1 | 3 | g.41265566_41266581del | c.1004A>T | p.Lys335Ile |
| CHC429T | ENSG00000168036 | CTNNB1 | 3 | g.41266103G>A | c.100G>A | p.Gly34Arg |
| CHC469T | ENSG00000168036 | CTNNB1 | 3 | g.41266103G>A | c.100G>A | p.Gly34Arg |
| CHC305T | ENSG00000168036 | CTNNB1 | 3 | g.41266106_41266117del | c.103_114del | p.Ile35_Gly38del |
| CHC168T | ENSG00000168036 | CTNNB1 | 3 | g.41266107T>G | c.104T>G | p.Ile35Ser |
| CHC1553T | ENSG00000168036 | CTNNB1 | 3 | g.41266110A>C | c.107A>C | p.His36Pro |
| CHC433T | ENSG00000168036 | CTNNB1 | 3 | g.41274899G>T | c.1149G>T | p.Trp383Cys |
| CHC983T | ENSG00000168036 | CTNNB1 | 3 | g.41274899G>T | c.1149G>T | p.Trp383Cys |
| CHC399T | ENSG00000168036 | CTNNB1 | 3 | g.41266124A>G | c.121A>G | p.Thr41Ala |
| CHC437T | ENSG00000168036 | CTNNB1 | 3 | g.41266124A>G | c.121A>G | p.Thr41Ala |
| CHC430T | ENSG00000168036 | CTNNB1 | 3 | g.41266125C>A | c.122C>A | p.Thr41Asn |
| CHC614T | ENSG00000168036 | CTNNB1 | 3 | g.41266125C>A | c.122C>A | p.Thr41Asn |
| CHC051T | ENSG00000168036 | CTNNB1 | 3 | g.41266129_41266134dup | c.126_131dup | p.Ala43_Pro44dup |
| CHC013T | ENSG00000168036 | CTNNB1 | 3 | g.41266136T>C | c.133T>C | p.Ser45Pro |
| CHC028T | ENSG00000168036 | CTNNB1 | 3 | g.41266136T>C | c.133T>C | p.Ser45Pro |

| HCC-ID | Ensembl Gene ID | Gene symbol | Chromosome | Nucleotide (genomic hg19) | cDNA | Amino Acid |
|---|---|---|---|---|---|---|
| CHC1040T | ENSG00000168036 | CTNNB1 | 3 | g.41266136T>C | c.133T>C | p.Ser45Pro |
| CHC115T | ENSG00000168036 | CTNNB1 | 3 | g.41266136T>C | c.133T>C | p.Ser45Pro |
| CHC164T | ENSG00000168036 | CTNNB1 | 3 | g.41266136T>C | c.133T>C | p.Ser45Pro |
| CHC302T | ENSG00000168036 | CTNNB1 | 3 | g.41266136T>C | c.133T>C | p.Ser45Pro |
| CHC1041T | ENSG00000168036 | CTNNB1 | 3 | g.41266136T>G | c.133T>G | p.Ser45Ala |
| CHC121T | ENSG00000168036 | CTNNB1 | 3 | g.41266136T>G | c.133T>G | p.Ser45Ala |
| CHC197T | ENSG00000168036 | CTNNB1 | 3 | g.41266137C>A | c.134C>A | p.Ser45Tyr |
| CHC1566T | ENSG00000168036 | CTNNB1 | 3 | g.41266137C>T | c.134C>T | p.Ser45Phe |
| CHC1584T | ENSG00000168036 | CTNNB1 | 3 | g.41265926_41266306del | c.14-91_241+62del | exon 3 deletion |
| CHC211T | ENSG00000168036 | CTNNB1 | 3 | g.41265926_41266306del | c.14-91_241+62del | exon 3 deletion |
| CHC433T | ENSG00000168036 | CTNNB1 | 3 | g.41242294A>G | c.874A>G | p.Lys292Glu |
| CHC059T | ENSG00000168036 | CTNNB1 | 3 | g.41266098A>C | c.95A>C | p.Asp32Ala |
| CHC1069T | ENSG00000168036 | CTNNB1 | 3 | g.41266098A>C | c.95A>C | p.Asp32Ala |
| CHC1545T | ENSG00000168036 | CTNNB1 | 3 | g.41266098A>G | c.95A>G | p.Asp32Gly |
| CHC361TA | ENSG00000168036 | CTNNB1 | 3 | g.41266098A>G | c.95A>G | p.Asp32Gly |
| CHC037T | ENSG00000168036 | CTNNB1 | 3 | g.41266098A>T | c.95A>T | p.Asp32Val |
| CHC097T | ENSG00000168036 | CTNNB1 | 3 | g.41266100T>C | c.97T>C | p.Ser33Pro |
| CHC1146T | ENSG00000168036 | CTNNB1 | 3 | g.41266100T>C | c.97T>C | p.Ser33Pro |
| CHC1565T | ENSG00000168036 | CTNNB1 | 3 | g.41266101C>G | c.98C>G | p.Ser33Cys |
| CHC242T | ENSG00000168036 | CTNNB1 | 3 | g.41266101C>G | c.98C>G | p.Ser33Cys |
| CHC301T | ENSG00000168036 | CTNNB1 | 3 | g.41266101C>G | c.98C>G | p.Ser33Cys |
| CHC432T | ENSG00000168036 | CTNNB1 | 3 | g.41266101C>G | c.98C>G | p.Ser33Cys |
| CHC798T | ENSG00000168036 | CTNNB1 | 3 | g.41266101C>G | c.98C>G | p.Ser33Cys |
| CHC1052T | ENSG00000168036 | CTNNB1 | 3 | g.41266101C>T | c.98C>T | p.Ser33Phe |
| CHC130T | ENSG00000168036 | CTNNB1 | 3 | g.41266101C>T | c.98C>T | p.Ser33Phe |
| CHC1983T | ENSG00000103126 | AXIN1 | 16 | g.360071T>A | c.1020-2A>T | splicing mutation |
| CHC079T | ENSG00000103126 | AXIN1 | 16 | g.354370_354373del | c.1185_1188del | p.Leu396ArgfsX17 |
| CHC736T | ENSG00000103126 | AXIN1 | 16 | g.348223G>T | c.1283C>A | p.Ser428X |
| CHC237T | ENSG00000103126 | AXIN1 | 16 | g.348122C>T | c.1384G>A | p.Ala462Thr |
| CHC1154T | ENSG00000103126 | AXIN1 | 16 | g.347929G>A | c.1577C>T | p.Ala526Val |
| CHC226T | ENSG00000103126 | AXIN1 | 16 | g.347764_347765del | c.1741_1742del | p.Ser581CysfsX9 |
| CHC250T | ENSG00000103126 | AXIN1 | 16 | g.347143G>T | c.1868C>A | p.Ser623X |
| CHC080T | ENSG00000103126 | AXIN1 | 16 | g.396664C>A | c.362G>T | p.Cys121Phe |
| CHC018T | ENSG00000103126 | AXIN1 | 16 | g.396458G>A | c.568C>T | p.Gln190X |
| CHC1758T | ENSG00000103126 | AXIN1 | 16 | g.396233C>A | c.793G>T | p.Gly265X |
| CHC2000T | ENSG00000103126 | AXIN1 | 16 | g.396941C>A | c.85G>T | p.Glu29X |
| CHC204T | ENSG00000103126 | AXIN1 | 16 | g.364661del | c.901del | p.Val301SerfsX113 |
| CHC237T | ENSG00000103126 | AXIN1 | 16 | g.364557_364563dup | c.999_1005dup | p.Thr336ValfsX17 |

| HCC-ID | Ensembl Gene ID | Gene symbol | Chromosome | Nucleotide (genomic hg19) | cDNA | Amino Acid |
|---|---|---|---|---|---|---|
| CHC012T | ENSG00000103126 | AXIN1 | 16 | Homozygous deletion | Deletion | Deletion |
| CHC314T | ENSG00000103126 | AXIN1 | 16 | g.360068del | c.1021del | p.Asp341MetfsX73 |
| CHC1052T | ENSG00000103126 | AXIN1 | 16 | g.397014T>G | c.12A>C | p.Gln4His |
| CHC1190T | ENSG00000103126 | AXIN1 | 16 | g.347909G>A | c.1597C>T | p.Arg533X |
| CHC1065T | ENSG00000103126 | AXIN1 | 16 | g.354326dup | c.1232dup | p.Arg412AlafsX12 |
| CHC1053T | ENSG00000103126 | AXIN1 | 16 | g.347930C>T | c.1576G>A | p.Ala526Thr |
| CHC1201T | ENSG00000103126 | AXIN1 | 16 | g.347903C>T | c.1603G>A | p.Val535Ile |
| CHC434T | ENSG00000103126 | AXIN1 | 16 | g.347772G>T | c.1734C>A | p.Tyr578X |
| CHC1545T | ENSG00000103126 | AXIN1 | 16 | g.347168C>T | c.1843G>A | p.Ala615Thr |
| CHC1585T | ENSG00000103126 | AXIN1 | 16 | g.341227del | c.2257del | p.Val753TyrfsX48 |
| CHC438T | ENSG00000103126 | AXIN1 | 16 | g.396962G>A | c.64C>T | p.Arg22X |
| CHC031T | ENSG00000103126 | AXIN1 | 16 | Homozygous deletion | Deletion | Deletion |
| CHC1199T | ENSG00000103126 | AXIN1 | 16 | Homozygous deletion | Deletion | Deletion |
| CHC1168T | ENSG00000117713 | ARID1A | 1 | g.27024008_27024009insG | c.1114_1115insG | p.Gln372ArgfsX28 |
| CHC1211T | ENSG00000117713 | ARID1A | 1 | g.27023020_27023022dup | c.126_128dup | p.Ala45dup |
| CHC1754T | ENSG00000117713 | ARID1A | 1 | g.27023020_27023022dup | c.126_128dup | p.Ala45dup |
| CHC241T | ENSG00000117713 | ARID1A | 1 | g.27089501dup | c.2457dup | p.Asn820GlnfsX52 |
| CHC014T | ENSG00000117713 | ARID1A | 1 | g.27089676C>T | c.2632C>T | p.Gln878X |
| CHC1758T | ENSG00000117713 | ARID1A | 1 | g.27023253_27023266del | c.359_372del | p.Pro120ArgfsX275 |
| CHC339T | ENSG00000117713 | ARID1A | 1 | g.27106159delinsAA | c.5770delinsAA | p.Glu1924LysfsX5 |
| CHC335T | ENSG00000117713 | ARID1A | 1 | g.27106649G>A | c.6260G>A | p.Gly2087Glu |
| CHC433T | ENSG00000117713 | ARID1A | 1 | g.27023909del | c.1015del | p.Ala339LeufsX24 |
| CHC434T | ENSG00000117713 | ARID1A | 1 | g.27023923_27023937del | c.1029_1043del | p.Ala345_Ala349del |
| CHC155T | ENSG00000117713 | ARID1A | 1 | g.27056173dup | c.1169dup | p.Met390IlefsX10 |
| CHC1040T | ENSG00000117713 | ARID1A | 1 | g.27056214C>T | c.1210C>T | p.Gln404X |
| CHC445T | ENSG00000117713 | ARID1A | 1 | g.27087897_27087911delinsT | c.2184_2198delinsT | p.Pro729GlyfsX83 |
| CHC121T | ENSG00000117713 | ARID1A | 1 | g.27023209_27023228del | c.315_334del | p.Asn106ProfsX4 |
| CHC1053T | ENSG00000117713 | ARID1A | 1 | g.27023214_27023223del | c.320_329del | p.Ala107GlyfsX4 |
| CHC437T | ENSG00000117713 | ARID1A | 1 | g.27097692dup | c.3281dup | p.Gln1095AlafsX10 |
| CHC983T | ENSG00000117713 | ARID1A | 1 | g.27099008C>T | c.3424C>T | p.Gln1142X |
| CHC205T | ENSG00000117713 | ARID1A | 1 | g.27099939_27099964delinsG | c.3818_3843delinsG | p.Met1273ArgfsX8 |
| CHC230T | ENSG00000117713 | ARID1A | 1 | g.27023381G>A | c.487G>A | p.Ala163Thr |
| CHC211T | ENSG00000117713 | ARID1A | 1 | g.27106934_27106953del | c.6545_6564del | p.Ala2182GlufsX36 |
| CHC013T | ENSG00000117713 | ARID1A | 1 | g.27106961_27106969del | c.6572_6580del | p.Ser2191_Gly2193del |
| CHC031T | ENSG00000117713 | ARID1A | 1 | Homozygous deletion | Deletion | Deletion |
| CHC079T | ENSG00000189079 | ARID2 | 12 | g.46231205G>C | c.1120+5G>C | Splicing mutation |

| HCC-ID | Ensembl Gene ID | Gene symbol | Chromosome | Nucleotide (genomic hg19) | cDNA | Amino Acid |
|---|---|---|---|---|---|---|
| CHC1999T | ENSG00000189079 | ARID2 | 12 | g.46231425T>C | c.1265T>C | p.Leu422Pro |
| CHC024T | ENSG00000189079 | ARID2 | 12 | g.46233199C>T | c.1418C>T | p.Ser473Phe |
| CHC123T | ENSG00000189079 | ARID2 | 12 | g.46243449G>A | c.1802G>A | p.Arg601Gln |
| CHC024T | ENSG00000189080 | ARID2 | 12 | g.46244230T>A | c.2324T>A | p.Leu775X |
| CHC1044T | ENSG00000189080 | ARID2 | 12 | g.46231153C>G | c.1073C>G, | p.Thr358Ser |
| CHC1584T | ENSG00000189080 | ARID2 | 12 | g.46231164_46231190del | c.1084_1110del | p.Cys362_Leu370del |
| CHC126T | ENSG00000189080 | ARID2 | 12 | g.46240646_46240647dup | c.1506_1507dup | p.Ala503GlufsX4 |
| CHC429T | ENSG00000189080 | ARID2 | 12 | g.46243361A>G | c.1716-2A>G | Splicing mutation |
| CHC429T | ENSG00000189080 | ARID2 | 12 | g.46244040G>T | c.2134G>T | p.Glu712X |
| CHC614T | ENSG00000189080 | ARID2 | 12 | g.46287229C>G | c.5174C>G | p.Ser1725X |
| CHC1199T | ENSG00000189080 | ARID2 | 12 | Homozygous deletion | Deletion | Deletion |
| CHC1986T | ENSG00000121879 | PIK3CA | 3 | g.178936091G>A | c.1633G>A | p.Glu545Lys |
| CHC1604T | ENSG00000121879 | PIK3CA | 3 | g.178952085A>G | c.3140A>G | p.His1047Arg |
| CHC235T | ENSG00000121879 | PIK3CA | 3 | g.178952085A>G | c.3140A>G | p.His1047Arg |
| CHC438T | ENSG00000121879 | PIK3CA | 3 | g.178952085A>T | c.3140A>T | p.His1047Leu |
| CHC1154T | ENSG00000177189 | RPS6KA3 | X | g.20185727C>A | c.1582G>T | p.Glu528X |
| CHC034T | ENSG00000177189 | RPS6KA3 | X | g.20185706C>T | c.1602+1G>A | p.? |
| CHC018T | ENSG00000177189 | RPS6KA3 | X | g.20183143_20183145delinsC | c.1636_1638delinsG | p.Leu546ValfsX4 |
| CHC1999T | ENSG00000177189 | RPS6KA3 | X | g.20181092T>C | c.1831A>G | p.Met611Val |
| CHC226T | ENSG00000177189 | RPS6KA3 | X | g.20206644A>G | c.602T>C | p.Leu201Pro |
| CHC037T | ENSG00000177189 | RPS6KA3 | X | g.20183078A>T | c.1703T>A | p.Leu568Gln |
| CHC1053T | ENSG00000177189 | RPS6KA3 | X | g.20181154A>C | c.1769T>G | p.Leu590X |
| CHC429T | ENSG00000177189 | RPS6KA3 | X | g.20222175T>C | c.290A>G | p.Tyr97Cys |
| CHC115T | ENSG00000177189 | RPS6KA3 | X | g.20206653C>G | c.594-1G>C | p.? |
| CHC1044T | ENSG00000177189 | RPS6KA3 | X | g.20284689C>A | c.62G>T | p.Ser21Ile |
| CHC434T | ENSG00000177189 | RPS6KA3 | X | g.20206044A>T | c.676T>A | p.Tyr226Asn |
| CHC1053T | ENSG00000177189 | RPS6KA3 | X | g.20205972C>T | c.748G>A | p.Asp250Asn |
| CHC1044T | ENSG00000177189 | RPS6KA3 | X | g.20205954C>T | c.766G>A | p.Val256Met |
| CHC258T | ENSG00000177189 | RPS6KA3 | X | g.20194611C>A | c.940G>T | p.Gly314X |
| CHC1041T | ENSG00000116044 | NFE2L2 | 2 | g.178098807T>G | c.238A>C | p.Thr80Pro |
| CHC059T | ENSG00000116044 | NFE2L2 | 2 | g.178098797_178098805del | c.240_248del | p.Gly81_Phe83del |
| CHC1069T | ENSG00000116044 | NFE2L2 | 2 | g.178098803C>A | c.242G>T | p.Gly81Val |
| CHC614T | ENSG00000116044 | NFE2L2 | 2 | g.178098800T>C | c.245A>G | p.Glu82Gly |
| CHC1040T | ENSG00000116044 | NFE2L2 | 2 | g.178098799T>G | c.246A>C | p.Glu82Asp |
| CHC614T | ENSG00000116044 | NFE2L2 | 2 | g.178098789T>A | c.256A>T | p.Ile86Phe |
| CHC1190T | ENSG00000116044 | NFE2L2 | 2 | g.178098960C>G | c.85G>C | p.Asp29His |
| CHC205T | ENSG00000116044 | NFE2L2 | 2 | g.178098959T>C | c.86A>G | p.Asp29Gly |