

Université de Paris

École doctorale : Hématologie, Oncogénèse, Biothérapies – ED561

*Inserm-UMR1138, Génomique fonctionnelle des tumeurs solides
Centre de Recherche des Cordeliers (CRC)*

Mécanismes moléculaires et conséquences oncogéniques des intégrations du Virus de l'Hépatite B dans les tissus hépatiques

Par Camille PÉNEAU

Thèse de doctorat en Oncogénèse

Dirigée par le Pr. Jessica ZUCMAN-ROSSI

Présentée et soutenue publiquement le 4 septembre 2020

Devant un jury composé de :

Dr. Hélène STRICK-MARCHAND, CR Pasteur, Institut Pasteur, Paris, Rapporteur
Dr. Romain PARENT, CR, Centre de Recherche en Cancérologie, Lyon, Rapporteur
Dr. Alain NICOLAS, DR émérite, Institut Curie, Paris, Examineur
Pr. Hugues de THÉ, PU-PH, Collège de France, Université de Paris, Examineur



Except where otherwise noted, this is work licensed under
<https://creativecommons.org/licenses/by-nc-nd/3.0/fr/>

RÉSUMÉ

Mécanismes moléculaires et conséquences oncogéniques des intégrations du Virus de l'Hépatite B dans les tissus hépatiques

Malgré l'existence d'un vaccin efficace et de traitements supprimant la réplication virale, l'infection par le Virus de l'Hépatite B (VHB) reste une des maladies chroniques les plus fréquentes, avec une mortalité associée dans 39% des cas au développement d'un carcinome hépatocellulaire (CHC), le cancer primitif du foie le plus courant et la troisième cause mondiale de décès par cancer. Le VHB est en effet le facteur de risque principal de survenue d'un CHC, chez des patients ayant généralement une cirrhose du foie induite par l'infection. Cependant, le fait que certains CHC liés au VHB surviennent sans inflammation chronique souligne les propriétés oncogéniques directes de ce virus à ADN, qui peut promouvoir la transformation cellulaire des hépatocytes en s'intégrant dans le génome humain. Ce projet a eu pour but de décrire les formes du VHB présentes dans des tissus hépatiques tumoraux et non-tumoraux de 177 patients majoritairement d'origine africaine et européenne, en utilisant des techniques de capture virale et de séquençage de nouvelle génération, et de caractériser les intégrations du virus en fonction des données génétiques et cliniques des patients. Nous avons montré que les tissus non-tumoraux contiennent plus fréquemment de l'ADN viral répliatif et un nombre total plus élevé d'insertions, principalement localisées dans des régions de chromatine ouverte mais sans conséquence fonctionnelle directe. Dans les tumeurs en revanche, les intégrations du VHB sont souvent clonales, enrichies à proximité de gènes impliqués dans la carcinogénèse hépatique comme *TERT* (dans un tiers des CHC liés au VHB), *CCNE1*, ou *KMT2B*, et peuvent entraîner directement le développement tumoral en activant ces gènes en *cis*. Les intégrations du VHB dans *CCNA2* ou *CCNE1* génèrent par exemple un stress répliatif et une signature de réarrangements structuraux spécifique, favorisant le développement de CHC agressifs en absence de cirrhose. Nous avons décrit par ailleurs un nouveau mécanisme oncogénique associé aux intégrations du VHB qui repose sur des réarrangements du génome humain délimités par des séquences virales intégrées, qui induisent des altérations du nombre de copies de gènes « driver » situés à distance comme *TP53* ou *MYC*. Nous avons donc approfondi la caractérisation des intégrations virales du VHB dans les CHC, mais également celles du virus adéno-associé (AAV) qui peut également s'intégrer dans l'ADN humain et favoriser le développement tumoral par mutagénèse insertionnelle en altérant les mêmes gènes que le VHB (*TERT*, *CCNA2*, *CCNE1*, *KMT2B*).

Mots-clés : Carcinome hépatocellulaire, Virus de l'Hépatite B, Oncovirus, Génomique du cancer, Capture virale, Intégration virale, Evolution clonale, Mutagénèse insertionnelle, Altérations du nombre de copies, Reconstruction de génomes

ABSTRACT

Molecular mechanisms and oncogenic consequences of Hepatitis B Virus integrations in liver tissues

Despite the existence of an effective vaccine and of treatments that suppress viral replication, Hepatitis B Virus (HBV) infection remains one of the most frequent chronic diseases. 39% of HBV-related deaths are associated with the development of hepatocellular carcinoma (HCC), the most common primary liver cancer and the third leading cause of cancer death worldwide. HBV is indeed the main risk factor of HCC development in patients who generally already have a liver cirrhosis induced by the infection. However, the fact that some HBV-related HCC occur without chronic inflammation underlines the direct oncogenic properties of this DNA virus, which can promote hepatocyte cell transformation through integration into the human genome. This project aimed to describe the HBV genomes in tumor and non-tumor liver tissues from 177 patients, mostly with African and European origin, using viral capture and next-generation sequencing techniques, and characterized viral integrations according to the genetic and clinical data of the patients. We showed that non-tumor tissues contain more frequently replicating HBV DNA and a higher number of insertions, mainly located in open chromatin regions but without direct functional consequences. In tumors, on the other hand, HBV integrations are often clonal and enriched in proximity of genes involved in hepatocarcinogenesis such as *TERT* (in one-third of HBV-related HCC), *CCNE1*, or *KMT2B*, and can directly lead to tumor development by activating these genes in *cis*. HBV integrations in *CCNA2* or *CCNE1*, for example, generate replicative stress and a specific signature of structural rearrangements, thus promoting the development of aggressive HCC in the absence of cirrhosis. We also described a novel oncogenic mechanism associated with HBV integrations based on rearrangements of the human genome delimited by integrated viral sequences, which induce copy number alterations of distant "driver" genes such as *TP53* or *MYC*. We have therefore further characterized the viral integrations of HBV in HCC, but also those of the adeno-associated virus (AAV) which can also integrate into human DNA and promote tumor development through insertional mutagenesis by altering the same genes as HBV (*TERT*, *CCNA2*, *CCNE1*, *KMT2B*).

Keywords : Hepatocellular carcinoma, Hepatitis B Virus, Oncovirus, Cancer genomics, Viral capture, Viral integration, Clonal evolution, Insertional mutagenesis, Copy number alterations, Genome reconstruction

REMERCIEMENTS

1er septembre 2016 – 4 septembre 2020, 4 années intenses et riches en émotion et je souhaite remercier tous ceux et celles qui ont contribué à ce projet, pour leur aide, leur soutien et leur amitié.

Je tiens tout d'abord à remercier le Dr Hélène Strick-Marchand et le Dr Romain Parent pour le temps que vous avez consacré à l'évaluation de mes travaux et pour l'intérêt que vous y avez porté. Merci également au Pr Hugues de Thé et au Dr Alain Nicolas d'avoir accepté d'être examinateurs de ce travail.

Cette thèse n'aurait pas pu avoir lieu si Jessica ne m'avait pas proposé ce projet en mars 2016, lorsque je faisais le tour des laboratoires parisiens à la recherche d'un sujet stimulant, et bien sûr je n'ai pas hésité très longtemps avant d'accepter. Au final, c'était certainement une aventure un peu ambitieuse, mais elle a été passionnante et m'a énormément appris sur le plan scientifique et personnel. Je ne connaissais pas grand-chose à la biologie lorsque j'ai commencé mon doctorat et encore moins à la bio-informatique, et j'ai souvent été un peu désemparée avec mon virus, mais je te suis extrêmement reconnaissante Jessica de m'avoir fait confiance tout au long de ces quatre années. Merci de m'avoir donné l'opportunité de naviguer en totale autonomie entre expérimental et informatique, de réaliser ce projet dans les meilleures conditions, d'apprendre la patience et la persévérance indispensables à la recherche, et d'aller présenter mes travaux à de nombreux congrès internationaux. Je te remercie aussi d'avoir toujours pris le temps d'écouter mes questionnements existentiels sur la suite de mon parcours professionnel pour me donner de précieux conseils car cela m'a permis d'avancer plus vite. Je suis très honorée d'avoir pu travailler sous ta direction et d'avoir pu être inspirée par ta science et ton énergie.

Merci à l'équipe « Virus », les irréductibles qui résistent encore à l'envahisseur ! Je suis arrivée comme un intrus HBV dans votre monde AAV2, mais vous m'avez accueillie chaleureusement et j'y ai facilement trouvé ma place. Je remercie Sandrine d'avoir eu la patience de m'apprendre les débuts de l'informatique et d'avoir toujours été prête à répondre à mes questions interminables et à améliorer mes scripts incompréhensibles. Merci à Iadh qui m'a initié aux expériences dès les premières semaines et qui m'a toujours donné des conseils avisés, et évidemment un immense merci à Tiziana qui m'a montré la voie et que j'ai essayé de suivre de près.

Un grand merci aux bio-infos qui ont eu un peu de mal à m'accepter dans leur groupe mais qui ont bien été obligés de reconnaître que j'étais une des leurs après la « Remontada » à Nantes. Je remercie tout particulièrement Stefano et Théo, pour m'avoir soutenue, conseillée et écoutée me plaindre pendant des heures lorsque que je ne savais plus comment avancer. Merci pour toutes nos longues discussions scientifiques et les moments partagés en écoutant du blues à Chicago ou en jouant à la Scopa à Prague : vous avez beaucoup contribué à cette thèse et je vous dois énormément ! Jean-Baptiste, je suis très heureuse que tu prennes la relève pour les

surveiller même si j'aurais vraiment aimé que l'on se croise plus longtemps au laboratoire. Nos apéros renversants du vendredi vont beaucoup me manquer et j'aurais voulu en profiter un peu plus.

Je tiens aussi à remercier les doctorants du laboratoire, avec qui j'ai partagé des expériences, des scripts, et beaucoup de moments forts. Merci Quentin et Léa pour m'avoir aidé à résoudre mes problèmes de virus insolubles et merci Jill, Jie et Massih pour m'avoir tenu compagnie pendant ces longues heures à chercher des échantillons disparus au -80 ou à percer des bulles dans des plaques Fluidigm en préPCR. Je remercie aussi Gabrielle, Alix, Clément, Guillaume, Eric et toute l'équipe du laboratoire pour m'avoir aidée dans mon projet et pour avoir rendu les heures de travail plus agréables au quotidien.

Et bien sûr merci à Anna-Line et Lisa, qui m'ont accompagnée jusqu'au bout pendant ces années importantes, en me soutenant dans mon projet de thèse mais aussi dans les activités de l'ADELIH : je savais que je pouvais toujours compter sur vous en cas de besoin ! Nous avons des caractères très différents et cela m'a beaucoup fait grandir de faire ce bout de chemin à vos côtés. Merci Lisa pour ta force et ton optimisme, et pour m'avoir entraînée dans tes folles aventures au ski, à Dijon, ou à la caravane. Merci Anna-Line pour ta détermination, ton écoute sans faille et ton soutien sportif. A force de te répéter que tout était possible, j'ai (presque) fini par me convaincre moi-même. Je suis la petite dernière à devenir docteur alors une page se tourne pour nous, mais j'espère que vous ne serez jamais loin.

Ces quatre années ont été par moments très intenses et j'ai aussi reçu beaucoup de soutien de la part de ma famille et de mes amis. Je remercie évidemment mes parents pour leur disponibilité infinie et leur aide en toute circonstance, et merci à David, Marie et Olivier pour nos rendez-vous familiaux malgré les décalages horaires. Même si la recherche et la biologie étaient des mondes assez mystérieux pour vous, vous avez toujours été curieux de savoir où en était mon projet et cela a renforcé ma passion et ma motivation et m'a redonné de l'énergie pour avancer.

Je remercie également mes amis qui ont été très présents dans ma vie parisienne au quotidien et qui m'ont permis de trouver mon équilibre. Merci aux « gars », Marc, Gildas, Quentin, Thibault et Vincent, pour les soirées et les week-end ensemble pendant lesquels on ne s'ennuie jamais, je sais que je peux compter sur vous pour me remonter le moral en deux minutes. Merci Simon pour ta curiosité communicative, Maxime pour les grands débats rue de la fidélité, et bien sûr merci les princesses, Priscille, Lisa et Elodie, pour votre amitié si précieuse depuis plus de dix ans.

Il me reste plus qu'à remercier Sophie, parce que c'était toi.

Et Lucas, parce que tu es la réalité qui me permet de rêver.

« En vérité, le chemin importe peu, la volonté d'arriver suffit à tout. » - Albert Camus

TABLE DES MATIERES

ABREVIATIONS	15
LISTE DES FIGURES ET TABLES	17
INTRODUCTION	19
I. LE CARCINOME HEPATOCELLULAIRE	21
A. Généralités sur le carcinome hépatocellulaire	22
1) Epidémiologie et facteurs de risque.....	22
2) Fibrose, cirrhose et carcinogénèse hépatique	25
3) Surveillance et diagnostic.....	29
4) Traitements du CHC.....	30
B. Analyse génomique des tissus tumoraux et non-tumoraux	33
1) Prédispositions génétiques.....	34
2) Mutations somatiques et signatures mutationnelles	36
3) Aberrations et réarrangements chromosomiques.....	38
4) Clonalité et chronologie des altérations.....	41
C. Les voies de signalisation dérégulées	43
1) Maintenance des télomères	44
2) Contrôle du cycle cellulaire.....	46
3) La voie WNT/ β -caténine	49
4) Autres voies de signalisation importantes dans la carcinogénèse hépatique	51
D. Classification moléculaire des CHC.....	53
II. LE VIRUS DE L'HEPATITE B	55
A. Biologie moléculaire du virus	55
1) Structure et organisation du génome.....	56
2) Eléments structuraux et régulateurs	57
3) Transcrits et protéines viraux	58
4) Cycle de réplication	61
B. Biologie de l'infection.....	64
1) Epidémiologie.....	65
2) Histoire naturelle de l'infection.....	67

3) Prophylaxie et traitements actuels.....	71
4) Enjeux et nouvelles cibles thérapeutiques.....	74
C. Influence des facteurs viraux sur l'infection : variabilité du génome et co-infections	76
1) Variabilité génotypique	77
2) Variabilité phénotypique	79
3) Co-infections virales	82
III. LA CARCINOGENESE INDUITE PAR LE VHB	87
A. Oncovirus et carcinogénèse viro-induite.....	87
1) Généralités sur les oncovirus	88
2) Mécanismes de carcinogénèse viro-induits.....	90
3) Rôle de la protéine HBx.....	96
B. Intégrations du VHB et conséquences oncogéniques	100
1) Place de l'intégration dans l'histoire naturelle de l'infection.....	101
2) Caractéristiques moléculaires des intégrations	104
3) Conséquences oncogéniques	108
C. Spécificités des CHC liés au VHB	113
1) Bilan des facteurs de risque et scores prédictifs	114
2) Caractéristiques génomiques	119
3) Enjeux thérapeutiques : infections occultes et intégrations	122
CONTEXTE DU PROJET DE RECHERCHE ET OBJECTIFS	127
RESULTATS	129
I. Caractérisation des intégrations du Virus de l'Hépatite B dans les tissus hépatiques tumoraux et non-tumoraux	131
Article 1: Hepatitis B virus integrations promote local and distant oncogenic driver alterations in hepatocarcinogenesis	135
II. Identification d'un sous-groupe spécifique de CHC agressifs présentant des altérations des gènes CCNA2 et CCNE1	177
Article 2: Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress.	179

III. Histoire naturelle et potentiel oncogénique de l'adénovirus-associé 2 (AAV2).....	211
Article 3: Adeno-associated virus in the liver: natural history and consequences in tumor development.....	213
DISCUSSION	243
I. Rôle de l'intégration virale dans l'infection et la carcinogénèse	245
A. Intégration ou réplication virales : deux destins incompatibles ?	245
B. Les séquences intégrées du VHB face à la réponse immunitaire	248
C. Intégration et réplication du VHB par rapport à la carcinogénèse	251
II. Conséquences oncogéniques des intégrations et diversité des CHC liés au VHB.....	253
A. Intégrations du VHB et instabilité chromosomique	253
B. Caractérisation et spécificité des réarrangements.....	257
C. Activation de TERT par la séquence intégrée du VHB.....	260
III. Similitudes et spécificités du VHB parmi les oncovirus	264
A. Intégrations virales et carcinogénèse : VHB versus VPH	264
B. Hépatocarcinogénèse et virus : VHB, VHC et aussi VHD et AAV	267
CONCLUSIONS ET PERSPECTIVES	271
BIBLIOGRAPHIE	273

ABREVIATIONS

AA	Acide Aristolochique	CRISPR-Cas9	CRISPR associated protein 9
AAV2	Virus Adéno-Associé 2		
ADN	Acide DésoxyriboNucléique	CTNNB1	Catenin (Cadherin-Associated Protein) Beta 1
ADNccc	ADN circulaire clos de façon covalente	CURS	Core Upstream Regulatory Region
ADNdsl	ADN linéaire double-brin	DDB1	DNA Damage-Binding protein 1
ADNrc	ADN circulaire relâché	DHBV	Duck Hepatitis B Virus
AFB1	Aflatoxine B1	DNMT	DNA méthyltransférases
AFP	Alpha-FœtoProtéine	DR	Direct Repeat
AgHBc	Antigène HBc	EASL	European Association for the Study of the Liver
AgHBe	Antigène HBe	ECOG	Eastern Cooperative Oncology Group
AgHBs	Antigène HBs	ELF4	E74-like ETS transcription factor 4
AgHD	Antigène HD	Enh	Enhancer
AHC	Adénome Hépatocellulaire	ERO	Espèces Réactives de l'Oxygène
ALD	Maladie hépatique liée à l'alcool (Alcoholic Liver Disease)	FAA	Fraction of Aberrant Arms
ALT	ALanine AminoTransférases	FDA	Food and Drug Administration
APOBEC	Apolipoprotein B mRNA editing enzyme catalytic polypeptide-like	FGF	Fibroblast Growth Factor
AR	Androgen Receptor	FN1	Fibronectine 1
ARN	Acide RiboNucléique	GWAS	Etude d'association pangénomique (Genome-Wide Association Study)
ARNm	ARN messenger	HBSP	Hepatitis B Spliced Protein
ARNpg	ARN pré-génomique	HHV	Virus de l'Herpès Humain
ASO	Oligonucléotides Antisens	HSD17B13	17-beta-hydroxysteroid dehydrogenase 13
AST	ASpartate aminoTransférases	HSPG	Heparan Sulfate ProteoGlycans
BCLC	Barcelona Clinic Liver Cancer	HTLV-1	Virus du lymphome T humain de type 1
BSB	Breakage-Fusion-Bridge	IFN	Interféron
CCN	Cycline	IL	InterLeukine
CDK	Cyclin Dependent Kinase	IRF2	Interferon regulatory factor 2
CHC	Carcinome Hépatocellulaire	IRM	Imagerie par Résonance Magnétique
CIRC	Centre international de Recherche sur le Cancer	ISG	Interferon Stimulated Genes
CNA	Altération du Nombre de Copies (Copy Number Alteration)		
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats		

KC	Cellules de Kupffer	PD-1	Programmed cell Death 1
KMT2B	Lysine Methyltransferase 2B (aussi nommé MLL4)	PD-L1	Programmed cell Death Ligand 1
L-HBs	Large-HBs	PNPLA3	Patatin-like phospholipase domain-containing 3
LINE	Long Interspersed Nuclear Elements	PS	Performance Status
LSEC	Cellules endothéliales sinusoïdales	RB	Rétinoblastome
MCP	Microenvironnement Cell Population	RNA-Seq	RNA Sequencing
MEC	Matrice ExtraCellulaire	RS	Réarrangement Structural
M-HBs	Medium-HBs	S-HBs	Small-HBs
MLL4	Mixed-Lineage Leukemia 4 (aussi nommé KMT2B)	SIDA	Syndrome de l'Immunodéficience Acquise
MMEJ	Microhomology-Mediated End Joining	SINE	Short Interspersed Nuclear Elements
NA	Analogues nucleos(t)idiques	SNP	Polymorphisme nucléotidique (Single Nucleotide Polymorphism)
NAFLD	Stéatose hépatique non liée à l'alcool (Nonalcoholic Fatty Liver Disease)	STAT4	Signal Transducer And Activator Of Transcription 4
NAP	Polymères d'acides nucléiques	T	Tumeur/Tumoral
NASH	Stéatohépatite non liée à l'alcool (Nonalcoholic SteatoHepatitis)	TACE	Chimio-Embolisation TransArtérielle
NDBG	Nodule Dysplasique de Bas Grade	TCGA	The Cancer Genome Atlas
NDHG	Nodule Dysplasique de Haut Grade	TCR	T Cell Receptor
NF-κB	Nuclear Factor-kappa B	TDM	TomoDensitoMétrie
NHEJ	Non-Homologous End Joining	TERC	Telomerase RNA Component
NK	Natural Killer	TERT	Telomerase Reverse Transcriptase
NRE	Negative Regulatory Element	TGF	Transforming Growth Factor
NT	Non-Tumoral (Tissu)	TLR	Récepteurs Toll-Like
NTCP	Sodium-tauchlorate cotransporting polypeptide	TM6SF2	Transmembrane 6 superfamily 2
OBI	Infection B occulte	TNF	Tumor Necrosis Factor
OMS	Organisation Mondiale de la Santé	VEB	Virus Epstein-Barr
ONU	Organisation des Nations Unies	VEGF	Vascular Endothelial Growth Factor
ORF	Cadre de lecture ouvert (Open Reading Frame)	VEGFR2	Vascular Endothelial Growth Factor Receptor 2
PBC	Promoteur Basal du Core	VHB	Virus de l'Hépatite B
PCR	Polymerase Chain Reaction	VHC	Virus de l'Hépatite C
		VHD	Virus de l'Hépatite D
		VIH	Virus de l'Immunodéficience Humaine
		VPH	PapillomaVirus Humain

LISTE DES FIGURES ET TABLES

Figure 1 : Taux d'incidence du CHC dans le monde standardisés sur l'âge	23
Figure 2 : Histoire naturelle de la maladie hépatique chronique	26
Figure 3 : Mécanismes d'hépatocarcinogénèse sur foie cirrhotique (70-80%) et foie non cirrhotique (20-30%).....	27
Figure 4 : Algorithme clinique pour la gestion des CHC.....	31
Figure 5 : Prédipositions génétiques et processus mutationnels	36
Figure 6 : Les dix signatures mutationnelles observées dans les CHC	38
Figure 7 : Les altérations du « tronc » impliquées dans l'hépatocarcinogénèse.....	41
Figure 8 : CHC et voies de signalisation dérégulées	43
Figure 9 : Mécanismes d'activation de la télomérase via une stimulation de l'expression du gène TERT dans le CHC.....	45
Figure 10 : Fréquence des altérations du gène CTNNB1 impliqués dans l'hépatocarcinogénèse et contexte génomique.....	50
Figure 11 : Classification moléculaire et pathologique des CHC	54
Figure 12 : Structure moléculaire du VHB.....	56
Figure 13 : Description des 5 transcrits et des 7 protéines virales du VHB	59
Figure 14 : Cycle de réplication du VHB	61
Figure 15 : Etape de rétro-transcription de l'ARNpg pour former l'ADNrc (90%) ou l'ADNdsi (10%).....	63
Figure 16 : Epidémiologie du VHB dans le monde	65
Figure 17 : Mortalité liée au VHB dans les six régions définies par l'OMS	66
Figure 18 : Histoire naturelle de l'infection par le VHB.....	67
Figure 19 : Les 5 phases de l'infection B chronique	69
Figure 20 : Cibles thérapeutiques actuelles et futures du VHB.....	74
Figure 21 : Différences entre les séquences nucléotidiques des différents génotypes du VHB.....	78
Figure 22 : Répartition des génotypes du VHB dans le monde	79
Figure 23 : Structure des virions du VHB et du VHD	84
Figure 24 : Les différents mécanismes impliqués dans la carcinogénèse liée au VHB	95

Figure 25 : Interactions de la protéine HBx et conséquences sur la réplication virale et sur l'hépto-carcinogénèse	96
Figure 26 : Modèle du processus d'intégration de l'ADN du VHB au cours de l'hépatite chronique	103
Figure 27 : Position des ORF dans la forme ADNdsI du VHB.....	108
Figure 28 : Représentation schématique des différents sous-groupes de CHC liés au VHB définis par analyse transcriptomique	122
Figure 29 : Niveau d'expression des transcrits HBx en 5' (gauche) ou 3' (droite) dans 157 tissus non-tumoraux de notre série analysée en Capture.	247
Figure 30 : Mécanismes de suppression des réponses immunitaires innées et adaptatives par les AgHBs.....	249
Figure 31 : Niveau d'expression des transcrits préS1 dans 157 tissus non-tumoraux de notre série analysée en Capture	249
Figure 32 : Prédiction des transcrits et analyse par western-blot du gène <i>CCNA2</i> dans le CHC #3798T (intégration du VHB dans l'intron 2).	256
Figure 33 : Exemples de réarrangements chromosomiques complexes : cycle BSB (gauche) et iso-chromosome (droite).	259
Figure 34 : Fréquence des gains et délétions dans 35 CHC.....	259
Figure 35 : Associations entre les mécanismes d'altérations de TERT et les caractéristiques des patients et des tumeurs	262
Figure 36 : Profil génomique et schéma du mécanisme d'amplification en « rolling circle » lié à l'intégration d'ADN du VPH	266
Figure 37 : Profil des réarrangements structuraux internes aux génomes du VHB présents dans la tumeur #1987T	267
Table 1 : Principaux variants de résistance du VHB face aux antiviraux actuels	82
Table 2 : Caractéristiques des six virus classés carcinogéniques pour l'humain par le CIRC.	89
Table 3 : Etudes d'intégrations du VHB à partir de séquençage haut-débit	106
Table 4 : Liste des facteurs de risque associés au développement d'un CHC chez les patients infectés par le VHB.....	114

INTRODUCTION

Le foie est un organe essentiel du corps humain qui remplit plus de 300 fonctions vitales. C'est le deuxième organe le plus actif après le cerveau et il a la particularité de pouvoir se régénérer rapidement pour retrouver toutes ses capacités lorsqu'il est réduit à une fraction de sa taille. Parmi ses fonctions, le foie stocke les vitamines, le sucre et le fer, contrôle la production et l'élimination du cholestérol, filtre les déchets et substances toxiques du sang, synthétise des facteurs de coagulation et des protéines de la réponse immunitaire, et produit la bile qui permet la digestion et l'absorption des nutriments. De par son rôle de filtre, le foie est en contact avec de nombreux agents pathogènes et les cellules hépatiques peuvent notamment être infectées par le virus de l'Hépatite B. Cette infection peut affecter considérablement la fonctionnalité du foie et contribuer au développement d'un cancer hépatique, le carcinome hépatocellulaire. L'objectif de cette introduction est de présenter d'un côté le carcinome hépatocellulaire et de l'autre le virus de l'Hépatite B afin de mettre en évidence dans une troisième partie comment l'un peut être la conséquence de l'autre.

I. LE CARCINOME HEPATOCELLULAIRE

Le cancer primitif du foie est à l'heure actuelle le sixième cancer en termes d'incidence (841 000 nouveaux cas par an) et le quatrième cancer en termes de mortalité (782 000 décès par an) dans le monde. Entre 1990 et 2015, l'incidence du cancer du foie a augmenté de 75% et les projections effectuées par le Centre International de Recherche sur le Cancer (CIRC) estiment que cette augmentation sera de 64% entre 2018 et 2040 (Bray et al., 2018; Global Burden of Disease Liver Cancer Collaboration et al., 2017). Avec un ratio incidence sur mortalité proche de 1 et une survie moyenne à 5 ans de 18%, le cancer du foie est le deuxième cancer le plus létal après le cancer du pancréas (Jemal et al., 2017).

Les cancers primitifs, à savoir principalement les carcinomes hépatocellulaires (75-85%) se développant à partir des hépatocytes et les cholangiocarcinomes (10-15%) qui sont issus des cellules épithéliales des canaux biliaires, sont à distinguer des cancers métastatiques qui sont beaucoup plus fréquents mais qui sont des événements secondaires de tumeurs originaires d'autres organes (intestin, peau, sein, prostate, etc...).

A. Généralités sur le carcinome hépatocellulaire

Le carcinome hépatocellulaire (CHC) est la tumeur maligne primitive du foie la plus fréquente. A lui seul, il représente 9,2% des nouveaux cas de cancers par an dans le monde. Cependant, les taux d'incidence et de mortalité du CHC sont très hétérogènes en fonction des régions, et la répartition des cas évolue rapidement à l'heure actuelle avec la croissance démographique mondiale, le vieillissement de la population dans les pays développés et l'augmentation ou la diminution de certains facteurs de risque. Dans l'ensemble, le nombre de CHC va néanmoins continuer à augmenter pendant les prochaines décennies, et étant donné qu'une proportion importante est détectée à des stades trop tardifs pour bénéficier d'un traitement curatif, c'est un enjeu de santé publique majeur pour beaucoup de pays (Sagnelli et al., 2020; Singal et al., 2020).

1) Epidémiologie et facteurs de risque

L'incidence du carcinome hépatocellulaire dans les différents pays dépend de la prévalence locale des facteurs de risque qui y sont associés. La répartition mondiale en 2018 était la suivante : 72% des CHC ont été diagnostiqués en Asie (et plus de 50% en Chine), 10% en Europe, 7,8% en Afrique, 5,1% en Amérique du Nord, 4,6% en Amérique latine et Caraïbes et 0,5% en Océanie (Singal et al., 2020). L'incidence standardisée sur l'âge (Figure 1) a augmenté de plus de 100% dans beaucoup de pays développés en Europe et Amérique du Nord, alors qu'à l'inverse, elle a diminué de plus de 20% en Chine et dans les pays d'Afrique sub-saharienne.

L'âge moyen du diagnostic et la survie moyenne varient aussi considérablement en fonction des régions. Le CHC se développe tardivement chez les patients originaires d'Europe, d'Amérique du Nord ou du Japon, chez qui le diagnostic a lieu généralement après 60 ans. La survie moyenne y est respectivement de 24, 33 et 60 mois. En revanche, dans la plupart des pays asiatiques et africains, les CHC sont diagnostiqués chez des patients ayant entre 30 et 60 ans. Les CHC liés au virus de l'Hépatite B en Afrique surviennent même fréquemment avant 40 ans. Et la survie moyenne chez ces patients africains ne dépasse pas 3 mois, à cause de l'absence de programmes de surveillance et de la difficulté d'accès aux traitements (J. D. Yang et al., 2019).

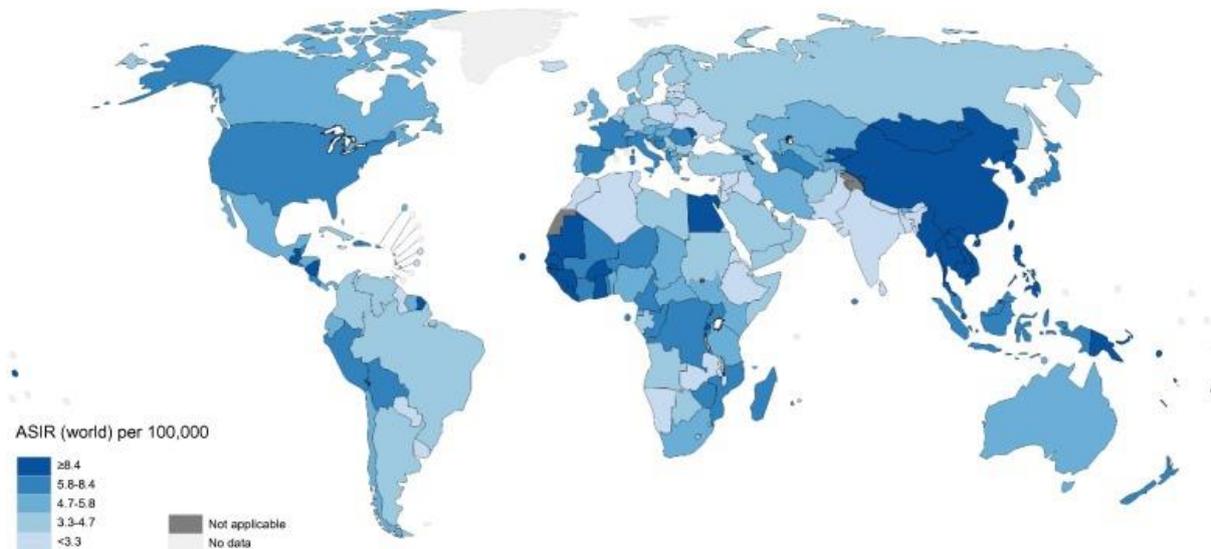


Figure 1 : Taux d'incidence du CHC dans le monde standardisés sur l'âge.
Source des données : GLOBOCAN 2018 (<http://gco.iarc.fr/today>) (Singal et al., 2020)

Globalement, le carcinome hépatocellulaire est 2 à 2,5 fois plus fréquent chez les hommes que chez les femmes. Cela peut s'expliquer par un regroupement de facteurs de risque parmi les hommes, et potentiellement par une fonction des androgènes qui favoriserait la cancérogénèse ou par un rôle protecteur des œstrogènes. Par ailleurs, il est rare qu'un CHC se développe sans maladie hépatique sous-jacente. En effet, les patients cirrhotiques présentent un risque de 2 à 4% par an de développer un CHC, ce risque pouvant varier en fonction de la cause et du degré des dommages hépatiques, de la région géographique, du sexe et de l'âge du patient. Il est estimé qu'un tiers des patients ayant une cirrhose vont développer un jour un cancer du foie (Singal et al., 2020; Villanueva, 2019).

A l'échelle mondiale, c'est le Virus de l'Hépatite B (VHB) qui est la première cause du CHC, suivi de l'Aflatoxine B1, du Virus de l'Hépatite C (VHC), de la consommation d'alcool et de la stéatose hépatique non liée à l'alcool (NAFLD – Nonalcoholic Fatty Liver Disease). Il y a par ailleurs un effet incrémental lié à la présence de plusieurs facteurs, dans le cas par exemple de patients avec une co-infection VHB/VHC ou de patients infectés par le VHC ayant une forte consommation d'alcool (Sagnelli et al., 2020). Les facteurs de risque sont détaillés ci-dessous (J. D. Yang et al., 2019) :

- Les hépatites virales : les infections chroniques par le VHB et le VHC sont à l'origine de 85% des cas de CHC dans le monde. Le VHB est majoritairement présent en Asie de l'Est et en Afrique sub-saharienne, alors que le VHC est fréquemment

retrouvé en Afrique du Nord, au Moyen-Orient et en Asie centrale, et également au Japon, en Europe et en Amérique du Nord. L'infection chronique du foie par un de ces deux virus entraîne une cirrhose chez 10 à 20% des patients, et l'incidence du CHC chez ces patients cirrhotiques est de 2 à 5% par an. Une particularité importante du VHB est qu'il peut entraîner la survenue d'un CHC en absence de cirrhose. Ces CHC sur foie sain sont globalement minoritaires : ils représentent 30 à 50% des CHC dans les régions où le VHB est endémique (Asie de l'Est et Afrique), et moins de 10% dans des pays où le virus est moins présent comme aux Etats-Unis. Grâce à la mise en place récente de programmes de vaccination contre le VHB, et le développement de nouvelles thérapies antivirales efficaces contre le VHC, le nombre de CHC liés aux hépatites virales va probablement diminuer dans les prochaines années.

- La maladie hépatique liée à l'alcool (ALD) : la cirrhose alcoolique est le deuxième facteur de risque associé au CHC en Europe et aux Etats-Unis (après la cirrhose liée au VHC), et est à l'origine de 30% des CHC dans le monde. L'effet carcinogène de l'alcool repose sur la formation d'acétaldéhyde qui a un effet toxique direct sur les cellules, et d'espèces réactives à l'oxygène (ERO) générant du stress oxydatif qui augmente les dommages cellulaires. La consommation excessive d'alcool induit ainsi une inflammation chronique du foie. Le risque de survenue d'un CHC sur cirrhose alcoolique est globalement plus faible que sur cirrhose issue d'une infection virale chronique mais il peut atteindre 2,9% dans certaines populations. Enfin, le diagnostic étant généralement réalisé plus tardivement chez ces patients, les CHC liés à l'alcool ont souvent un pronostic plus défavorable (Ganne-Carrié and Nahon, 2019).

- La stéatose hépatique non liée à l'alcool (NAFLD) : la NAFLD est devenue la maladie hépatique la plus courante dans les pays développés et un facteur de risque majeur pour le CHC. Sa prévalence est notamment de 30% aux Etats-Unis, où 10 à 20% des cas de CHC sont attribués à une NAFLD. Cette stéatose non provoquée par l'alcool est liée le plus souvent à des syndromes métaboliques tels que l'obésité ou le diabète de type II. Elle peut induire le développement d'une stéatohépatite non liée à l'alcool (NASH – Nonalcoholic SteatoHepatitis), une inflammation chronique du foie pouvant aller jusqu'à la cirrhose. L'incidence du CHC chez les patients ayant une NASH est également inférieure à celle liée aux infections virales et se situe entre 1 et 2% par an. Cependant, des études récentes ont montré qu'un quart des CHC induits par les NAFLD se développent sur foie non-cirrhotique. Il n'y a encore que peu

d'information sur le risque de CHC lié aux NAFLD mais le nombre patients étant en forte augmentation, ce facteur de risque va devenir de plus en plus important dans les années à venir (Kulik and El-Serag, 2019; Singal et al., 2020).

- Aflatoxine B1 et acide aristolochique : l'Aflatoxine B1 (AFB1) est une toxine carcinogène produite par un champignon parasite nommé *Aspergillus Flavus* et qui peut être présente dans les céréales et les arachides mal conservés. 90% de la population d'Afrique de l'Ouest sub-saharienne est exposée à cette toxine alors qu'elle est quasiment inexistante dans les pays occidentaux. L'acide aristolochique (AA) est un composé hautement mutagène contenu dans les plantes nommées *Aristolochia* ou *Asarum*. Ces plantes ont été utilisées dans la médecine chinoise traditionnelle pendant plusieurs siècles. Des études récentes de séquençage ont permis de montrer que les CHC de patients exposés à l'AFB1 ou à l'AA présentent une signature mutationnelle caractéristique de l'exposition. Ces patients provenant majoritairement de régions où le VHB est endémique, il existe potentiellement une interaction entre l'infection virale et l'exposition à ces agents mutagènes, qui sera développée dans la partie III.C.1.

- Virus adéno-associé de type 2 (AAV2) : le virus AAV2 a été ajouté récemment à ces facteurs de risque. Le laboratoire a en effet identifié que 2% des CHC ont pour origine une mutagénèse insertionnelle d'AAV2 à proximité d'oncogènes, dont la surexpression de ces gènes peut jouer un rôle dans la carcinogénèse. Il s'agit de CHC sur foie sain et sans autre étiologie (La Bella et al., 2020; Nault et al., 2015).

Enfin, d'autres maladies telles que l'hémochromatose, les maladies biliaires ou les hépatites auto-immunes peuvent induire une cirrhose et un CHC. Les patients ayant une cirrhose due à l'hémochromatose (accumulation de fer dans les hépatocytes) ont notamment un risque très élevé de développer CHC (dans 45% des cas). Le tabac semble aussi être un facteur favorisant la survenue d'un CHC (Singal et al., 2020).

2) Fibrose, cirrhose et carcinogénèse hépatique

La persistance des dommages hépatiques provoqués par la plupart des étiologies présentées dans le paragraphe précédent déclenche la mise en place dans le foie d'un

processus de cicatrisation avec inflammation chronique constitué de cycles continus de nécrose et régénération des hépatocytes (Farazi and DePinho, 2006).

Fibrose et cirrhose (Figure 2)

La fibrose hépatique est une conséquence histologique de ce processus de cicatrisation des lésions hépatiques, qui est caractérisée par l'accumulation de matrice extracellulaire (MEC). Ces changements sont réversibles si les dommages générés au foie sont transitoires, et l'architecture hépatique retrouve alors sa composition normale. Cependant, si les dommages persistent, l'inflammation devient chronique et l'accumulation de MEC persiste, entraînant une substitution progressive du parenchyme hépatique par du tissu cicatriciel (Hernandez-Gea and Friedman, 2011).

La cirrhose hépatique correspond au stade le plus avancé de la fibrose qui devient alors souvent irréversible. Elle se caractérise par la formation de nodules régénératifs séparés par du tissu fibreux, qui entraîne une dégradation de la fonction hépatique, des modifications vasculaires et une hypertension portale. En général, la progression vers ce stade final est lente. Chez des patients ayant une atteinte hépatique chronique, la maladie hépatique peut évoluer pendant 20 à 40 ans en fonction des facteurs génétiques et environnementaux (Hernandez-Gea and Friedman, 2011; Pellicoro et al., 2014; Schuppan and Afdhal, 2008).

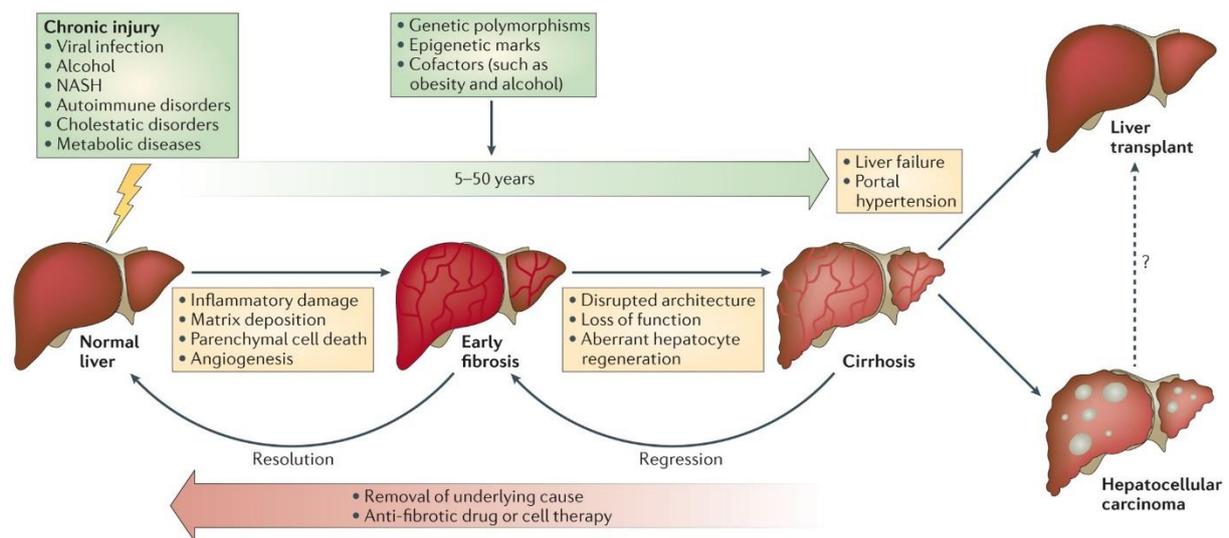


Figure 2 : Histoire naturelle de la maladie hépatique chronique
(Pellicoro et al., 2014)

Les patients demeurent souvent asymptomatiques jusqu'au développement de la cirrhose, les symptômes étant légers tant que le foie peut encore accomplir ses fonctions principales. En revanche, lorsque la cirrhose est dite « décompensée », des complications indiquant l'épuisement du foie apparaissent (ascites, jaunisse, varices œsophagiennes). La survie moyenne des patients avec une cirrhose compensée et décompensée est supérieure à 12 ans et inférieure à 2 ans respectivement. Le seul traitement pour la cirrhose est la transplantation hépatique (Muir, 2015).

Le stade de la maladie hépatique peut être exprimé avec un score semi-quantitatif appelé score METAVIR. Le niveau F0 correspond à une absence de fibrose, les niveaux F1 à F3 à une fibrose légère à sévère et le niveau F4 caractérise le stade final qu'est la cirrhose. Le score de Child-Pugh est également utilisé pour caractériser la sévérité des atteintes hépatiques et estimer le pronostic associé. Une cirrhose de grade A est associée à un meilleur pronostic qu'une cirrhose de grade B, le grade C étant attribué aux cirrhoses les plus sévères avec le moins bon pronostic (Muir, 2015).

Développement du CHC :

Au cours du développement d'un CHC, plusieurs mécanismes sont mis en jeu (Figure 3). Schématiquement, on doit distinguer les CHC survenant sur un foie cirrhotique des CHC se développant sur foie sain, les mécanismes de carcinogénèse étant différents.

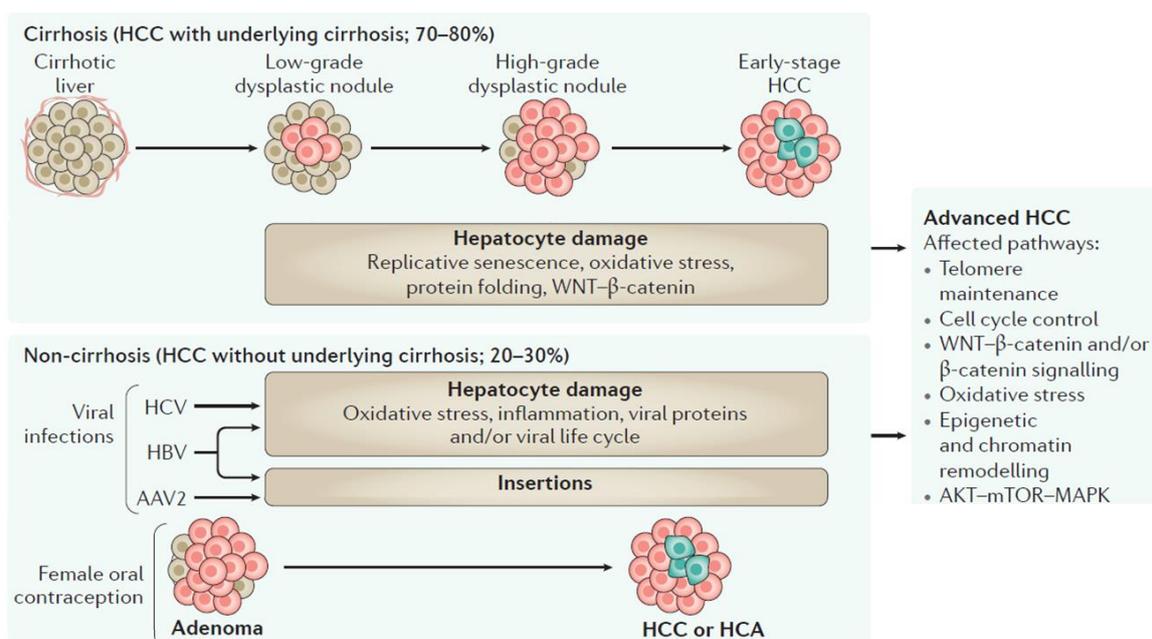


Figure 3 : Mécanismes d'hépatocarcinogénèse sur foie cirrhotique (70-80%) et foie non cirrhotique (20-30%) (Adaptée de Llovet et al., 2016).

- CHC sur foie cirrhotique :

Sur foie cirrhotique, la carcinogénèse hépatique est un processus en plusieurs étapes et consiste en l'apparition séquentielle des lésions tumorales suivantes : nodule dysplasique de bas grade (NDBG), nodule dysplasique de haut grade (NDHG), CHC précoce et CHC avancé. Les nodules dysplasiques sont définis comme des lésions précancéreuses qui mesurent entre 1 mm et 2 cm et qui ont des caractéristiques régénératrices. Un apport en sang artériel est généralement observé dans les NDHG, puis une invasion stromale dans les CHC précoces et enfin une invasion veineuse ainsi que de possibles métastases dans les CHC avancés (Di Tommaso et al., 2013).

Les caractéristiques histologiques principales associées à la transformation maligne des cellules hépatiques sont une augmentation de la densité cellulaire et du ratio noyau/cytoplasme. Les CHC sont souvent décrits en fonction de leur architecture (trabéculaire, pseudo-glandulaire, compacte) et de leur aspect cytologique. Suivant ces facteurs, ils peuvent également être classés selon leur niveau de différenciation en 4 grades selon la classification Edmondson and Steiner (numérotées de I à IV, I étant la classe des CHC les plus différenciés et IV celle des CHC les moins différenciés) ou en 3 catégories selon la classification de l'Organisation Mondiale de la Santé (les CHC bien différenciés, moyennement différenciés et peu différenciés) (Calderaro et al., 2019).

- CHC sur foie sain :

Dans 10% à 50% des cas en fonction des régions, les CHC se développent sur non-cirrhotique, voire sur foie sain, ils sont alors initiés par un événement oncogénique fort. L'infection virale par le VHB est notamment connue pour avoir la capacité d'entraîner une carcinogénèse sur foie sain. Le laboratoire a montré récemment que c'est aussi le cas de l'infection par AAV2 (La Bella et al., 2020; Rebouissou and Nault, 2020). Les mécanismes sous-jacents seront développés plus en détails dans la partie III.A.

Par ailleurs, des études ont montré que certains CHC diagnostiqués chez des patients ayant un syndrome métabolique se sont développés en l'absence de fibrose et de cirrhose suggérant un mécanisme de carcinogénèse direct comme pour l'infection par le VHB. Mais une proportion significative des CHC développés sur foie sain résulterait de la transformation maligne d'un adénome du foie (Margini and Dufour, 2016; Paradis et al., 2009).

En effet, la survenue d'un CHC sur foie sain peut également être due à la transformation maligne d'un adénome hépatocellulaire (AHC). Les AHC se développent principalement chez des femmes jeunes prenant une contraception orale et sans aucune maladie hépatique. La transformation d'un AHC en CHC est un événement rare qui n'est observée que pour 5% des AHC. Cependant, les AHC peuvent être classés en différents sous-groupes moléculaires et le risque de complications diffère fortement entre ces sous-groupes. 10 à 15% des AHC avec une mutation activatrice du gène *CTNNB1* dans l'exon 3 ont notamment un risque élevé de transformation maligne (Rebouissou and Nault, 2020).

3) Surveillance et diagnostic

L'objectif des programmes de surveillance est de pouvoir détecter les CHC à un stade précoce afin d'augmenter les chances de pouvoir administrer au patient un traitement curatif. Les personnes à risque pour lesquels une surveillance est recommandée sont les patients ayant une cirrhose légère ou modérée (Score de Child-Pugh A ou B), et les patients ayant une cirrhose sévère (Score de Child-Pugh C) en attente d'une transplantation hépatique. En cas d'absence de cirrhose, une surveillance est néanmoins suggérée pour les patients avec une infection du VHB rassemblant d'autres critères de risque (sexe masculin, âge élevé, forte réplication virale, fibrose hépatique...) ainsi que pour les patients avec une infection au VHC et une fibrose avancée (METAVIR F3). Les recommandations classiques indiquent qu'une surveillance doit être réalisée tous les 6 mois chez ces patients. Cet intervalle de temps correspond en effet à la période de doublement d'un nodule (Galle et al., 2018; Llovet et al., 2016; Singal et al., 2020; Villanueva, 2019).

Les tests de surveillance sont principalement réalisés par échographie, une technique d'imagerie à faible coût et non-invasive. La spécificité de cette technique est supérieure à 90% et sa sensibilité est de 84%. Cependant, la sensibilité pour détecter les CHC précoces n'est que de 47%, et l'échographie semble moins efficace chez certains patients atteints d'obésité ou de maladie hépatique non virale. C'est pourquoi d'autres techniques d'imagerie telles que la Tomodensitométrie (TDM) ou l'Imagerie par Résonance Magnétique (IRM) sont de plus en plus utilisées en clinique (Singal

et al., 2020; Villanueva, 2019). La mesure du niveau d'Alpha-FœtoProtéine (AFP) dans le sang est également couramment effectuée bien que la sensibilité de ce test pour détecter un CHC soit faible. Elle augmente cependant la sensibilité de l'échographie lorsque les deux techniques sont combinées (J. D. Yang et al., 2019).

Chez les patients cirrhotiques, le diagnostic peut être directement effectué à partir de l'imagerie si la lésion a une taille supérieure à 1 cm. En présence de nodules plus petits, le diagnostic s'avère plus compliqué et il est recommandé de surveiller par échographie tous les 3 à 4 mois. Si l'imagerie ne permet pas de conclure ou si le patient n'est pas cirrhotique, une biopsie est indispensable pour effectuer le diagnostic (Galle et al., 2018; Villanueva, 2019).

L'évaluation du stade de la carcinogénèse repose alors généralement sur l'algorithme de décision BCLC (Barcelona Clinic Liver Cancer). Cet algorithme classe les patients en 5 groupes distincts, avec des recommandations de traitements spécifiques à chaque stade (Figure 4). Il tient compte des caractéristiques tumorales données par l'imagerie (nombre et taille des nodules, envahissement vasculaire, propagation extra-hépatique), de la sévérité de la maladie hépatique indiquant si le foie a toujours la capacité de remplir ses fonctions (score Child-Pugh) et de l'état général de santé du patient (basé sur un score nommé Performance Status). Ce dernier score est une classification mise au point par l'ECOG (Eastern Cooperative Oncology Group) allant de 0 à 5 où le 0 représente un bon état de santé et 5 le décès (Villanueva, 2019).

4) Traitements du CHC

La classification BCLC est donc l'outil utilisé dans la plupart des régions du monde pour déterminer le traitement approprié pour chaque patient. A l'heure actuelle, il existe 4 types de traitements : les traitements chirurgicaux (résection ou transplantation), l'ablation tumorale (radiofréquence percutanée), les traitements transartériels (chimio-embolisation) et les thérapies systémiques. Seuls les traitements chirurgicaux et l'ablation peuvent être curatifs et ils ne sont possibles que pour des CHC diagnostiqués à un stade précoce (BCLC 0 ou A). Aucun traitement n'existe pour les CHC diagnostiqués à un stade terminal (BCLC D).

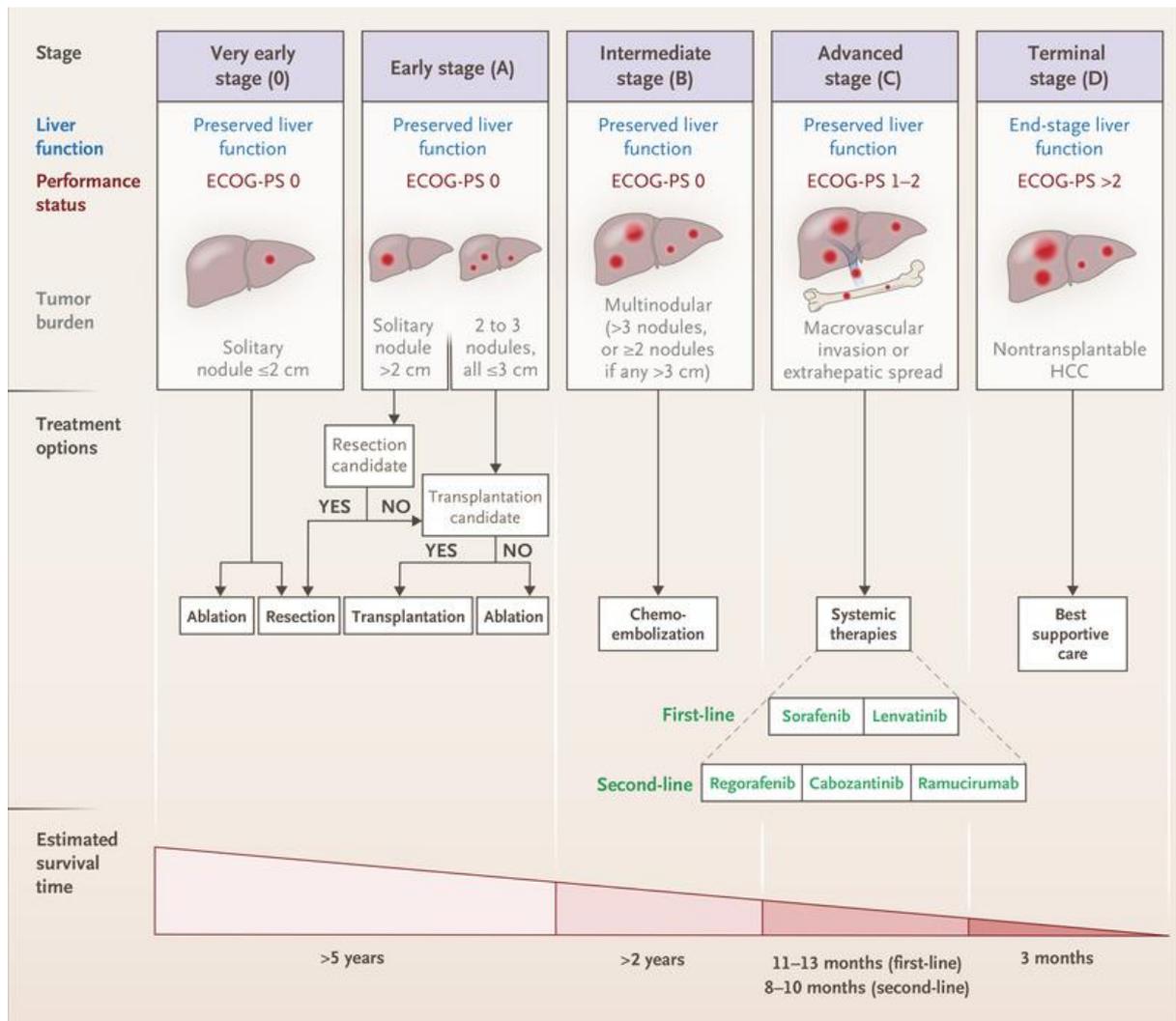


Figure 4 : Algorithme clinique pour la gestion des CHC (Villanueva, 2019).

Les différents traitements sont détaillés dans les paragraphes suivants (Villanueva, 2019; J. D. Yang et al., 2019) :

- Les traitements chirurgicaux :

Ils sont recommandés pour les patients dont le CHC a été diagnostiqué à un stade précoce (BCLC 0 ou A). La résection est associée à une survie moyenne à 5 ans supérieure à 60%, c'est un traitement qui est potentiellement curatif mais une récurrence survient néanmoins dans les 5 ans chez plus de 70% de ces patients. C'est le traitement le plus utilisé pour les CHC développés sur foie non-cirrhotique.

La transplantation est réalisée chez des patients qui correspondent au critère de Milan (nodule unique de moins de 5 cm, ou un à trois nodules dont aucun ne dépasse 3 cm).

Le taux de survie chez ces patients après transplantation est de 60 à 80% après 5 ans et de 50% après 10 ans, et le taux de récurrence est inférieur à 15%. La limite principale de la transplantation reste la différence importante qui existe entre le nombre de donneurs et le nombre de patients en attente de transplantation.

- L'ablation tumorale :

L'ablation est recommandée chez les patients dont le CHC a également été diagnostiqué à un stade précoce (BCLC 0 ou A) mais qui ne peuvent pas bénéficier d'une chirurgie à cause de comorbidités ou d'une dysfonction hépatique. La méthode la plus utilisée est l'ablation par radiofréquence percutanée qui consiste à augmenter fortement la température intra-tumorale pour induire la nécrose des cellules. L'ablation entraîne moins de complications que la résection mais permet un moins bon contrôle à l'échelle locale pour les tumeurs de grande taille. L'ablation peut par ailleurs être utilisée comme traitement néo-adjuvant pour éviter une progression tumorale chez des patients en attente de transplantation ou pour réduire la taille des nodules chez des patients qui pourraient ainsi devenir éligibles à la transplantation (Lee et al., 2017).

- Les traitements transartériels :

Ils sont recommandés pour les patients avec un CHC à un stade intermédiaire (BCLC B). La méthode la plus utilisée est la chimio-embolisation transartérielle (TACE), qui consiste en une perfusion intra-artérielle d'un agent cytotoxique, immédiatement suivie d'une embolisation des vaisseaux alimentant la tumeur, afin de provoquer une nécrose ischémique. La survie moyenne après TACE varie entre 26 et 40 mois, en fonction de l'étendue de la tumeur et de la sévérité des atteintes de la fonction hépatique. Ce traitement peut également être administré aux patients en attente de transplantation pour éviter une progression tumorale.

- Les thérapies systémiques :

Les thérapies systémiques sont recommandées en cas de CHC avancé (BCLC C). Le sorafenib a été la première molécule adoptée en 2008 par la FDA (Food and Drug Administration) pour le traitement du CHC. Elle a été longtemps la seule molécule utilisée en première ligne de traitement, jusqu'au développement du lenvatinib en 2018. Ces deux molécules sont des inhibiteurs multi-kinase (Kudo et al., 2018; Llovet et al., 2008). La plupart des patients ayant néanmoins montré une progression de leur

CHC malgré le traitement au sorafenib, des molécules ont été approuvées pour une seconde ligne de traitement. C'est le cas du regorafenib et du cabozantinib qui sont également des inhibiteurs multi-kinase et du ramucirumab qui est un anticorps ciblant le récepteur 2 du facteur de croissance de l'endothélium vasculaire (VEGFR2) (Abou-Alfa et al., 2018; Bruix et al., 2017; A. X. Zhu et al., 2019). L'ensemble de ces thérapies de première ou deuxième ligne permet une augmentation de la survie moyenne des patients de quelques mois avec de faibles taux de réponses.

Des immunothérapies utilisant des inhibiteurs de PD-1 (Programmed cell Death 1), le pembrolizumab et le nivolumab, ont été testées dans des essais cliniques de phase II et ont montré des augmentations de la survie moyenne des patients encourageantes mais des taux de réponse partiel (17% et 20% respectivement). Cependant, les résultats récents des essais de phase III indiquent que les objectifs prédéfinis n'ont pas été atteints (El-Khoueiry et al., 2017; Zhu et al., 2018). Récemment un traitement combinant le bevacizumab (antiangiogénique anti-VEGF) associé à l'atezolizumab (immunomodulateur anti-PD-L1) dans un essai de phase 3 a démontré sa supériorité en comparaison avec le sorafenib en première ligne (Finn et al., 2020; von Felden, 2020). Ainsi, l'association bevacizumab/atezolizumab est maintenant devenue le traitement de première intention des CHC avancés en première ligne malgré le risque hémorragique significatif lié à l'usage d'anti-VEGF.

B. Analyse génomique des tissus tumoraux et non-tumoraux

Le développement des techniques de séquençage a engendré des avancées majeures dans la connaissance de la génomique liée au CHC, rendant possible une description moléculaire précise des tissus tumoraux et non-tumoraux.

D'un côté, la caractérisation génomique de tissus hépatiques non-tumoraux ont mis en évidence l'existence de polymorphismes associés à un risque plus élevé de développer un CHC (Nahon and Zucman-Rossi, 2012; Rebouissou and Nault, 2020). En parallèle, la comparaison du génome tumoral avec le génome du tissu hépatique non-tumoral adjacent a permis d'identifier des altérations génétiques présentes uniquement dans les tumeurs et appelées altérations somatiques. Il peut s'agir de mutations (substitution d'un nucléotide, insertions ou délétions de quelques

nucléotides) ou d'altérations chromosomiques (délétions homozygotes, amplifications focales, pertes ou gains de chromosomes et translocations) (Schulze et al., 2016). Des intégrations virales ont également pu être identifiées et jouent un rôle important dans la carcinogénèse hépatique, mais ce type d'altérations étant majoritairement lié au Virus de l'Hépatite B, il ne sera pas abordé dans cette partie mais dans la partie III.

1) Prédispositions génétiques

Les SNP (Single Nucleotide Polymorphism) sont des variants génétiques présents dans la population générale qui peuvent influencer l'occurrence, la sévérité et les caractéristiques cliniques des maladies. Ils sont généralement identifiés dans plus de 5% de la population et ils n'apportent qu'une prédisposition modérée au développement de la maladie considérée (Rebouissou and Nault, 2020).

De nombreuses études ont montré des associations entre différentes SNPs et la survenue d'un CHC. La plupart d'entre elles testent l'hypothèse qu'un variant connu soit représenté en plus grande proportion chez des patients ayant déjà développé un CHC (études cas-témoins) ou soit associé à un risque plus élevé de développer ce cancer chez des patients cirrhotiques (études prospectives). Malheureusement, la majorité de ces études sont impactées par des biais méthodologiques majeurs dus à l'hétérogénéité des populations considérées. Les résultats obtenus dans ces séries ont donc rarement pu être reproduits dans d'autres publications, suggérant l'existence de nombreux résultats faussement positifs (Nahon and Zucman-Rossi, 2012).

Parmi les SNPs identifiés (Figure 5), un certain nombre d'entre eux altèrent des voies de signalisation biologiques impliquées dans la carcinogénèse, telles que l'inflammation ($TNF-\alpha$, $IL-1\beta$, $IL-10$ ou $TGF-\beta$), le stress oxydatif ($GSTM1$, $SOD2$, MPO), le métabolisme du fer ($HFE1$) ou les mécanismes de réparation de l'ADN ($MTHFR$, $MDM2$, $TP53$). Ces SNPs impacteraient le risque de survenue un CHC en influant directement la transformation maligne et la progression tumorale ou en prédisposant en amont à des facteurs de risque spécifiques, au développement d'une maladie hépatique chronique ou à son évolution vers la cirrhose. Une validation de ces associations reste néanmoins nécessaire chez des patients aux étiologies diverses et avec différents degrés d'atteintes hépatiques (Zucman-Rossi et al., 2015).

Une autre technique a été mise au point plus récemment pour identifier des prédispositions génétiques : les études d'association pangénomique GWAS (Genome-Wide Association Study). Elles permettent de comparer la distribution génotypique de milliers de SNP entre des patients atteints de CHC et une population témoin, sans hypothèse préalable. Neuf SNPs ont ainsi été identifiés comme associés à un risque élevé de survenue d'un CHC dans des populations spécifiques. Il s'agit de variants des gènes *DDX18* (patients cirrhotiques), *DEPDC5* et *MICA* (patients infectés par le VHC), *KIF1B*, *GRIK1*, *STAT4* et *DLC1* (patients infectés par le VHB), et *PNPLA3* et *TM6SF2* (patients avec une cirrhose alcoolique). Une étude du laboratoire a testé ces associations dans une population européenne avec différents degrés de maladie hépatique due à différentes étiologies et n'a pu valider que 3 SNPs parmi cette liste : le SNP *STAT4* rs7574865 associé au CHC chez les patients infectés par le VHB et les SNPs *PNPLA3* rs738409 et *TM6SF2* rs58542926 associés au CHC chez les patients ayant une maladie hépatique liée à l'alcool. Ces deux derniers polymorphismes ont un effet additif et semblent également associés à un risque plus élevé de développer une NAFLD même si cela n'a pas pu être validé dans cette étude à cause du nombre limité de patients ayant cette étiologie. *STAT4* (Signal Transducer And Activator Of Transcription 4) est une protéine de l'inflammation alors que *PNPLA3* (patatin-like phospholipase domain-containing 3) et *TM6SF2* (transmembrane 6 superfamily 2) sont impliqués dans le métabolisme des lipides (J. Yang et al., 2019b).

Enfin, le GWAS peut aussi mettre en évidence l'existence de variants protecteurs, diminuant le risque de développer un CHC. C'est le cas du SNP *HSD17B13* rs72613567 qui a un rôle protecteur contre le développement d'une ALD ou NAFLD et potentiellement d'une hépatite liée au VHC. Il a également un rôle protecteur contre le développement d'un CHC lié à une ALD. Le gène *HSD17B13* (17-beta-hydroxysteroid dehydrogenase 13) est également impliqué dans le métabolisme lipidique mais sa fonction précise est encore inconnue (Abul-Husn et al., 2018; J. Yang et al., 2019a).

A l'heure actuelle, il n'existe pas de données permettant d'utiliser les SNPs en clinique. Les prédispositions génétiques observées semblent trop faibles pour organiser un dépistage des patients et les classer en fonction de leur risque de développer un CHC. Des combinaisons de plusieurs SNPs pourraient néanmoins être considérées, en association avec des caractéristiques cliniques, afin de mieux définir les populations à risque et d'adopter les stratégies de surveillance (Rebouissou and Nault, 2020).

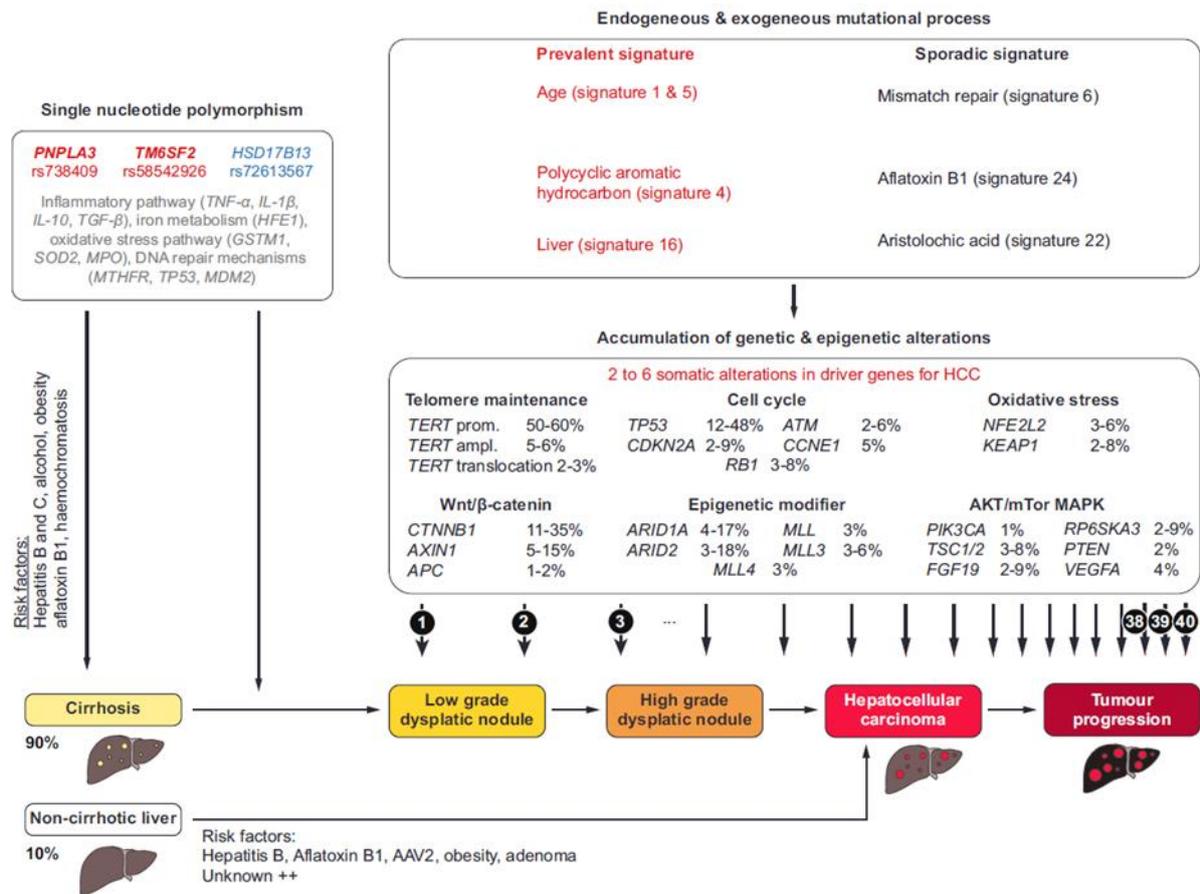


Figure 5 : Prédipositions génétiques et processus mutationnels. Les SNP favorisant ou protégeant contre la cirrhose et le développement d'un CHC sont indiqués respectivement en rouge et en bleu. Ceux nécessitant une validation complémentaire sont indiqués en gris (Adaptée de Rebouissou and Nault, 2020)

2) Mutations somatiques et signatures mutationnelles

La mutagénèse suscite la transformation d'une cellule normale en cellule tumorale. Un génome tumoral reflète donc l'ensemble des processus mutagènes qui se sont déroulés tout au long du cycle de vie de la cellule, qu'ils aient eu lieu avant ou après la transformation néoplasique. Ces processus peuvent être endogènes (liés à l'âge ou à des erreurs dans les mécanismes de réparation de l'ADN) ou exogènes (liés principalement aux expositions environnementales à des carcinogènes comme le tabac ou les rayons ultraviolets). Le nombre moyen de mutations dans une tumeur est très variable en fonction du type de tumeurs. Alors qu'il est inférieur à 10 dans les cancers pédiatriques ou hématologiques, il peut être de l'ordre de centaines ou milliers dans les cancers du poumon, les mélanomes ou les cancers colorectaux présentant

une instabilité des microsatellites. En ce qui concerne le CHC, le nombre de mutations codantes identifiées dans des exomes est de 40 en moyenne (Schulze et al., 2016).

Il est considéré qu'une petite proportion de ces mutations (moins de 10) confèrent à la cellule un avantage prolifératif et sont nommées mutations « driver ». Elles contribuent à l'évolution de la tumeur à un stade donné en influant sur l'initiation, la progression tumorale, la métastase ou encore la résistance au traitement. Les autres mutations accumulées qui sont largement majoritaires sont dites « passager », ce sont des événements secondaires sans conséquences fonctionnelles connues mais qui témoignent des processus mutagènes qu'a subi la cellule. Chacun de ces processus laisse en effet une signature caractéristique sur le génome que l'on appelle signature mutationnelle, et qui tient compte non seulement du type de substitution nucléotidique mais également du contexte tri-nucléotidique dans lequel elle se trouve (c'est-à-dire des deux nucléotides adjacents en 3' et 5'). La signature d'un échantillon tumoral est générée en indiquant la fréquence de mutation pour chacun des 96 triplets nucléotidiques (Nik-Zainal et al., 2015; Schulze et al., 2016).

Cette analyse a d'abord été réalisée lors d'une étude pan-cancer qui a mis en évidence l'existence d'une vingtaine de signatures (Alexandrov et al., 2013) et a ensuite été affinée plus spécifiquement sur une série de tumeurs hépatiques du laboratoire à partir d'analyse d'exomes (Schulze et al., 2015) ou de génomes (Letouzé et al., 2017). De nouvelles signatures sont décrites au fur et à mesure que le nombre d'échantillons analysés augmente. La publication la plus récente fait ainsi état de 49 signatures identifiées en pan-cancer qui ont une origine biologique (Alexandrov et al., 2020).

Concernant les tumeurs hépatiques, dix signatures mutationnelles permettent d'expliquer la quasi-totalité des mutations identifiées dans les CHC (Figures 5 et 6). Le processus sous-jacent a pu être déterminé pour 6 d'entre elles. Cinq signatures sont ubiquitaires et retrouvées dans plus de 50% des CHC : les signatures 1 et 5 sont associées à l'âge, la signature 4 aux hydrocarbures aromatiques polycycliques (HAP) présents notamment dans le tabac, et les signatures 12 et 16 dont l'origine est encore inconnue. La signature 16 a la particularité d'avoir une spécificité hépatique car elle n'est retrouvée que dans les cancers primitifs du foie, et bien que le mécanisme mutationnel reste encore à déterminer, elle serait potentiellement due à un produit du métabolisme hépatique retrouvé en plus grande quantité chez les hommes, les fumeurs et les consommateurs d'alcool. Les cinq autres signatures sont présentes

dans moins de 5% des tumeurs : la signature 6 est associée à des défauts dans la réparation de l'ADN, les signatures 22 et 24 sont caractéristiques de l'exposition à l'Acide Aristolochique et à l'Aflatoxine B1 respectivement, et les signatures 17 et 23 restent inexpliquées (Letouzé et al., 2017).

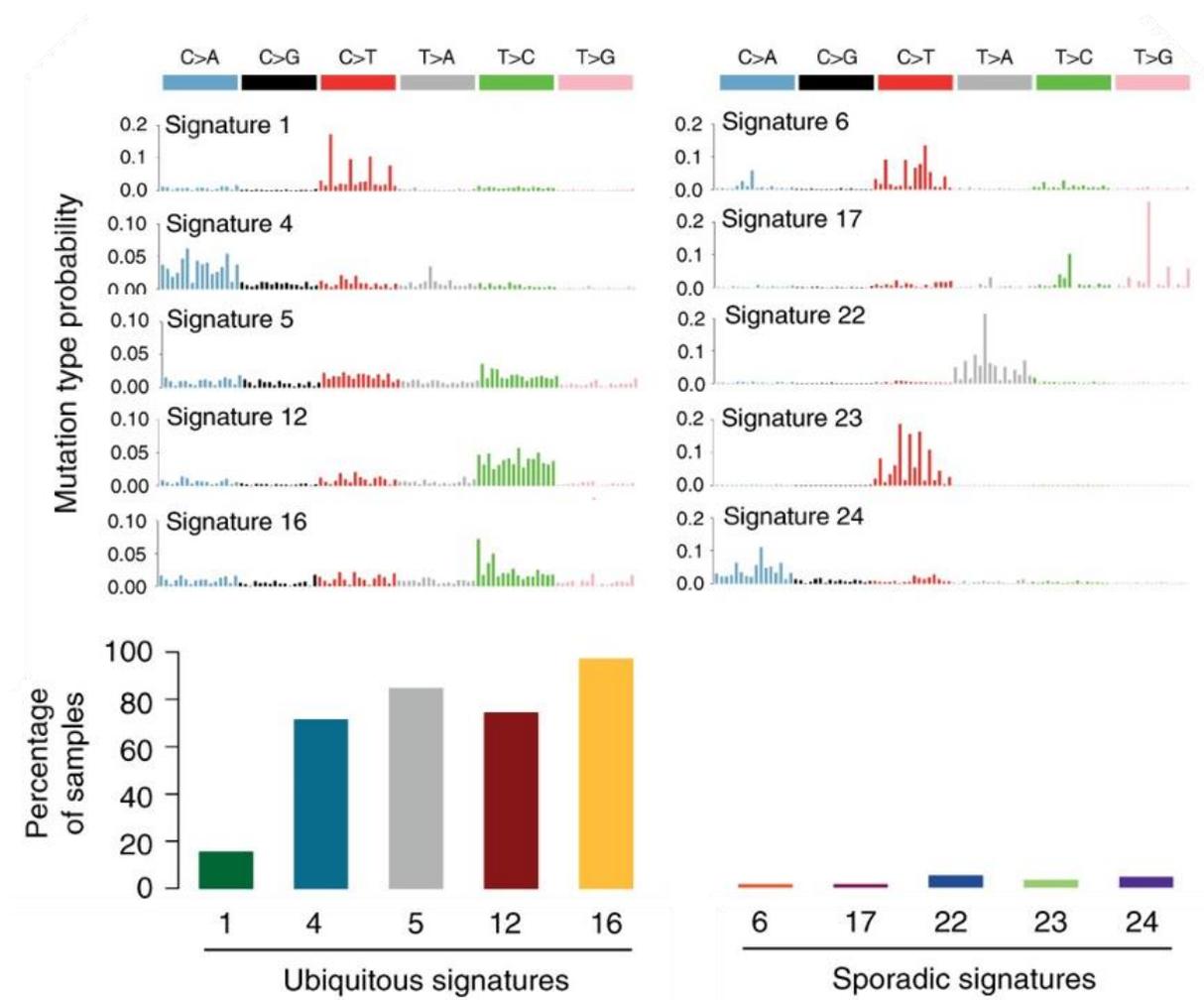


Figure 6 : Les dix signatures mutationnelles observées dans les CHC
(Adaptée de Letouzé et al., 2017).

3) Aberrations et réarrangements chromosomiques

Les altérations chromosomiques peuvent être classées en deux groupes distincts : les aberrations numériques et les altérations structurales. La première catégorie regroupe les phénomènes de gain ou de perte de chromosomes entiers (ou de bras de chromosomes) qui sont en général générés lors de la division cellulaire et qui entraînent une aneuploidie des cellules. La deuxième catégorie concerne les réarrangements chromosomiques tels que les translocations, inversions, délétions et

duplications, qui sont principalement dus à des cassures d'ADN non réparées ou mal réparées (Tahmasebi-Birgani et al., 2019).

Ces modifications du génome conduisent fréquemment à des altérations du nombre de copies (CNA – Copy Number Alterations) de fragments du génome tumoral, qui peuvent être facilement identifiées. Ainsi, une étude pan-cancer a analysé les CNA identifiés dans un grand nombre d'échantillons tumoraux et a observé que leur fréquence est inversement corrélée à leur taille : les CNA de chromosomes entiers sont moins fréquents que les CNA impliquant des bras de chromosomes, qui sont eux-mêmes moins fréquents que les CNA focales. De plus, alors que la plupart des pertes ou gains de bras de chromosomes impliquent une seule copie, certaines CNA focales peuvent avoir des amplitudes très grandes (Beroukhim et al., 2010).

- Aberrations numériques :

Les aberrations numériques fréquemment observées dans les CHC sont des gains des bras 1q, 5p, 6p, 7q, 8q, 17q, et 20q et des pertes des bras 1p, 4q, 6q, 8p, 9p, 13q, 14q, 16p-q, 17p, 21p-q, et 22q. La conséquence fonctionnelle de ces événements peut être le gain d'un oncogène pouvant induire sa surexpression (comme *TERT* sur le bras 5p, ou *MYC* sur le bras 8q) ou la perte d'un gène suppresseur de tumeur annulant son expression (comme *RB1* sur le bras 13q, *AXIN1* sur le bras 16p ou encore *TP53* sur le bras 17p). Parmi ces événements, les plus fréquents sont les gains des bras 1q et 8q (dans plus de 50% des CHC) et les pertes des bras 4q et 8p (Niu et al., 2016).

Ces aberrations numériques peuvent être quantifiés par un score nommé Fraction de Bras de chromosomes Aberrants (FAA – Fraction of Aberrant Arms) correspondant à la proportion des bras de chromosomes qui sont altérés sur plus de 40% de leur longueur. Il a été observé que ce FAA est plus élevé dans les tumeurs présentant des caractéristiques d'agressivité : CHC sur foie non-cirrhotique, liées au VHB, de grande taille, peu différenciés et avec un niveau d'AFP sérique élevé (Guichard et al., 2012).

Par ailleurs, une étude récente sur la polypléidisation a montré que la polypléidie nucléaire est caractéristique du tissu tumoral hépatique, alors que la polypléidie cellulaire y est au contraire moins observée que dans les tissus non-tumoraux. De plus, la quantité d'hépatocytes mononucléaires polypléides est plus importante dans les CHC liés au VHB, peu différenciés et présentant une mutation du gène *TP53*. Ces

CHC polyploïdes sont très prolifératifs et de mauvais pronostic. L'ensemble de ces données suggère que la polyploïdisation serait un moteur du développement du CHC, au moins dans un contexte où le gène *TP53* est muté (Bou-Nader et al., 2020). Les données génomiques montrent également cette polyploïdie des hépatocytes tumoraux (dans plus de 20% des CHC) (Guichard et al., 2012). Plus spécifiquement, des duplications du génome entier ont été retrouvées de façon récurrente dans les analyses de séquençage d'exome et génome (Letouzé et al., 2017; Totoki et al., 2014).

- Aberrations chromosomiques structurales :

En dehors des larges altérations impliquant des bras de chromosomes ou des chromosomes entiers, des CNA spécifiques de certaines régions ont été mis en évidence. Des amplifications focales ont été identifiées dans 25-30% des CHC, les deux loci les plus fréquemment amplifiés étant *TERT* et *CCND1-FGF19*. La fréquence des délétions homozygotes est plus variable (17% et 40% des CHC) mais ces événements entraînent tous des délétions de gènes suppresseurs de tumeurs tels que *CDKN2A-CDKN2B*, *AXIN1* ou encore *PTEN* (Guichard et al., 2012; Totoki et al., 2014).

Ces altérations focales du nombre de copies impactent généralement l'expression des gènes contenus dans les fragments amplifiés ou supprimés. Ces CNA sont par ailleurs souvent la conséquence de réarrangements structuraux tels que des délétions, duplications, inversions ou translocations, qui peuvent également impacter la structure des gènes (via la coupure d'un gène ou la fusion entre deux gènes) ou leur état chromatinien (ce qui influencera éventuellement leur expression).

De la même façon que pour les mutations somatiques, il est possible de déterminer des signatures de réarrangements structuraux (RS). Ces réarrangements sont classés en 38 sous-catégories en fonction de la nature (délétion, duplication, inversion, translocation) et de la taille de l'événement. Une analyse des réarrangements présents dans 35 CHC du laboratoire par séquençage de génome entier a permis d'identifier 6 signatures de variants structuraux, mais la détermination des mécanismes sous-jacents est encore en cours. Néanmoins, la signature RS1 a été associée à des altérations des gènes *CCNA2* et *CCNE1* comme cela sera présenté dans la partie Résultats, et la signature RS2 est présente majoritairement dans les tumeurs présentant une mutation du gène *TP53* (Bayard et al., 2018; Letouzé et al., 2017).

4) Clonalité et chronologie des altérations

Les tumeurs sont la conséquence d'un processus séquentiel d'expansion et de sélection clonales de cellules. Au cours du temps, les cellules accumulent une série d'altérations génomiques et épi-génomiques qui vont déterminer leur comportement face aux pressions de sélection. Malgré l'hétérogénéité qui est donc observée dans le CHC, il est cependant largement admis que les cellules tumorales proviennent à l'origine d'une cellule commune portant les premiers événements moléculaires pro-oncogéniques. Cet ancêtre commun possède donc déjà des altérations génétiques importantes qui seront nécessairement clonales et présentes dans toutes les cellules de la tumeur. Ces altérations du « tronc » peuvent être classées en deux types : les altérations « gatekeeper » (gardiennes) qui sont nécessaires mais pas suffisantes pour l'initiation tumorale et qui sont déjà présentes dans les nodules dysplasiques, et les altérations « driver » qui vont donner ensuite à la cellule un avantage prolifératif et qui sont donc présentes dans les CHC précoces. Lors de la progression tumorale, les cellules vont continuer à accumuler des altérations et seront plus ou moins sélectionnées en fonction des pressions de sélection, générant au final un ensemble de sous-populations avec des altérations dites « sous-clonales » qui donnent au CHC son hétérogénéité (Nik-Zainal et al., 2015; Torrecilla et al., 2017) (Figure 7).

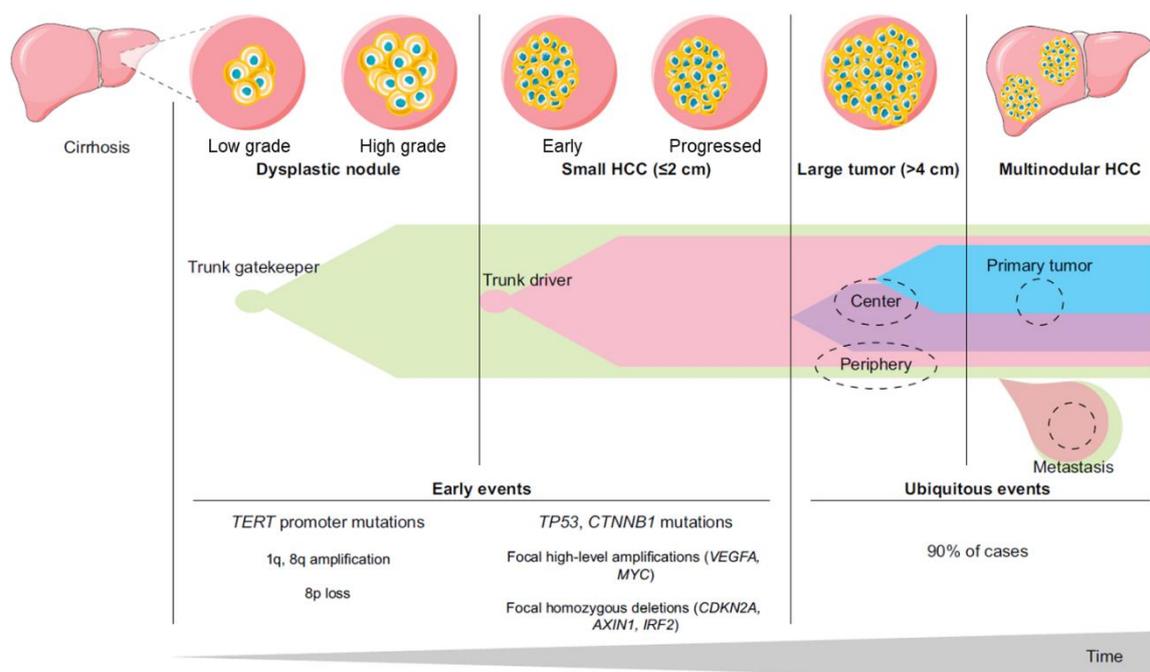


Figure 7 : Les altérations du « tronc » impliquées dans l'hépatocarcinogénèse
(Adaptée de Torrecilla et al., 2017).

Une analyse des mutations clonales et sous-clonales a permis de déterminer que la plupart des mutations sous-clonales ne sont présentes que dans moins de 20% des cellules tumorales et que la proportion de cette sous-clonalité est plus élevée chez les patients jeunes et de sexe féminin et dans les tumeurs du foie bénignes suggérant que ce sont donc des tumeurs plus hétérogènes. La grande majorité des mutations dans les gènes connus comme étant « driver » est clonale (Letouzé et al., 2017).

Plusieurs études ont étudié les altérations moléculaires présentes aux étapes initiales de la carcinogénèse hépatique, en comparant des nodules dysplasiques à des CHC de différents grades. Dans l'ensemble, le nombre de mutations somatiques augmente avec la progression tumorale, ainsi que le nombre d'aberrations chromosomiques plus tardivement (Schulze et al., 2015). La principale altération « gardienne » qui a été décrite est l'altération du promoteur de *TERT*, présente dans 25% des nodules dysplasiques (Nault et al., 2014). Il a aussi été montré que ces nodules précancéreux ont un génome très stable avec peu d'altérations chromosomiques par rapport aux CHC précoces. Seules des gains des bras 1q et 8q et des pertes du 8p ont été identifiées avec une fréquence de 3 à 7% dans ces nodules, et pourraient donc être également des événements initiateurs en altérant des oncogènes tels que *MYC* (8q), *PARP1* (1q) et/ou *MDM4* (1q). Des CNA focales ont en revanche été identifiées de façon récurrente dans des CHC précoces. Bien que plus tardive que pour *TERT*, les altérations de *TP53* et *CTNNB1* semblent exister aussi bien aux stades précoces qu'aux stades avancés et sont aussi des acteurs majeurs de la progression tumorale (Nault et al., 2020; Torrecilla et al., 2017). Les aberrations chromosomiques de grande ampleur telles que les gains synchrones de plusieurs chromosomes ou les duplications du génome entier, sont en revanche des événements très tardifs (Letouzé et al., 2017).

Enfin, il est important de souligner que l'analyse des altérations somatiques a souvent été réalisée dans des tissus tumoraux uniquement. De récentes études ont cependant montré la présence de mutations somatiques sous-clonales dans le tissu non-tumoral, mettant ainsi en évidence l'apparition de mutations récurrentes dans certains gènes du génome d'hépatocytes non transformés. Ces gènes ne sont cependant pas tous connus comme jouant un rôle dans l'initiation ou la progression tumorale. Les mutations sous-clonales du tissu non-tumoral favoriseraient donc la régénération hépatique mais pas nécessairement la transformation maligne, elles pourraient même au contraire avoir une fonction protectrice (Yizhak et al., 2019; M. Zhu et al., 2019).

C. Les voies de signalisation dérégulées

Les études des altérations « driver » ont permis d'identifier les principaux gènes altérés dans les CHC et impliqués dans 6 grandes voies de signalisation (Schulze et al., 2015) (Figures 5 et 8). Les altérations des gènes *TERT*, *TP53* et *CTNNB1* sont les événements les plus récurrents. Elles sont retrouvées de façon ubiquitaire dans les cellules d'une même tumeur et fréquemment conservées entre une tumeur primaire et sa métastase intra-hépatique (dans plus de 80% des cas) (Torrecilla et al., 2017).

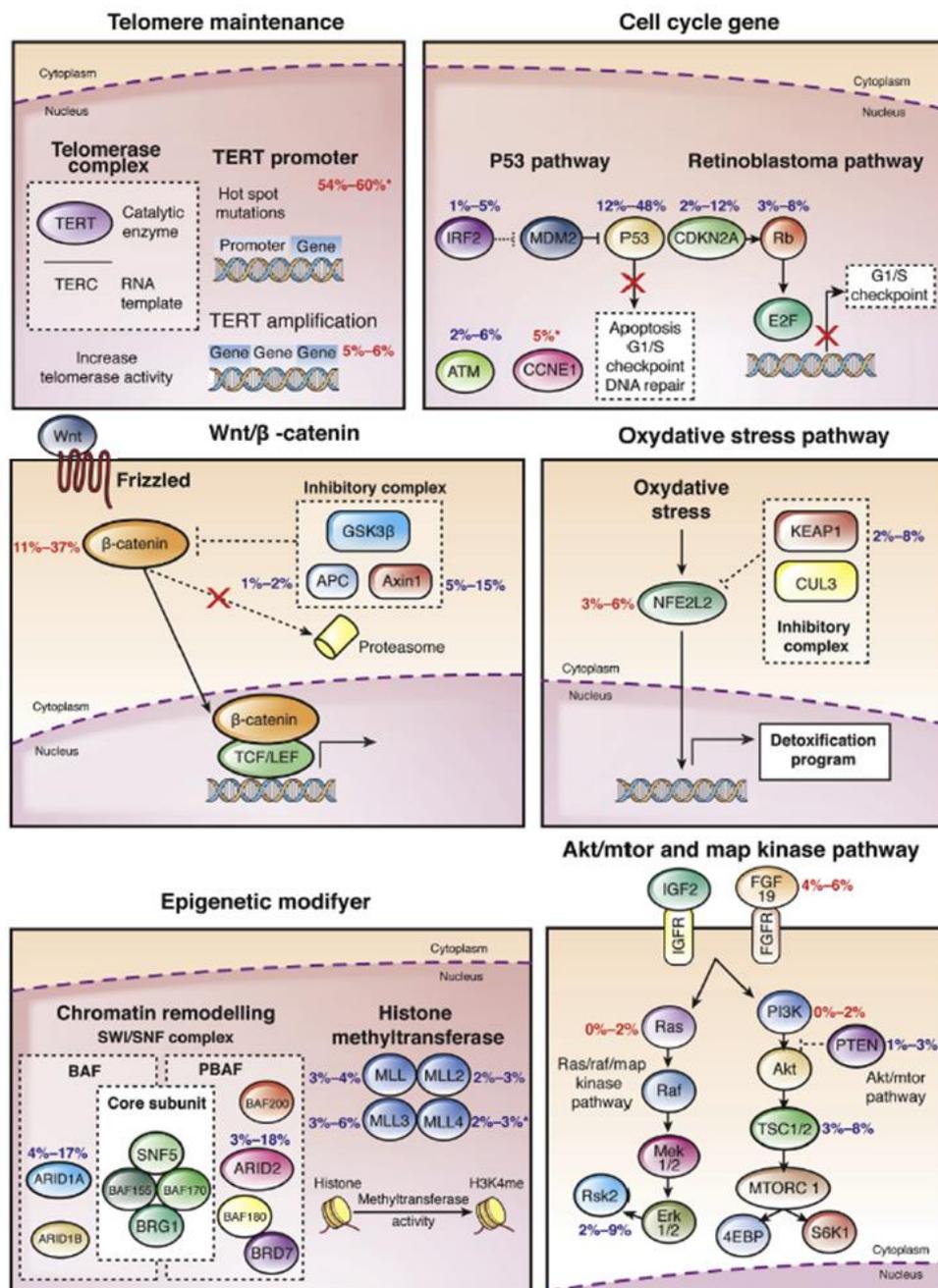


Figure 8 : CHC et voies de signalisation dérégulées (Zucman-Rossi et al., 2015).

1) Maintenance des télomères

Les télomères sont des séquences répétées en tandem de 6 nucléotides (TTAGGG) situées à chaque extrémité chromosomique et liées à des protéines « shelterines » qui empêchent les mécanismes de réparation de l'ADN de considérer l'extrémité de ces séquences terminales comme des cassures double-brins. La réplication de l'ADN avant une division cellulaire induisant systématiquement une perte des nucléotides situés à l'extrémité de chaque chromosome, la présence de séquences télomériques répétées permet de prévenir la perte d'information génétique importante. Le raccourcissement des télomères lors des divisions cellulaires est donc un phénomène physiologique, qui déclenche, après un certain nombre de divisions, l'arrêt de la prolifération cellulaire, la sénescence et l'apoptose de la cellule. La télomérase est une enzyme clé de ce processus qui permet la survie des cellules en synthétisant de nouvelles séquences télomériques. Le gène *TERT* (Telomerase Reverse Transcriptase) code pour l'enzyme du complexe de la télomérase, également composé d'une amorce ARN (TERC) et de la dyskerine, une protéine qui stabilise le complexe. C'est l'expression de *TERT* qui garantit l'expression de la télomérase dans une cellule et permet ainsi son auto-renouvellement (Calado and Young, 2009).

Dans la plupart des cancers humains, la réactivation de la télomérase est nécessaire pour permettre la prolifération des cellules tumorales. Plusieurs types d'altérations ont été identifiées dans les CHC et entraînent l'activation de la télomérase via la surexpression du gène *TERT* (Nault et al., 2019) (Figure 9) :

- Les mutations du promoteur (30-60% des CHC), principalement à deux sites principaux, en positions -124 et -146 par rapport à l'ATG du gène, la mutation en -124 étant plus fréquente (95% des cas) que celle en -146. Ces mutations activent le gène en créant un site de fixation pour des facteurs de transcription des familles ETS (E-Twenty-Six) et TCF (Ternary Complex Factor).
- Les insertions virales du VHB ou du virus AAV2 (10% des CHC), qui seront décrites plus en détail dans la partie III et dans la partie Résultats.
- Les amplifications du gène (5-10% des CHC) qui sont provoquées par des mécanismes encore inconnus.
- Les translocations du gène (3-4% des CHC) qui déplacent le gène et le positionnent sous le contrôle de promoteurs fortement exprimés dans le foie.

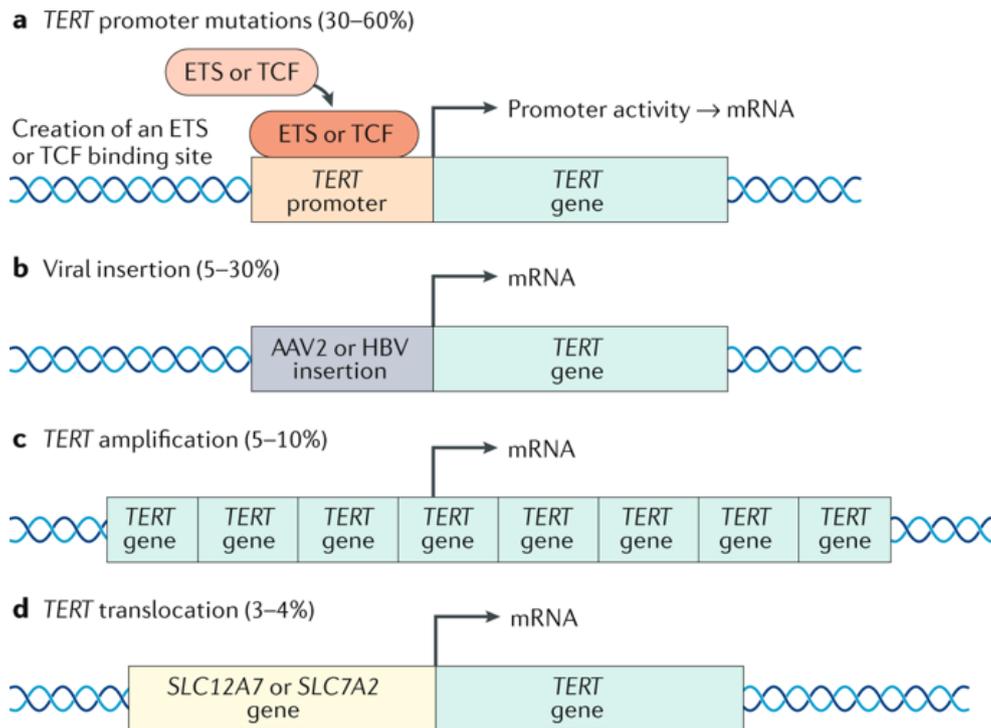


Figure 9 : Mécanismes d'activation de la télomérase via une stimulation de l'expression du gène *TERT* dans le CHC (Nault et al., 2019).

Les mutations du promoteur de *TERT* constituent l'altération la plus récurrente mais elles n'induisent qu'une surexpression modérée du gène par rapport aux autres altérations dont elles sont exclusives. Ces mutations sont plus fréquentes chez des patients âgés, de sexe masculin, avec infection du VHC, mais significativement moins fréquentes chez les patients infectés par le VHB. Elles sont souvent associées à des mutations activatrices du gène *CTNNB1*, suggérant une potentielle coopération oncogénique entre les deux voies de signalisation (Ally et al., 2017; Totoki et al., 2014).

La réactivation de la télomérase est retrouvée dans 90-95% des CHC, ce qui souligne son rôle clé dans la carcinogénèse hépatique. Elle est décrite à la fois dans les CHC sur foie cirrhotique et sur foie sain. Elle est notamment nécessaire pour la transformation maligne des adénomes hépatocellulaires, mais elle n'intervient que tardivement dans ce processus, alors qu'elle est au contraire un des mécanismes les plus précoces de la carcinogénèse sur foie cirrhotique (Zucman-Rossi et al., 2015).

Les nodules cirrhotiques ont en effet la particularité d'être constitués d'hépatocytes sénescents avec des télomères courts. Comme cela a été mentionné précédemment, les maladies chroniques du foie consistent en des cycles répétés de nécrose et de

prolifération des hépatocytes, qui amènent à la cirrhose. Le foie adulte n'ayant pas d'activité télomérase, cette prolifération anormalement élevée induit la sénescence des cellules via le raccourcissement de leurs télomères. Dans ce contexte particulier, la réactivation de la télomérase est donc nécessaire pour empêcher la sénescence des hépatocytes et leur permettre de garder leur capacité de prolifération. L'altération de *TERT* est donc un événement « gatekeeper » de l'initiation tumorale (Rebouissou and Nault, 2020). Cela explique pourquoi *TERT* est le seul gène, parmi ceux qui sont fréquemment altérés dans les CHC, qui présente des altérations récurrentes dans les lésions précancéreuses. Des mutations de ce gène sont en effet observées dans 6% des nodules dysplasiques de bas grade, 19% des nodules de haut grade et 61% des CHC précoces (Nault et al., 2014, 2013). Dans les 5-10% des CHC sans altérations du gène *TERT*, la maintenance des télomères est probablement assurée par un mécanisme de prolongation alternative des télomères comme cela a été observé dans d'autres cancers, mais cela reste à investiguer dans le foie (Nault et al., 2019).

La maintenance des télomères est également primordiale pour éviter la survenue d'une crise télomérique, un état de grande instabilité chromosomique. Cette crise survient lorsque les voies de signalisation des gènes suppresseurs de tumeur *TP53* et *RB1* sont altérées dans une cellule dont les télomères deviennent trop courts. La cellule ne recevant pas la signalisation pour entrer en sénescence, elle subit des altérations chromosomiques de grande ampleur provoquées par la structure anormale de ses télomères. Il peut s'agir de gains ou pertes de chromosomes, de translocations, d'amplifications de larges régions via des cycles BSB (Breakage-Fusion-Bridge), de duplications du génome entier, d'événements de kataegis (hypermutations groupées) ou de chromotripsis (multiples réarrangements groupés). La résolution d'une crise télomérique nécessite la réactivation de la télomérase, et génère ainsi un clone tumoral avec un génome fortement réarrangé (Maciejowski and de Lange, 2017).

2) Contrôle du cycle cellulaire

Des altérations des voies contrôlant le cycle cellulaire sont fréquemment observées dans les CHC et notamment des mutations des gènes suppresseurs de tumeurs *TP53* (12-48%) codant pour la protéine p53 et *RB1* (3-8%) codant pour la protéine Rb (Rétinoblastome) (Zucman-Rossi et al., 2015).

- La protéine p53 agit comme un suppresseur de tumeur en déclenchant l'arrêt du cycle cellulaire, l'apoptose et la sénescence en réponse au stress cellulaire afin de maintenir l'intégrité du génome. La moitié des tumeurs humaines sont porteuses d'altérations codant pour une protéine p53 mutante qui favorise la carcinogénèse via l'augmentation de la prolifération et de la survie cellulaires ou qui régule négativement l'activité des protéines non mutées. La régulation de p53 est principalement effectuée par la protéine MDM2 (Murine Double Minute 2). En interagissant avec son domaine de transactivation, MDM2 bloque la régulation transcriptionnelle médiée par p53, et induit sa poly-ubiquitinylation et sa dégradation par le protéasome. p53 stimule par ailleurs la transcription de MDM2, formant ainsi une boucle de rétroaction autorégulatrice qui contrôle l'équilibre entre les deux protéines (Meng et al., 2014).

Les mutations de *TP53* identifiées dans les CHC se produisent principalement dans son domaine de liaison avec l'ADN, ce qui diminue son affinité pour se fixer sur les séquences spécifiques de ses gènes cibles, et impacte notamment la transcription de MDM2. La diminution de la régulation exercée par MDM2 entraîne alors par conséquent une quantité plus élevée de protéines p53 mutantes dans les cellules. Un « hotspot » principal a été décrit sur le troisième nucléotide du codon 249 du gène *TP53* (substitution G>C induisant la mutation R249S). Parmi les CHC avec des mutations de *TP53*, cette mutation a été observée dans 30% des cas à l'échelle mondiale, et dans plus de 96% des cas en Afrique et en Asie du Sud-Est. Elle est ainsi fortement associée à l'exposition à l'Aflatoxine B1 (AFB1) et à l'infection par le VHB, deux facteurs très présents dans ces régions (Meng et al., 2014; Wang et al., 2019). Cette mutation est aussi associée à un moins bon pronostic par rapport aux autres mutations de *TP53*, potentiellement car elle arrive plus précocement dans la carcinogénèse si elle est induite par l'AFB1 (Woo et al., 2011). Le rôle synergique entre l'exposition à l'AFB1 et le VHB sur l'activation de p53 sera abordé dans la partie III.C.1.

En utilisant l'expression de ses gènes cibles pour évaluer la fonctionnalité de p53, une étude a montré que les tumeurs ayant une protéine p53 faiblement fonctionnelle n'avaient pas toutes une mutation du gène *TP53*, suggérant l'existence d'autres mécanismes inactivant la protéine (Ally et al., 2017a). Des mutations ou des délétions homozygotes du gène *IRF2* (interferon regulatory factor 2) (1-5%) et du gène *ATM* (2-6%) ont notamment été observées dans des CHC. Les altérations de ces deux gènes induisant également une inactivation fonctionnelle de p53, elles sont exclusives des

mutations de *TP53* (Fujimoto et al., 2012; Guichard et al., 2012). Les tumeurs ayant une protéine p53 inactive présentent une plus forte instabilité chromosomique (comme de fréquentes pertes du bras 4q) et une expression plus faible des gènes exprimés dans les hépatocytes matures. L'acquisition d'un profil d'expression semblable à celui des cellules progénitrices pourrait expliquer les caractéristiques agressives de ces tumeurs qui sont généralement associées chez les patients à un risque plus élevé de récurrence et à un taux de survie plus faible (Ally et al., 2017a; Woo et al., 2011).

- La protéine Rb (Rétinoblastome) codée par le gène *RB1*, contrôle la progression cellulaire en régulant la transition G1/S au cours du cycle. Lorsque cette protéine est active, elle forme un complexe de répression transcriptionnelle avec les facteurs de transcription de la famille E2F et diverses protéines modificatrices de la chromatine. Ce complexe bloque les cellules en phase G1 via l'inhibition de la transcription des gènes cibles des protéines E2F tels que *CCNA2* ou *CCNE1*. En réponse à des stimuli de prolifération, le complexe CCND-CDK4/6 (cyclin D-cyclin-dependent kinases 4/6) induit la phosphorylation de Rb pour la rendre inactive, libérant ainsi les facteurs de transcription E2F qui facilitent alors la transition vers la phase S. La phosphorylation de Rb peut notamment être inhibée par la protéine p16 codée par le gène *CDKN2A* (Kitajima and Takahashi, 2017; Sherr and McCormick, 2002). En dehors des mutations de *RB1*, des délétions homozygotes de *CDKN2A* (2-12%) et des amplifications du locus de *CCND1* (5-14%) peuvent ainsi également altérer la régulation du cycle cellulaire effectuée par Rb et favoriser la prolifération des cellules tumorales. Comme pour la voie p53, les altérations de la voie Rb sont associées à des CHC de stades avancés et de mauvais pronostic et sont retrouvées plus fréquemment dans les CHC induits par le VHB (Ahn et al., 2014; Zucman-Rossi et al., 2015).

Cette observation a été confirmée dans une étude récente du laboratoire, qui a identifié un pourcentage plus important de mutations des gènes *RB1* et *TP53* dans des CHC avancés que dans des CHC aux stades précoces, avec une cooccurrence fréquente des deux inactivations. Cela confirme que ces mutations semblent entraîner le développement de tumeurs particulièrement agressives (Nault et al., 2020).

- Enfin, les gènes codant pour les protéines cycline-A2 et cycline-E1 ont aussi été décrits comme altérés dans les CHC et notamment par des insertions virales (Bayard et al., 2018). Cela sera détaillé dans la partie III et dans la partie Résultats.

3) La voie WNT/ β -caténine

La voie WNT/ β -caténine a de très nombreuses fonctions physiologiques, elle est impliquée dans l'embryogenèse et notamment dans le développement hépatobiliaire, la maturation et la zonation hépatique, mais également dans la prolifération et la différenciation cellulaire, et dans l'homéostasie du foie. Dans les hépatocytes matures sains, la voie est inactive et la quantité cytoplasmique de β -caténine est maintenue à un niveau faible grâce à sa dégradation continue contrôlée par un complexe inhibiteur composé des protéines Axine, APC (Adenomatous Polyposis Coli), GSK3 β (Glycogen Synthase Kinase 3 β) et CK1 α (Casein Kinase I α). Ce complexe permet à la β -caténine d'être phosphorylée progressivement au niveau de plusieurs résidus spécifiques (S45, T41, S37 et S33), puis ubiquitinée par l'enzyme β -TrCP et enfin dégradée par le protéasome. La voie est activée lorsqu'un ligand Wnt se lie à un récepteur frizzled, qui entraîne une dislocation du complexe inhibiteur. La β -caténine n'est plus dégradée et transloque dans le noyau où elle favorise la transcription de gènes impliqués notamment dans la prolifération et la migration cellulaires (Perugorria et al., 2019).

La protéine β -caténine est codée par le gène *CTNNB1*. L'activation de la voie WNT/ β -caténine dans les CHC est due à des altérations de ce gène dans 11 à 37% des CHC. Il s'agit de mutations à des « hotspots » des exons 3, 7 et 8 ou de délétions de grande taille impactant l'exon 3, qui sont classées en trois groupes distincts associés à une activation de la voie plus ou moins importante (Rebouissou et al., 2016) (Figure 10) :

- Le premier groupe (54% des CHC mutés *CTNNB1*) rassemble les altérations fréquentes : les mutations des résidus D32 et S37 de l'exon 3 et les larges délétions, qui induisent une forte activation de la voie en modifiant directement le site de liaison avec β -TrCP, empêchant alors la dégradation de la protéine.
- Le deuxième groupe (14% des CHC mutés *CTNNB1*), composé des mutations de l'acide aminé T41 de l'exon 3, entraîne une activation modérée
- Le troisième groupe (25% des CHC mutés *CTNNB1*) contenant les mutations des résidus S45 de l'exon 3, K335 de l'exon 7 et N387 de l'exon 8 ne déclenche qu'une activation faible. Une deuxième altération a cependant été retrouvée dans ces CHC, et consiste souvent en une duplication de l'allèle muté de *CTNNB1*, permettant ainsi d'augmenter la production de la protéine mutée et donc l'activation du signal. Par ailleurs, les mutations des exons 7 et 8, qui sont

très fréquentes dans les AHC, ne sont présentes que dans 5% des CHC mutés *CTNNB1* et sont fréquemment associées à des mutations de *TP53*.

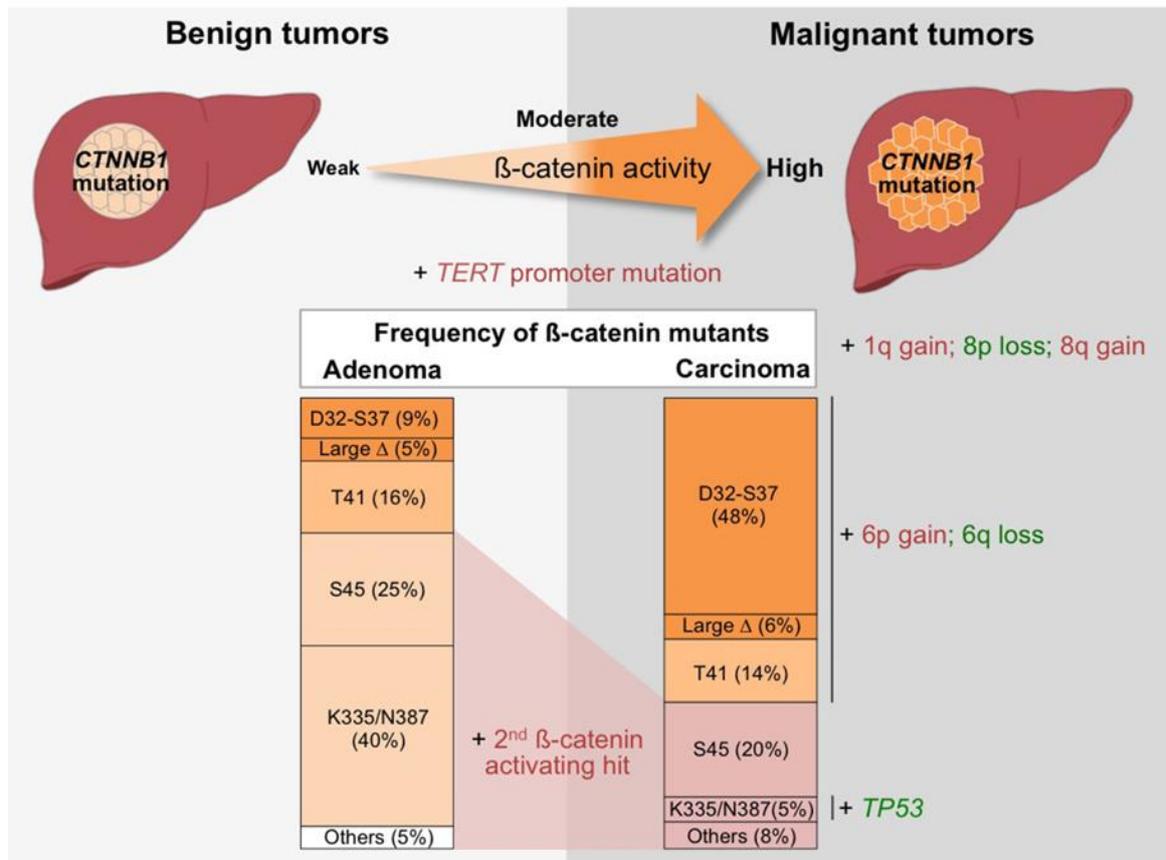


Figure 10 : Fréquence des altérations du gène *CTNNB1* impliqués dans l'hépatocarcinogénèse et contexte génomique (Rebouissou et al., 2016).

Ainsi, une activation modérée ou forte de la voie semble nécessaire pour induire le développement d'un CHC, soulignant le rôle « dose-dépendant » qu'a la signalisation de la voie WNT/ β -caténine dans la carcinogénèse hépatique (Rebouissou et al., 2016).

En dehors des altérations de *CTNNB1*, des mutations inactivatrices des gènes *AXIN1* (5-15% des CHC) ou *APC* (1-2% des CHC) entraînent aussi l'activation de cette voie de signalisation (Llovet et al., 2016). Les CHC avec des mutations *CTNNB1* et *AXIN1* ont cependant des caractéristiques différentes. En effet, alors que les CHC mutés *CTNNB1* sont associés aux maladies liées à l'alcool et présentent une surexpression des gènes cibles classiques de la voie WNT/ β -caténine comme *GLUL* (codant pour la Glutamine Synthase), les CHC mutés *AXIN1* sont associés au VHB, présentent une instabilité chromosomique et appartiennent à un groupe transcriptomique différent comme cela sera décrit dans la partie I.D (Boyault et al., 2007; Perugorria et al., 2019).

4) Autres voies de signalisation importantes dans la carcinogenèse hépatique

Trois autres voies de signalisation ont été identifiées comme étant fréquemment altérées dans les CHC (Schulze et al., 2015; Zucman-Rossi et al., 2015) :

- Les régulateurs épigénétiques : ils sont altérés dans les CHC via des mutations inactivatrices des gènes *ARID1A* (4-17%) et *ARID2* (3-18%), deux gènes jouant un rôle dans le complexe de remodelage de la chromatine SWI/SNF (SWItch/Sucrose Non-Fermentable). Le rôle de ce complexe est de modifier la structure de la chromatine et la position de nucléosomes, ils peuvent donc avoir un rôle suppresseur de tumeurs en modulant la transcription des gènes. Des altérations récurrentes ont également été décrites dans les gènes *MLL* (3-4%), *MLL2* (2-3%), *MLL3* (3-6%) et *MLL4* (2-3%), qui codent pour des protéines modifiant la méthylation des histones (Zucman-Rossi et al., 2015). Par ailleurs, une étude récente du laboratoire a décrit un sous-groupe homogène de tumeurs hépatiques présentant des altérations du gène *BAP1* impliqué dans la modification des histones, qui est inactivé dans 5% des CHC par des mutations ou des translocations. Ces CHC montrent également une activation de la voie PKA (Protéine Kinase A) due à des gains ou pertes chromosomiques, ils se développent majoritairement sur des foies sains en l'absence d'hépatite chronique et présentent des caractéristiques histologiques fibrolamellaires (Hirsch et al., 2019).

- La voie du stress oxydatif : une des voies de réponse au stress oxydatif est la voie NRF2/KEAP1 qui est altérée dans 5-15% des CHC par des mutations activatrices du gène *NFE2L2* (codant pour la protéine NRF2, NFE2-Related Factor 2) ou inactivatrices du gène *KEAP1* (Zucman-Rossi et al., 2015). NRF2 est un facteur de transcription sensible aux signaux de stress cellulaire comme la présence intracellulaire d'espèces réactives de l'oxygène (ERO), qui permet de maintenir l'homéostasie en déclenchant la transcription de gènes impliqués dans la réponse anti-oxydante. La voie est normalement maintenue inactive grâce à la dégradation de NRF2 par le protéasome suite à son ubiquitinylation par le complexe formé par les protéines KEAP1 (Kelch-like ECH-associated protein 1) et CUL3 (Cullin 3). Les mutations induisant l'augmentation de l'expression de *NFE2L2* ou la diminution de l'expression de *KEAP1* entraînent donc une activation constitutive de la voie via

l'inhibition de la dégradation de NRF2. L'activité de cette protéine augmente alors la protection des cellules contre le stress induit par le microenvironnement (notamment en cas de maladie hépatique chronique) et diminue les niveaux d'ERO toxiques, permettant la prolifération et la survie des cellules tumorales (Sporn and Liby, 2012).

- La voie AKT/mTOR et map kinase : ces deux voies de signalisation intracellulaires peuvent être activées par des facteurs de croissance via des récepteurs tyrosine kinase. Le gène *FGF19* code pour un de ces facteurs de croissance, et le gène *MET* pour un de ces récepteurs. Des amplifications des loci de *FGF19* et *MET* ont été identifiées dans 5-10% et 1% des CHC respectivement et entraînent donc l'activation des deux voies de signalisation en aval (Nault et al., 2020; Zucman-Rossi et al., 2015).

La voie PIK3/AKT/MTOR régule le métabolisme, l'inflammation, la motilité et la survie cellulaires. PI3K (Phosphatidylinositol-3-kinases) est un groupe de kinases qui peut déclencher une cascade de signalisation impliquant notamment l'activation d'AKT qui supprime l'effet inhibiteur qu'exerce le complexe TSC1/TSC2 sur un groupe protéique contenant la kinase mTOR (mammalian Target Of Rapamycin). Ce dernier complexe induit alors la synthèse de protéines et la survie cellulaire via l'activation en aval des protéines S6K (Kinases S6) et 4EBP (eukaryotic initiation factor-4E Binding Proteins). Cette signalisation peut être inhibée par le gène suppresseur de tumeur *PTEN* (Phosphatase and TENSing homolog) (Fruman and Rommel, 2014). En dehors des amplifications de *FGF19*, cette voie de signalisation est dérégulée dans certains CHC par des mutations activatrices de *PIK3CA* (0-2%) ou inactivatrices de *TSC1* ou *TSC2* (3-8%) ou des délétions homozygotes de *PTEN* (1-3%) (Zucman-Rossi et al., 2015).

La voie de signalisation RAS/RAF/MAP kinase consiste en l'activation successive des protéines RAF, RAS, MEK1/2, et ERK1/2, qui induit en aval la phosphorylation d'un grand nombre de protéines suscitant diverses réponses cellulaires telles que la prolifération, la différenciation, la survie et la migration (Lavoie and Therrien, 2015). ERK1/2 entraîne notamment la phosphorylation de la protéine RSK2, qui effectue un rétrocontrôle négatif de la voie. En dehors des amplifications de *FGF19*, une dérégulation de cette voie est générée dans certains CHC par de rares mutations activatrices des gènes de la famille RAS (0-2%) ou par des mutations inactivatrices du gène *RPS6KA3* (2-9%) codant pour RSK2 qui suppriment le rétrocontrôle exercée par cette protéine (Zucman-Rossi et al., 2015).

D. Classification moléculaire des CHC

L'ensemble des données moléculaires a permis d'effectuer une classification des CHC afin de répertorier les événements biologiques à l'origine de l'hépto-carcinogénèse. L'objectif est de définir d'éventuels biomarqueurs spécifiques en fonction des voies de signalisation altérées et d'en déduire si possible les traitements adaptés. Les CHC peuvent être divisés en deux groupes principaux : les CHC prolifératifs et les CHC non-prolifératifs, regroupant respectivement les groupes G1-G3 et G4-G6 initialement définis par le laboratoire (Boyault et al., 2007; Calderaro et al., 2019; Hoshida et al., 2009; Rebouissou and Nault, 2020; Zucman-Rossi et al., 2015) (Figure 11).

Les CHC prolifératifs : tumeurs agressives, instables, peu différenciées (G1-G3)

Cette première catégorie de CHC regroupe des tumeurs agressives ayant un mauvais pronostic, souvent associées à une infection par le VHB. Elles sont peu différenciées, et présentent une invasion vasculaire fréquente. Ces tumeurs sont également caractérisées par une instabilité chromosomique élevée et une hypo-méthylation globale de leur ADN, et par des mutations inactivatrices de *TP53*, et des amplifications du locus *CCND1/FGF19*. Elles montrent une activation fréquente des voies de signalisation favorisant la survie cellulaire, comme les voies liées au cycle cellulaire, les voies mTOR, RAS-MAPK et MET. Ce groupe est sous-divisé en deux classes :

- Une classe « Wnt-TGF β » (G2-G3) caractérisée par l'activation de ces deux voies de signalisation et par une réponse immunitaire « épuisée ». Elle contient un sous-groupe de tumeurs plus agressives qui ont un phénotype histologique « trabéculaire massif » (G3).
- Une classe « progéniteur » (G1) définie par la surexpression de gènes et de protéines liés aux marqueurs des cellules hépatiques progénitrices, et associée à des mutations inactivatrices fréquentes de *RPS6KA3* et *AXIN1*.

Les CHC non-prolifératifs : tumeurs stables et différenciées (G4-G6)

Cette deuxième catégorie plus hétérogène regroupe des tumeurs moins agressives, avec une stabilité chromosomique et principalement associées à une infection par le VHC ou à la consommation d'alcool. Ce groupe est aussi sous-divisé en deux classes :

INTRODUCTION

- Une classe avec des CHC mutés *CTNNB1* (G5-G6) caractérisée par une activation de la voie Wnt/ β -caténine modérée ou forte, avec de fréquentes mutations du promoteur de *TERT*, et qui sont immunologiquement froides.
- Une classe moins bien définie (G4) rassemblant notamment des tumeurs avec un phénotype histologique stéatohépatique, des tumeurs présentant une polysomie du chromosome 7, et des tumeurs avec une réponse immunitaire active sur-exprimant des gènes stimulés par l'interféron.

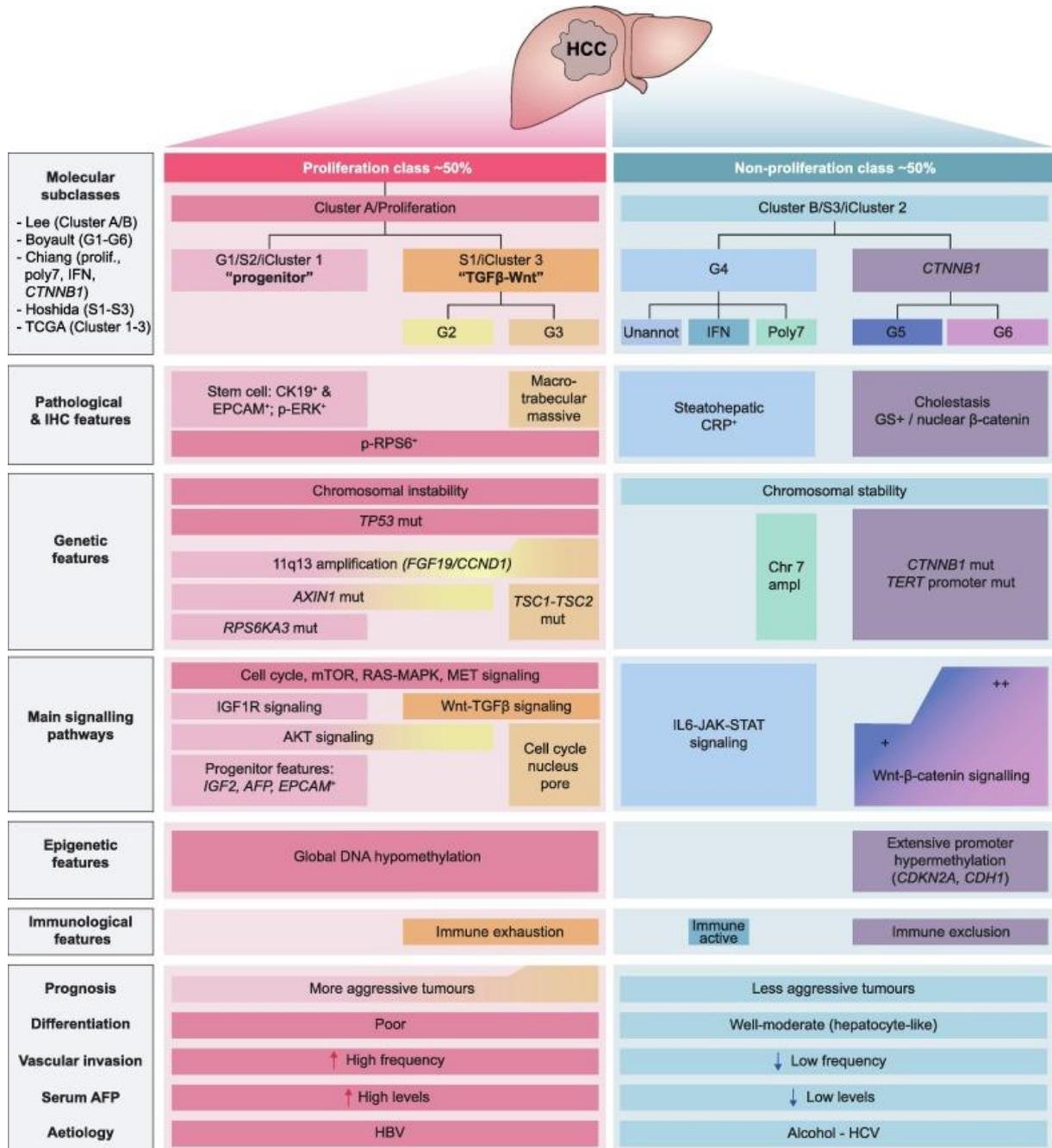


Figure 11 : Classification moléculaire et pathologique des CHC (Rebouissou and Nault, 2020).

II. LE VIRUS DE L'HEPATITE B

L'origine virale de certaines hépatites a été décrite pour la première fois en 1944 et deux types d'infections liées à deux agents pathogènes distincts furent caractérisées en 1947 et nommées Hépatite A et Hépatite B (Block et al., 2016). Mais ce n'est qu'en 1965 que le généticien Baruch Blumberg identifie l'antigène 'Australia' (renommé plus tard antigène HBs) permettant de détecter spécifiquement l'infection par le virus de l'Hépatite B chez les patients. Il révéla ainsi l'existence de nombreux porteurs chroniques dans la population, mit en évidence la relation causale entre l'Hépatite B et le carcinome hépatocellulaire, et élaborera le premier vaccin. Ses travaux ont été récompensés par le prix Nobel de physiologie ou médecine en 1976 (Blumberg, 2002).

A. Biologie moléculaire du virus

Le Virus de l'Hépatite B (VHB) appartient à la famille des *Hepadnaviridae* (Hepatotropic DNA virus). Ces virus au tropisme hépatique ont une organisation génétique et un cycle de réplication communs, et ont la spécificité d'avoir un génome d'ADN circulaire partiellement double-brin et une activité de transcription inverse réalisée par une polymérase virale (Hu and Seeger, 2015a). Cette famille est constituée de trois grands sous-groupes : les *orthohépadnavirus* qui infectent les mammifères et dont fait partie le VHB, les *avihépadnavirus* qui infectent les oiseaux, et un nouveau sous-groupe découvert récemment d'*hépadnavirus* infectant les poissons et les amphibiens (Dill et al., 2016).

Le chimpanzé est le seul modèle d'infection naturelle par le VHB mais les expériences sur ces animaux étant interdites depuis 2013, deux autres modèles d'animaux infectés par des virus de la même famille ont beaucoup été utilisés pour étudier certaines caractéristiques de l'infection et notamment les mécanismes d'intégration : la marmotte infectée par le WHV (Woodchuck Hepatitis Virus) et le canard infecté par le DHBV (Duck Hepatitis B Virus). Récemment, des modèles de souris à foie humanisé ou de souris doublement humanisées (possédant un foie et un système immunitaire humain) ont été développés et permettent désormais d'étudier *in vivo* l'infection par le VHB d'hépatocytes humains (Allweiss and Dandri, 2016; Allweiss and Strick-Marchand, 2020; Strick-Marchand et al., 2015).

1) Structure et organisation du génome

Les particules infectieuses du VHB, nommées particules de Dane, ont été caractérisées dès 1970 par microscopie électronique et mesurent 42 nm de diamètre (Dane et al., 1970; Patient et al., 2009). Elles sont constituées (Figure 12A) :

- d'une enveloppe composée de lipides et de 3 types de protéines de surface appelées Small (S-HBs ou antigène HBs), Medium (M-HBs) et Large (L-HBs)
- contenant une nucléocapside composée de dimères (antigènes HBc)
- à l'intérieur de laquelle se trouve le génome d'ADN viral lié de manière covalente à la polymérase virale.

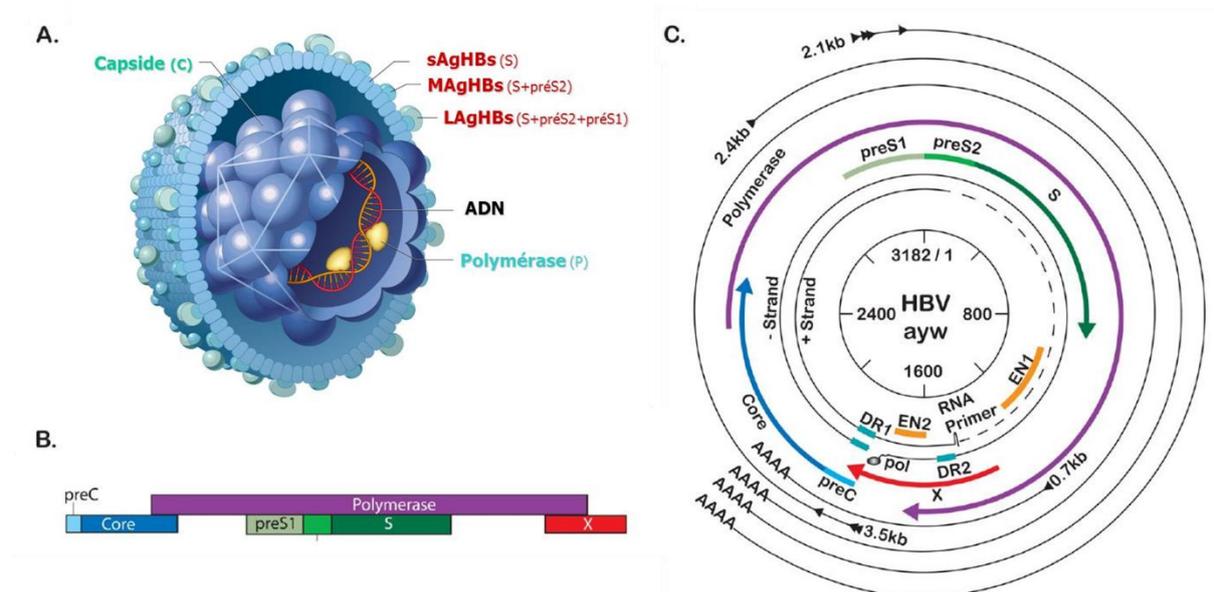


Figure 12 : Structure moléculaire du VHB. (A) Composition d'une particule de Dane. (B) Représentation schématique des ORFs. (C) Génome viral (sérotypage ayw) avec les ORFs (en couleur) et les transcrits (en noir) (Lamontagne et al., 2016).

Une particularité des infections par les *hépadnavirus* est la très grande production de particules subvirales qu'elles engendrent. Il s'agit de particules défectives, non-infectieuses, constituée uniquement de l'enveloppe du virus (AgHBs) et donc sans génome viral. Ces particules subvirales sont produites en quantité 1000 à 100 000 fois supérieure par rapport aux particules complètes. Il existe deux types de particules subvirales : les sphères constituées uniquement des protéines M-HBs et S-HBs, et les filaments constituées des trois protéines de surface dans les mêmes proportions que dans des particules virales complètes (Valaydon and Locarnini, 2017).

Le Virus de l'Hépatite B (VHB) est un virus à ADN qui a une taille de 3,2 kilobases. Dans les particules virales infectieuses, l'ADN du VHB est circulaire et partiellement double-brin, sous une forme nommée ADN circulaire relâché (ADNrc). Le brin (-) est complet et contient une région répétée de 7 à 9 nucléotides à ses extrémités et le brin (+) est incomplet, de longueur variable (entre 50% et 100% de celle du brin (-)) et possède donc un « gap » au niveau de l'extrémité 3'. La polymérase virale est attachée de manière covalente à l'extrémité 5' du brin (-), et une amorce ARN de 18 nucléotides est présente à l'extrémité 5' du brin (+) (Hu and Seeger, 2015). Le génome viral est entièrement codant et composé de quatre cadres de lectures ouverts (Open Reading Frame, ORF) sur le brin (-) qui se superposent : l'ORF préS/S code pour les 3 protéines de surface, l'ORF préC/C code pour les antigènes HBe et HBc, l'ORF P code pour la polymérase virale et l'ORF X code pour une protéine impliquée dans la réplication virale nommée HBx (Valaydon and Locarnini, 2017) (Figures 12B et 12C).

L'ADN viral peut aussi se trouver sous une forme linéaire double-brin (ADNds), ou sous une forme circulaire entièrement double-brin, appelée forme circulaire close de façon covalente (ADNccc). Ces formes seront détaillées dans la partie II.A.4.

2) Eléments structuraux et régulateurs

Le génome du VHB est le plus petit génome parmi les virus à ADN connus. Etant donné que ce génome est entièrement codant, les séquences structurales et régulatrices se doivent d'être intégrées parmi les différents cadres de lecture.

Eléments structuraux : Communs à tous les *hepadnavirus*, les éléments structuraux du VHB sont rassemblés au sein d'une courte séquence du génome qui se situe entre les extrémités 5' des deux brins. La complémentarité du brin (-) complet et du brin (+) partiel sur cette région garantit la circularisation du génome viral. Cette zone est nommée « région terminale cohésive » et contient les motifs structuraux suivants :

- Deux séquences répétées de 11 nucléotides appelés DR1 et DR2 (« Direct Repeat ») qui encadrent la région terminale cohésive et qui jouent un rôle essentiel dans le cycle de réplication du virus (Wei et al., 2010) (Figure 12C).
- Un signal d'encapsidation ϵ (epsilon) situé en amont de la séquence DR1 à l'extrémité 5' de l'ARNpg. C'est une structure secondaire tige-boucle nécessaire

à l'initiation de la transcription inverse de l'ARNpg et à l'encapsidation simultanée de l'ARNpg et de la polymérase virale (Pollack and Ganem, 1994).

- Un signal de polyadénylation situé au début de l'ORF C. Tous les transcrits sont polyadénylés à ce site et ils ont donc tous la même extrémité 3' (Figure 12C).

Eléments régulateurs : Le génome viral contient des éléments régulateurs agissant en *cis* : quatre régions promotrices situées en amont des régions Core, X, PréS1 et PréS2, ainsi que deux séquences « Enhancer » nommées Enh I (270 nucléotides) et Enh II (105 nucléotides) permettant le recrutement de facteurs de transcription ubiquitaires ou spécifiques au foie (Moolla et al., 2002; Quasdorff and Protzer, 2010).

- Le promoteur Core contrôle la transcription de l'ARNpg et de l'ARNm de la protéine PréCore. Il est constitué du Promoteur Basal du Core (PBC) précédé d'une région régulatrice. Cette dernière contient une région CURS (Core Upstream Regulatory Region) pouvant activer le promoteur qui se superpose à la séquence de l'Enh II, et une région NRE (Negative Regulatory Element) encore plus en amont qui régule négativement l'activité du promoteur.
- Les promoteurs préS1 et préS2 contrôlent la transcription des ARNs codant pour les protéines de surface. Le promoteur préS1 est plus faible et négativement régulé par le promoteur préS2. Cette différence entre les deux promoteurs garantit l'équilibre dans la synthèse des 3 protéines de surface.
- Le promoteur X a peu été étudié et son activité semble être faible par rapport aux autres promoteurs viraux.

3) Transcrits et protéines viraux

Transcrits viraux : Le virus utilise l'ARN polymérase II de la cellule hôte afin de générer 5 ARNs messagers (ARNm) de longueurs variables à partir de l'ADNccc. Ces ARNm se terminent tous en un site unique de polyadénylation (Figures 12C et 13).

- Le transcrit utilisé comme matrice pour la réplication virale est l'ARN pré-génomique (ARNpg), il mesure 3,5kb et sa taille est donc supérieure à celle du génome entier. Cet ARN permet également de synthétiser la polymérase virale et la protéine de capsidie nommée antigène HBc (AgHBc).
- L'ARNm préC/C mesure également 3,5kb mais contient quelques bases de plus que l'ARNpg. Il code pour une protéine virale nommée antigène HBe (AgHBe).

- L'ARNm généré par les régions préS2/préS1/S et préS1/S font 2,4kb et 2,1kb respectivement. Le premier code pour la protéine de surface L-HBs, et le deuxième code pour les deux autres protéines de surface M-HBs et S-HBs.
- Le plus petit transcrit de 0,7kb code pour la protéine HBx.

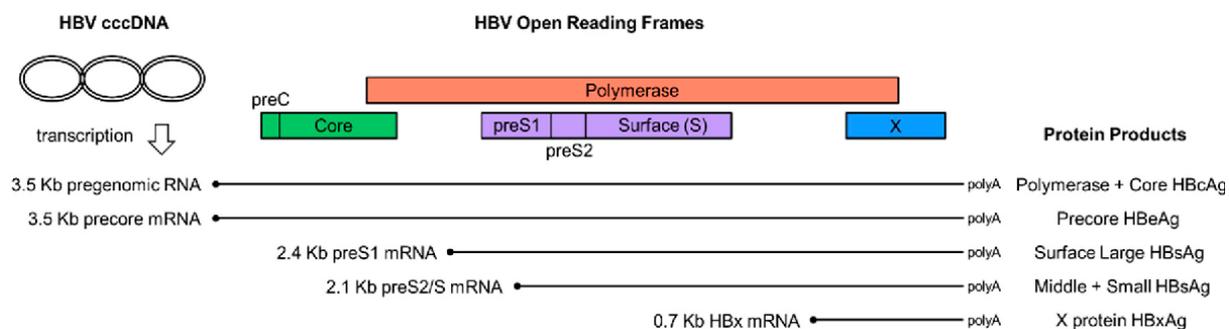


Figure 13 : Description des 5 transcrits et des 7 protéines virales du VHB
(Hadziyannis and Laras, 2018).

D'autres transcrits peuvent être générés suite à l'épissage alternatif de l'ARNpg et représentés jusqu'à 30% des ARNpg totaux. Ces ARN épissés peuvent être rétro-transcrits de la même façon que l'ARNpg et être retrouvés dans des particules virales qui sont alors défectives. Il a été montré que le nombre de particules défectives dans le sérum de patients infectés est corrélé avec la maladie hépatique et qu'il augmente avant le développement d'un carcinome hépatocellulaire (Bayliss et al., 2013).

Protéines virales : Les ARNm décrits permettent la synthèse de 7 protéines virales.

- L'ORF P correspond à 70% du génome viral et code pour la polymérase virale. Cette protéine est composée de quatre sous-domaines : une transcriptase inverse permettant la synthèse d'ADN à partir de l'ARNpg, une Ribonucléase H permettant de dégrader l'ARNpg ayant servi de matrice à la réplication, une région « Spacer » assurant la flexibilité de la protéine et une région terminale qui se lie de manière covalente à l'amorce d'ADN pendant la synthèse du brin (-) (Nassal, 2008).

- L'ORF C code pour la protéine AgHBc qui a comme propriétés de gérer l'auto-assemblage de la capsid et de créer une interaction avec les protéines de surface pour former les particules virales. La séquence protéique contient également un motif de localisation nucléaire afin de faire entrer le génome viral dans le noyau des cellules hôtes et un domaine pour fixer l'ARNpg lors de l'encapsidation (Ludgate et al., 2016). Avec quelques acides aminés supplémentaires, la protéine préCore est codée par

l'ORF préC/C. Un clivage post-traductionnelle des régions N- et C-terminales de cette protéine donne lieu à une protéine soluble nommée antigène HBe (AgHBe). Le rôle de l'AgHBe n'est pas encore entièrement compris, elle n'est pas nécessaire à la réplication virale mais elle semble participer à la modulation de la réponse immunitaire de l'hôte et à l'établissement de l'infection chronique (Chen et al., 2005).

- L'ORF préS1/préS2/S code pour les protéines de surface, les antigènes HBs (AgHBs) qui existent sous trois formes : S-HBs, M-HBs et L-HBs. La région S qui est commune aux trois protéines contient les domaines transmembranaires ainsi qu'un domaine antigénique qui rend ces protéines fortement immunogènes (Valaydon and Locarnini, 2017). La protéine S-HBs qui représente la majeure partie des protéines de l'enveloppe est indispensable à l'assemblage de l'enveloppe pour former les particules virales ou subvirales. Cela n'est en revanche pas le cas pour la protéine M-HBs dont le rôle est encore inconnu (Patient et al., 2009). Par ailleurs, la région préS1 qui est présente uniquement dans la plus grande protéine de surface L-HBs contient un site de myristoylation à son extrémité N-terminale qui permet la reconnaissance et l'entrée des particules virales dans les hépatocytes via le récepteur NTCP (sodium-tauchlorate cotransporting polypeptide). Dans une autre conformation, la protéine L-HBs permet aussi l'enveloppement de la nucléocapside (Glebe, 2007).

- La région X permet la synthèse de la plus petite protéine virale, la protéine HBx. Le rôle de cette protéine a longtemps été débattu au sein de la communauté scientifique et de nombreuses fonctions lui ont été attribuées en fonction du contexte cellulaire et du modèle utilisés et de la quantité de protéines HBx exprimée. En réalité, HBx est synthétisée en très faible quantité au cours de l'infection virale (Wei et al., 2010). Néanmoins, il est désormais avéré que cette protéine est indispensable à la transcription de l'ADNccc (Lucifora et al., 2011). HBx entraîne notamment la dégradation par le protéasome du complexe Smc5/6, et empêche ainsi l'interaction du complexe avec l'ADNccc qui inhibe la réplication (Decorsière et al., 2016). Par ailleurs, la protéine HBx jouerait également un rôle dans l'hépatocarcinogénèse, qui sera détaillé dans la partie III.A.3 de l'introduction.

A ces 7 protéines, il faut ajouter une dernière protéine qui a été observée dans des biopsies du foie de patients atteints d'Hépatite B chronique. Cette protéine est issue d'un transcrit appelé SP1 généré par l'épissage alternatif de l'ARNpg et elle a donc été nommée HBSP (Hepatitis B Spliced Protein) (Soussan et al., 2003). Sa fonction n'a

pas encore été totalement déterminée mais elle favoriserait la persistance de l'infection du VHB en inhibant la réponse des cellules immunitaires (Duriez et al., 2017).

4) Cycle de réplication

Le cycle de réplication du VHB (Figure 14) peut être décomposé en plusieurs étapes regroupées en 3 groupes (étapes précoces, intermédiaires et finales) qui vont être détaillées ci-dessous (Nassal, 2015; Tu et al., 2017; Valaydon and Locarnini, 2017).

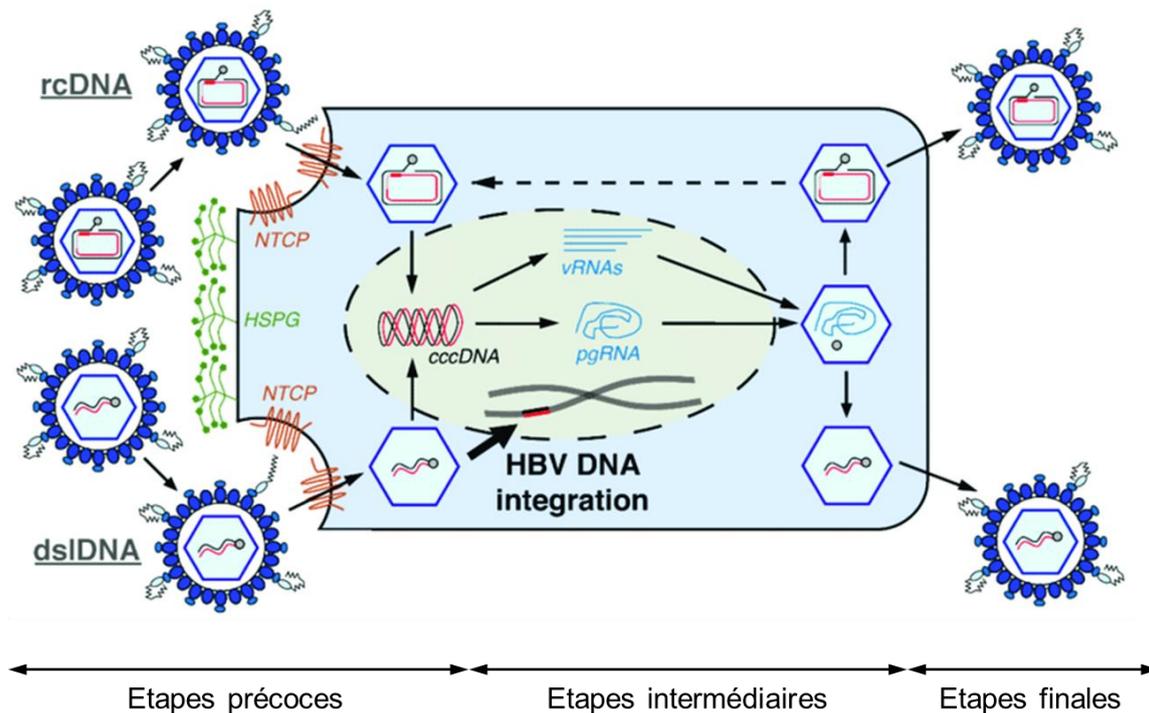


Figure 14 : Cycle de réplication du VHB (Adaptée de Budzinska et al., 2018a).

Etapes précoces : entrée du virus et formation de l'ADNccc

- La première étape est l'entrée du VHB dans la cellule hôte. La particule virale s'attache d'abord à l'hépatocyte via des récepteurs HSPG (Heparan Sulfate ProteoGlycans) puis via le récepteur NTCP qui n'est exprimé que par les cellules du foie. C'est l'interaction de forte spécificité entre le domaine préS1 de la protéine L-HBs et ce récepteur NTCP situé à la surface des hépatocytes qui permet l'endocytose de la particule virale.
- Une fois dans le cytoplasme, les protéines de surface du virion sont clivées, permettant la libération de la nucléocapside qui va être transportée jusqu'au noyau par l'intermédiaire des microtubules.

- La capsid étant de petite taille, elle pénètre dans le noyau à travers les pores nucléaires et libère le génome viral qu'elle contient sous forme d'ADNrc.
- L'ADNrc partiellement double-brin est alors converti en ADNccc entièrement double-brin. Après détachement de la polymérase virale liée à l'extrémité 5' du brin (-), l'amorce ARN en 5' du brin (+) et la séquence redondante du brin (-) sont éliminées. Le brin (+) est ensuite complété et ses extrémités liguées grâce à l'intervention de facteurs cellulaires de la cellule hôte et notamment de protéines impliquées dans les mécanismes de détection et de réparation des dommages à l'ADN. L'ADNccc se lie à des protéines virales et cellulaires (histones et non-histones) pour se compacter sous forme de mini-chromosome. Cette forme stable de l'ADN viral est à l'origine de la réplication et permet la persistance de l'infection dans les hépatocytes. L'ADNccc est produit très précocement, dans les 24 heures suivant l'infection, et sa quantité moyenne est d'une copie par cellule (alors que celle de l'ADNrc peut atteindre plus de 1000 copies par cellule). La durée de vie d'une molécule d'ADNccc n'est pas encore connue précisément bien que des modèles d'hépatocytes primaires ont permis de déterminer une demi-vie de l'ordre de 30 à 60 jours, et sa persistance au cours des divisions cellulaires est encore largement débattue.

Etapas intermédiaires : transcription de l'ADNccc et transcription inverse de l'ARNpg

- L'ADNccc sert de matrice pour la transcription par l'ARN polymérase II de la cellule hôte des 5 transcrits décrits précédemment. Ces ARNm viraux sont ensuite transportés vers le cytoplasme pour être traduits par la cellule hôte et permettre ainsi la synthèse des différentes protéines virales.
- Une fois synthétisée, la polymérase virale se lie de façon covalente à l'extrémité 5' de l'ARNpg au niveau de la structure secondaire ϵ , ce qui déclenche le recrutement des protéines de capsid et l'encapsidation du complexe ARNpg/polymérase dans le cytoplasme.
- A l'intérieur de la capsid et en parallèle de sa maturation, l'ARNpg est rétro-transcrit afin de reformer l'ADN viral (Figure 15). Cela est initié par la synthèse d'une amorce ADN de 3 nucléotides effectuée par la protéine terminale de la polymérase virale au niveau de la structure ϵ de l'ARNpg. Tout en restant liée de façon covalente à la polymérase, cette amorce est ensuite transloquée vers le motif DR1 à l'extrémité 3' de l'ARNpg et la polymérase peut alors effectuer

une transcription inverse pour générer le brin (-) de l'ADN viral, tout en hydrolysant au fur et à mesure l'ARNpg ayant servi de matrice grâce à son domaine RNase H. Seule une séquence de 18 nucléotides en 5' de l'ARNpg est préservée de la dégradation afin d'être utilisée comme amorce pour la synthèse du brin (+). Dans 90% des cas, cette amorce est transloquée vers le motif DR2 permettant ainsi la formation de l'ADNrc, qui est circularisé grâce à la complémentarité des séquences des deux brins dans la région terminale cohésive. Dans les 10% restants, l'amorce est transloquée vers le motif DR1, ce qui aboutit à la formation de la forme ADNdsl qui reste linéaire.

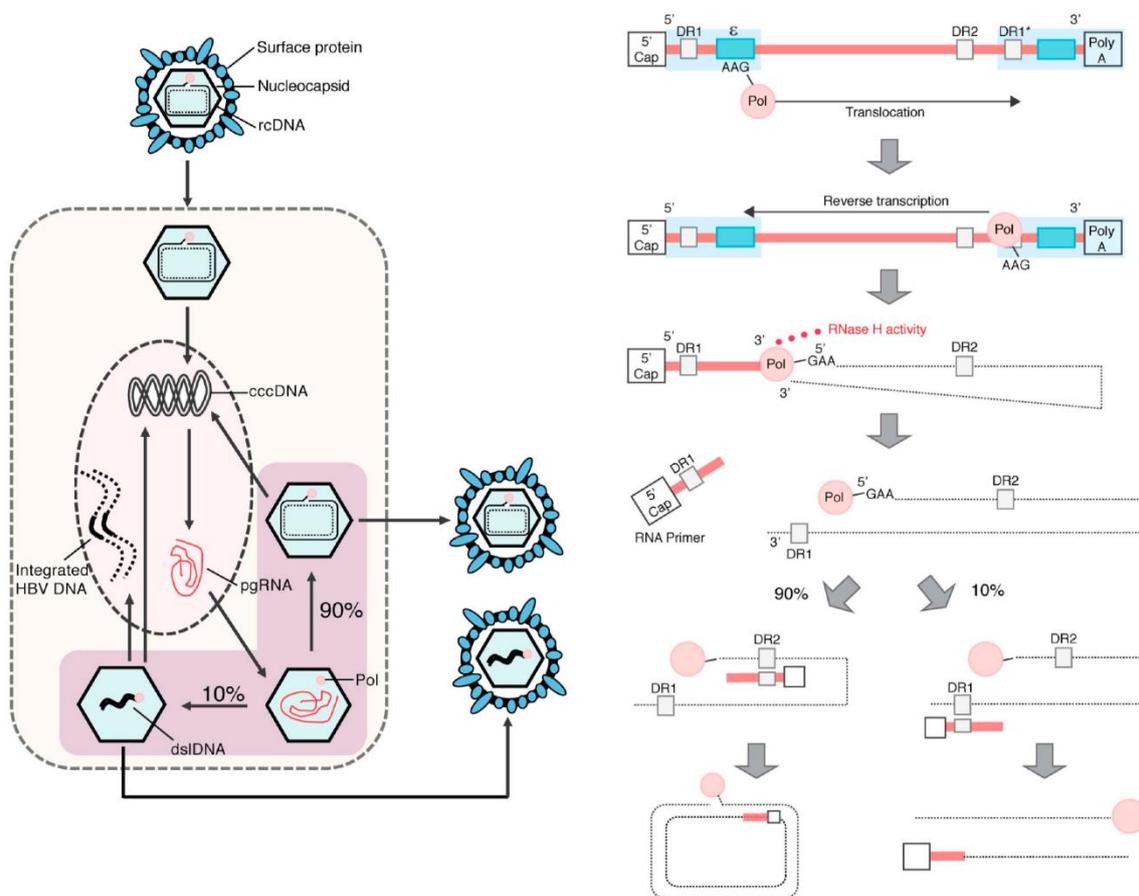


Figure 15 : Etape de rétro-transcription de l'ARNpg pour former l'ADNrc (90%) ou l'ADNdsl (10%) (Tu et al., 2017).

- D'autres formes intermédiaires de particules virales peuvent également être formées et parfois sécrétées, telles que des nucléocapsides non-enveloppées contenant de l'ADNrc ou des nucléocapsides enveloppées contenant de l'ARNpg, mais le rôle de ces formes mineures n'est pas connu à l'heure actuelle.

- Une partie des nucléocapsides contenant de l'ADNrc ou ADNdsI va retourner dans le noyau de la cellule hôte pour être recyclées et ainsi permettre la maintenance d'un pool d'ADNccc dans le noyau.
- Dans le noyau, l'ADN viral sous forme ADNdsI peut également être intégré dans le génome de la cellule hôte. Ce processus d'intégration ne joue cependant aucun rôle dans le cycle de réplication virale et sera détaillé dans la partie III.B.

Etapas finales : assemblage et libération du virion

- L'autre partie des nucléocapsides qui ne retourne pas vers le noyau migre vers le réticulum endoplasmique où les trois protéines de surface du virus sont synthétisées. L'assemblage est rendu possible par l'interaction entre les protéines HBc et les protéines L-HBs.
- Les virions enveloppés sont alors conduits à la membrane cytoplasmique où ils sont relâchés à l'extérieur de la cellule. Ils peuvent contenir l'ADN viral sous forme ADNrc ou sous forme ADNdsI.
- Des particules subvirales formées par les antigènes HBs ainsi que des antigènes HBe sont également sécrétés à l'extérieur de la cellule hôte.

Dans le cas d'une infection par un virion contenant de l'ADNdsI, ce dernier pourrait être complété pour former de l'ADNccc comme cela a été montré dans les modèles animaux. Ce nouvel ADNccc contenant des insertions, il ne permet pas la synthèse d'ADNrc mais pourrait représenter néanmoins une large proportion du pool d'ADNccc dans le noyau des cellules infectées. Par ailleurs, cet ADNccc modifié pourrait potentiellement être transformé en ADNccc « classique » via un mécanisme de recombinaison homologue (Yang et al., 1996; Yang and Summers, 1998).

B. Biologie de l'infection

Longtemps négligées par rapport à d'autres maladies infectieuses telles que la tuberculose, le paludisme ou le VIH, les hépatites virales ont été ajoutées en 2015 à la liste des maladies à combattre dans le cadre du Programme de développement durable à l'horizon 2030 de l'Organisation des Nations Unies (ONU).

L'Organisation Mondiale de la Santé (OMS) a donc adopté en mai 2016 une stratégie mondiale sur les hépatites virales pour les années 2016-2021. Le but est d'éliminer les

hépatites virales des problèmes de santé publique majeurs, en mettant l'accent sur les Hépatites B et C qui regroupent 95% des décès liés aux hépatites virales. L'OMS souhaite atteindre cet objectif en réduisant de 90% les nouvelles infections par ces deux virus et de 65% le taux de mortalité associé. Les moyens mis à avant pour l'Hépatite B sont l'augmentation de la couverture vaccinale dès la naissance, le test systématique des dons de sang, l'amélioration des conditions d'injections, l'augmentation du nombre de personnes diagnostiqués ainsi que du nombre de personnes sous traitement (World Health Organization, 2016a).

1) Épidémiologie

L'ensemble des hépatites virales (A, B, C, D et E) a été la cause en 2015 de 1,45 millions de décès dans le monde. Environ 47% de ces décès étaient imputables au VHB, 48% au VHC et le reste aux virus des Hépatites A et E. Cette mortalité est comparable à celle liée au VIH ou à la tuberculose mais alors que dans ces deux derniers cas les chiffres sont en diminution d'année en année, le nombre de décès liés aux hépatites virales est encore en augmentation (Figure 16A). Ce nombre a en effet augmenté de 63% entre 1990 et 2013 (Stanaway et al., 2016). En 2015, le nombre de patients ayant une hépatite B chronique était estimé à 3,5% de la population mondiale (257 millions de personnes). Seulement 9% d'entre eux avaient été diagnostiqué (22 millions), et seulement 8% de ces derniers (1,7 millions) bénéficiaient d'un traitement contre l'infection (Thomas, 2019; World Health Organization, 2017) (Figure 16B).

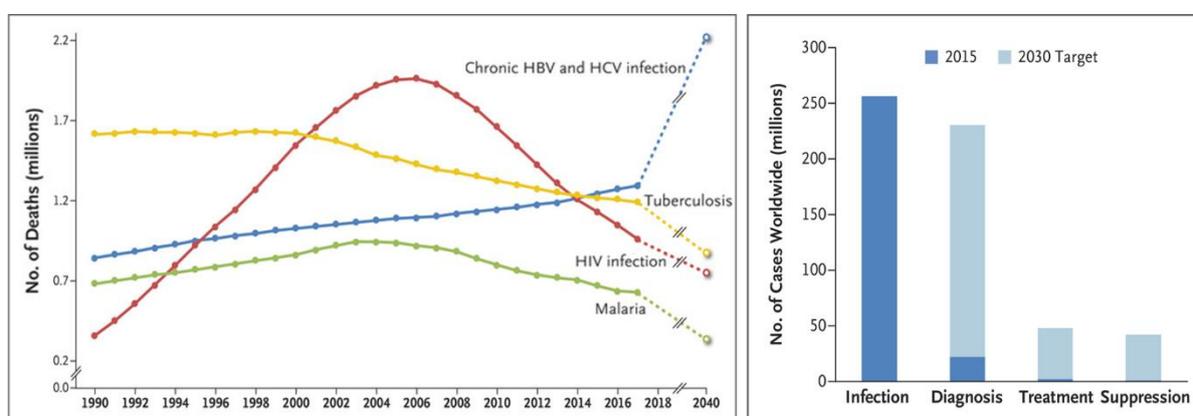


Figure 16 : Épidémiologie du VHB dans le monde. (A) Evolution de la mortalité annuelle mondiale des principales maladies infectieuses. (B) Situation en 2015 et objectifs pour 2030 de la prise en charge de l'infection par le VHB (Thomas, 2019).

De plus, ces chiffres ont une grande disparité en fonction des régions du monde. Les régions de forte endémicité sont l'Afrique et le Pacifique occidental (regroupant l'Asie de l'Est et l'Océanie) qui rassemblent 68% des personnes infectées. Alors que la prévalence de l'infection est respectivement de 6,1% et 6,2% dans les deux régions mentionnées, elle n'est que de 0,7% en Amérique, 1,6% en Europe et 2% en Asie du Sud-Est. La prévalence et l'incidence de l'infection par le VHB sont respectivement 7,4 fois et 9,2 fois plus élevées dans les pays à faible revenu que dans les pays à revenu élevé. Cependant la proportion de personnes diagnostiquées est de 0,8% dans les premiers contre 18% dans les derniers, et parmi ces personnes, la proportion qui est sous traitement est de 9% et 14% respectivement (World Health Organization, 2017).

La mortalité liée à l'infection par le VHB peut être due directement à une infection aiguë mais elle est plus fréquemment causée par le développement d'une cirrhose ou d'un CHC chez les patients ayant une infection chronique, comme cela sera développé dans la partie suivante sur l'histoire naturelle de l'infection. En 2016, 11% des décès liés à l'Hépatite B faisaient suite à une infection aiguë, 50% à une cirrhose et 39% à un cancer du foie. Mais ces pourcentages sont également très variables suivant les régions du monde (World Health Organization, 2016b) (Figure 17).

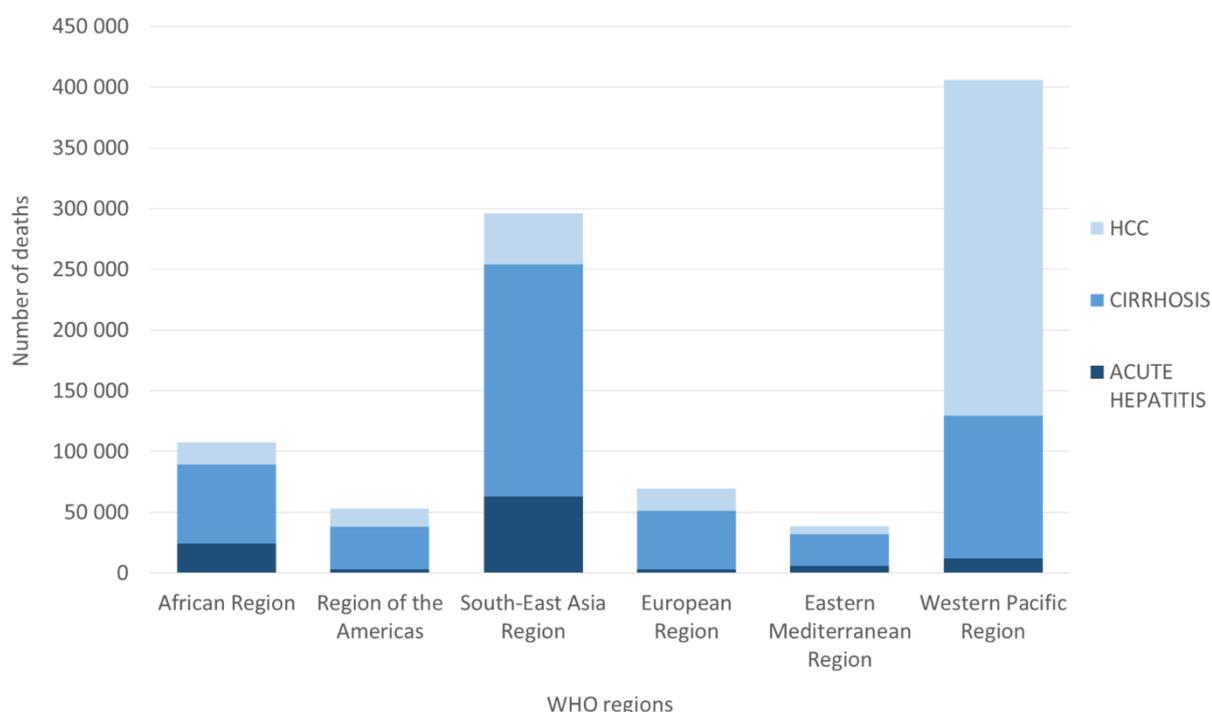


Figure 17 : Mortalité liée au VHB dans les six régions définies par l'OMS (World Health Organization, 2016b).

2) Histoire naturelle de l'infection

Les virions peuvent être présents en grande quantité dans le sang (entre 10^5 et 10^{10} copies/mL) mais également dans la salive, la sueur et dans les sécrétions (sperme, sécrétions vaginales, selles et urine) dans une moindre mesure (inférieur à 10^6 copies/mL) (Kidd-Ljunggren et al., 2006). Le VHB se transmet donc via un contact direct ou indirect avec du sang ou du sperme de personnes infectées. Trois modes de transmission principaux ont été identifiés. Dans les régions très endémiques, le VHB est transmis majoritairement par voie périnatale, alors que dans les régions peu endémiques, ce sont les transmissions par voie sexuelle qui prédominent. La troisième source d'infection est liée à la réutilisation d'aiguilles ou de seringues ayant été en contact avec du sang contaminé ou à la transfusion de sang contaminé. Le VHB est un virus non-cytopathique : les dommages observés sur le foie sont provoqués par la réponse immunitaire de l'hôte infecté (Trépo et al., 2014).

Au total, près de 30% de la population mondiale présente des marques d'infection par le VHB passée ou présente. L'histoire naturelle de l'infection est schématisée sur la Figure 18 ci-dessous.

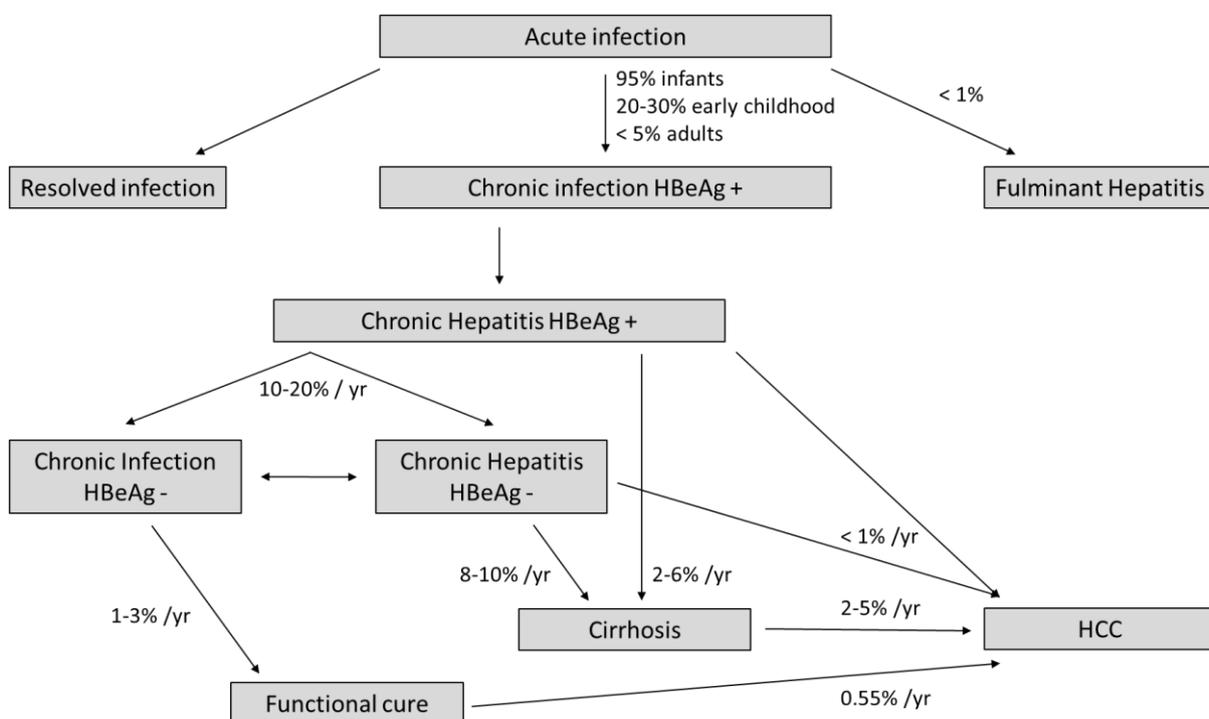


Figure 18 : Histoire naturelle de l'infection par le VHB
(Adaptée de Lampertico et al., 2017; Trépo et al., 2014; Yim and Lok, 2006).

Infection aigüe :

Après contamination par le VHB, le temps d'incubation varie entre 2 et 3 mois (avec des extrêmes allant de 1 à 6 mois). L'incubation est suivie d'une période prodromique qui dure moins de deux semaines et dont les signes sont une légère fièvre, de la fatigue, des nausées et des douleurs abdominales. Il est très fréquent que l'infection passe du coup inaperçue auprès des patients. En effet, l'infection aigüe n'entraîne aucun symptôme clinique dans 60 à 80% des cas, ou souvent uniquement les symptômes non spécifiques mentionnés. Ils peuvent être accompagnés d'une jaunisse associée à une urine sombre (dans moins d'un tiers des cas) (Dény and Zoulim, 2010).

Au niveau biologique, le premier marqueur à apparaître est l'AgHBs qui augmente avec la quantité d'ADN viral dans le sang et l'AgHBe qui est également détectable de façon transitoire. Les anticorps anti-HBc apparaissent peu après et restent stables tout au long de la vie du patient, ils témoignent du contact avec le virus. Simultanément, la cytolysse du foie est reflétée par une augmentation du niveau des enzymes transaminases dans le sérum : les ALanine AminoTransférases (ALT) et les ASpartate AminoTransférases (AST). La résolution de l'infection est caractérisée par la disparition de l'ADN viral dans le sérum et la séroconversion HBs, c'est-à-dire la disparition des AgHBs et l'apparition d'anticorps anti-HBs. Dans moins de 1% des cas, l'hépatite aigüe progresse en hépatite fulminante, qui a un taux de mortalité de 80% en l'absence de transplantation hépatique (Trépo et al., 2014).

L'infection aigüe peut se transformer en infection chronique si l'AgHBs persiste dans le sang pendant plus de six mois. Le risque de transformation en hépatite chronique diminue avec l'âge : il est de 95% chez les nouveau-nés, de 20-30% chez les enfants (âgés de 1 à 5 ans) et de moins de 5% chez les adultes (Trépo et al., 2014).

Infection chronique :

L'infection chronique est un processus dynamique qui reflète l'interaction entre la réplication virale et la réponse du système immunitaire de la personne infectée. Tous les patients ayant une infection chronique du VHB ne développent pas une hépatite chronique (Lampertico et al., 2017). L'évolution de la maladie chronique peut être décomposée en 5 phases qui ne sont pas nécessairement séquentielles et dont la terminologie a été modifiée en 2017 par l'EASL (European Association for the Study of the Liver) (Lampertico et al., 2017) (Figure 19).

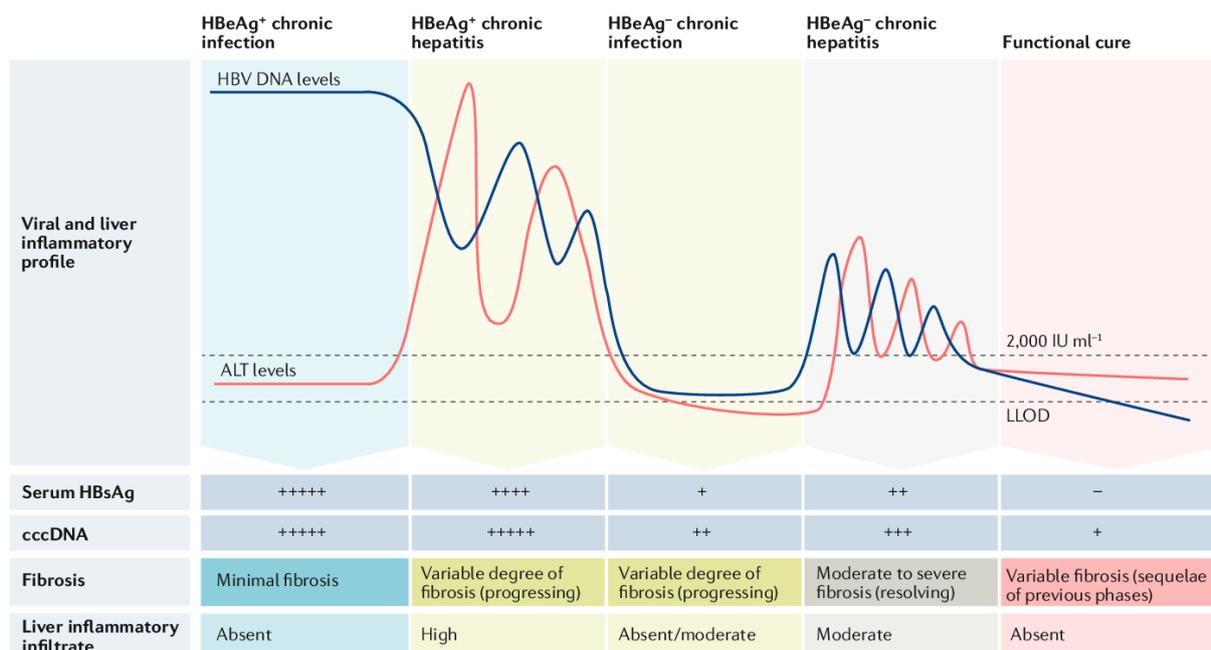


Figure 19 : Les 5 phases de l'infection B chronique (Fanning et al., 2019).

- Phase 1 : Infection chronique avec AgHBe positif, caractérisée par la présence d'AgHBe et d'une quantité d'ADN viral très élevée dans le sérum, ainsi que d'un niveau d'ALT normal. Pendant cette phase généralement asymptomatique, le virus réplique fortement et les patients sont très contagieux étant donné l'importance de leur charge virale. Il n'y a cependant pas ou peu d'inflammation et de fibrose dans le foie. Le taux de séroconversion spontanée de l'AgHBe à ce stade est très rare.
- Phase 2 : Hépatite chronique avec AgHBe positif, caractérisée par la présence d'AgHBe et d'une quantité d'ADN viral élevée dans le sang, ainsi que d'un niveau élevé d'ALT. La réponse immunitaire s'intensifie et entraîne l'apparition de dommages hépatiques (inflammation et progression de la fibrose). La plupart des patients vont alors avoir une séroconversion de l'AgHBe (10-20% par an), mais avec ou sans contrôle de la répllication virale.
- Phase 3 : Infection chronique avec AgHBe négatif, caractérisée par la présence d'anticorps anti-HBe, d'une quantité d'ADN viral faible ou indétectable, ainsi que d'un niveau d'ALT normal. Pendant cette phase, un équilibre s'installe entre la répllication virale et la réponse du système immunitaire. La répllication est donc limitée, l'inflammation et la fibrose dans le foie sont minimales, et les patients

n'ont donc qu'un risque faible de développer une cirrhose ou un CHC pendant cette phase de l'infection. Chez 1 à 3% de ces patients, l'AgHBs devient négatif spontanément, avec ou sans séroconversion.

- Phase 4 : Hépatite chronique avec AgHBe négatif, caractérisée par la présence d'anticorps anti-HBe, d'une quantité d'ADN viral modérée ou élevée, ainsi que d'un niveau d'ALT élevé (de façon persistante ou intermittente). Le système immunitaire ne permet plus le contrôle de la réplication virale, l'histologie du foie témoigne d'une atteinte hépatique modérée ou sévère avec nécro-inflammation et fibrose. Le taux de séroconversion HBs et donc de rémission spontanée des patients à ce stade est très faible.
- Phase 5 : Infection occulte (OBI – Occult B Infection), caractérisée par l'absence d'AgHBs dans le sérum et de la présence d'anticorps anti-HBc mais pas nécessairement d'anticorps anti-HBs. Le niveau d'ALT est normal, et l'ADN viral est en général indétectable dans le sérum. Cependant, de l'ADNccc reste présent dans le foie. Si l'AgHBs devient négatif avant le développement de la cirrhose, le risque pour les patients d'avoir une décompensation hépatique ou un CHC est minimal, mais le risque reste élevé pour les patients cirrhotiques. Enfin, une réactivation du virus peut être observée en cas d'immunosuppression.

Cette histoire naturelle de l'infection B chronique est déterminée par un ensemble de facteurs liés au virus (génotype, mutations ponctuelles, quantité d'ADN viral), à l'individu infecté (sexe, âge, histoire familiale) et à l'environnement (co-infection avec d'autres virus comme les virus de l'Hépatite C, de l'Hépatite D ou le VIH, expositions à des toxines comme l'Aflatoxine B1, autres sources d'inflammation du foie comme l'alcool ou les syndromes métaboliques) (Croagh, 2014).

Par ailleurs le risque d'évolution de l'infection chronique vers la cirrhose et le développement d'un carcinome hépatocellulaire (CHC) est variable et dépend fortement de la réponse immunitaire de l'individu infecté. Chez les patients atteints d'une Hépatite B chronique non traitée, l'incidence cumulée sur 5 ans de la cirrhose varie entre 8 et 20%. Chez les patients ayant déjà développé une cirrhose, le risque cumulé sur 5 ans de décompensation hépatique est de 20% et le risque annuel de développer un CHC est de 2 à 5% (Lampertico et al., 2017).

3) Prophylaxie et traitements actuels

Vaccination :

Les premiers vaccins ont été développés à la fin des années 1970 aux Etats-Unis et en France. Il s'agissait de vaccins contenant des AgHBs issus du sérum de patients ayant été infectés. Ils ont peu à peu été remplacés par des vaccins recombinants contenant les protéines d'enveloppe du virus. Une seconde génération a été mise au point dans les années 1980 à partir de la transfection de séquences codant pour la protéine S-HBs dans des levures. Ces vaccins recombinants sont désormais utilisés à l'échelle mondiale et leur efficacité est proche de 100%. Des facteurs peuvent cependant diminuer la réponse vaccinale chez certains patients (âge avancé, insuffisance rénale, co-infection par le VIH et immunosuppression). Les régions préS1/préS2 ayant aussi été décrites comme immunogènes, une troisième génération de vaccins contenant les deux protéines S-HBs et M-HBs avec ou sans L-HBs, a alors été générée et s'est avéré efficace pour ces groupes de patients (Shouval et al., 2015).

La vaccination universelle à la naissance a été recommandée par l'OMS en 1991, et elle est désormais obligatoire dans de nombreux pays. En 2015, la couverture vaccinale mondiale était de 84%. Ces politiques ont permis une réduction considérable du nombre de transmissions du VHB. En effet, depuis que l'administration d'un vaccin a été largement mise en place dans le début des années 2000, la proportion d'enfants de moins de 5 ans devenant porteurs chroniques est passée de 4,7% à 1,3%. Cependant, la couverture vaccinale mondiale reste faible parmi les nouveau-nés, elle n'était que de 39% en 2015. Ces vaccinations à la naissance sont pourtant la meilleure prévention contre la transmission du virus de la mère à l'enfant qui constitue actuellement la majorité des nouvelles infections. Cela reste notamment un problème de santé publique très important en Afrique puisqu'en 2015 la couverture vaccinale des nouveau-nés n'était que de 10% et la proportion d'enfants de moins de 5 ans devenant porteurs chroniques était encore de 3% (World Health Organization, 2017).

La vaccination est donc la meilleure arme pour lutter contre les nouvelles infections mais elle n'a aucun effet sur les personnes déjà infectées. La majorité des personnes ayant une hépatite B chronique actuellement est en effet née avant qu'un vaccin contre le virus ait été largement diffusé et administré dès l'enfance.

Marqueurs biologiques :

Un test de dépistage du VHB analyse la positivité des marqueurs sérologiques suivants : AgHBs, anticorps anti-HBc et anticorps anti-HBs. La positivité des AgHBs indique que la personne est en cours d'infection, la positivité des anticorps anti-HBc témoigne d'un contact présent ou passé avec le virus, la positivité des anticorps anti-HBs prouve que la personne est immunisée grâce à la vaccination ou suite à la résolution d'une infection passée (Davison and Strasser, 2014).

Pour les patients atteints d'Hépatite B chronique, la décision de débiter un traitement dépend principalement des trois marqueurs suivants : la quantité d'ADN viral et les niveaux d'ALT dans le sérum et la sévérité des atteintes hépatiques. Les recommandations de l'EASL indiquent qu'il faut administrer un traitement aux patients en phase d'hépatite chronique, c'est-à-dire avec un niveau d'ALT élevé ou avec des atteintes hépatiques déjà importantes (qu'ils soient AgHBe positif ou négatif, et avec ou sans cirrhose). En revanche, les indications sont plus flexibles si le patient ne présente pas de dommage au foie. Il est néanmoins recommandé de mettre sous traitement un patient s'il a une charge virale élevée et qu'il est âgé de plus de 30ans, ou s'il a des antécédents familiaux de cirrhose ou de CHC (Lampertico et al., 2017).

Traitements actuels :

L'objectif du traitement est multiple : supprimer ou réduire à un niveau indétectable la quantité d'ADN viral dans le sang, induire la perte de l'AgHBe (indicateur d'un contrôle partiel de l'infection par le système immunitaire), stopper la progression des atteintes hépatiques (normalisation du niveau des ALT) et idéalement induire la perte de l'AgHBs (indicateur de la « guérison fonctionnelle ») (Lampertico et al., 2017).

Il existe actuellement deux stratégies de traitement : par interféron- α pégylé (PegIFN α) et par analogues nucleos(t)idiques (NA), dont les principes et les avantages et inconvénients sont détaillés ci-dessous (Lampertico et al., 2017; Zoulim et al., 2016).

- L'IFN- α pégylé est une cytokine de l'immunité innée qui induit l'expression des gènes ISGs (Interferon Stimulated Genes) codant pour des protéines aux propriétés antivirales, et qui favorise la différenciation et l'activation de cellules immunitaires. Chez les patients infectés par le VHB, l'IFN- α pégylé va notamment activer les cellules NK (Natural Killer) pour éliminer les cellules infectées, inhiber la transcription de l'ADNccc

grâce à une régulation épigénétique, bloquer l'assemblage de la capsid, mais aussi directement dégrader l'ADNccc via l'activation du gène *APOBEC3A*. L'avantage principal du traitement par IFN- α pégylé est qu'il induit une réponse immunitaire à long-terme via une stimulation immunitaire innée tout en étant administré sur une durée limitée (48 semaines). Il est cependant moins bien toléré, et les réponses sont très variables en fonction des patients. Il ne peut notamment pas être utilisé chez des personnes ayant eu une décompensation hépatique, une transplantation du foie ou sous immunosuppression. La suppression virale est atteinte chez 10 à 40% des patients, et la normalisation des niveaux d'ALT chez 35 à 50% des patients. La perte de l'AgHBs est observée dans 10% des cas au bout de 3 ans après traitement.

- Les analogues nucleos(t)idiques (NA) ciblent directement le virus en inhibant l'activité de transcription inverse de la polymérase virale : ils entravent l'élongation de l'ADN viral simple-brin qui est synthétisé à partir de l'ARNpg. Les conséquences sont la diminution de la production de virions et du recyclage des capsides ainsi que théoriquement de la quantité d'ADNccc même si cette dernière affirmation n'est observée qu'après plusieurs années de traitement. Les NA n'inhibent pas cependant la formation de novo d'ADNccc dans les cellules nouvellement infectées. Ces traitements sont classés en deux catégories suivant l'émergence de mutations dans le génome du VHB qui rend le virus résistant : la lamiduvine, l'adéfovir et le telbivudine ont une barrière faible face à ces résistances alors que l'entécavir et le ténofovir ont une barrière forte. Ces derniers sont donc désormais largement recommandés, ils sont bien tolérés et ont une forte efficacité antivirale à long-terme chez une grande majorité des patients en diminuant rapidement la quantité d'ADN viral dans le sérum. La suppression virale est atteinte chez 60 à 75% des patients après 3 ans de thérapie, et la normalisation des ALT chez 75 à 80% des patients. La perte de l'AgHBs est observée dans 12% des cas après 7 ans de traitement au ténofovir.

La négativité de l'AgHBs, principalement liée à une séroconversion et donc associée à l'apparition d'anticorps anti-HBs, est appelé « Guérison fonctionnelle » et permet en général l'arrêt des traitements antiviraux même si elle n'est que rarement atteinte. L'infection n'est cependant pas éradiquée à cause de la persistance dans le noyau des cellules hépatiques de l'ADNccc et d'ADN viral intégré, qui ne sont pas ciblé par les thérapies actuelles. Les patients restent donc à risque de développer un CHC et cela sera développé dans la partie III.

4) Enjeux et nouvelles cibles thérapeutiques

L'enjeu des nouveaux traitements est à la fois de cibler le pool d'ADNccc qui persiste dans le noyau des cellules infectées pour supprimer le virus, et de parer au dysfonctionnement de la réponse immunitaire pour l'aider à supprimer les cellules infectées. Le fait qu'une « guérison complète » nécessite l'élimination du virus à l'intérieur des cellules infectées ou bien l'élimination totale de ces cellules n'est pas encore déterminé. De nouvelles cibles thérapeutiques ont été identifiées par la recherche académique et industrielle et ont fait l'objet de nombreuses publications qui vont être résumées dans les paragraphes qui suivent (Fanning et al., 2019; Martinez et al., 2019; Revill et al., 2019) et qui sont représentées sur la Figure 20 ci-dessous. Certaines molécules sont d'ores et déjà évaluées dans des modèles précliniques ou dans les premières phases du développement clinique (Hepatitis B Foundation, 2020; Martinez et al., 2020).

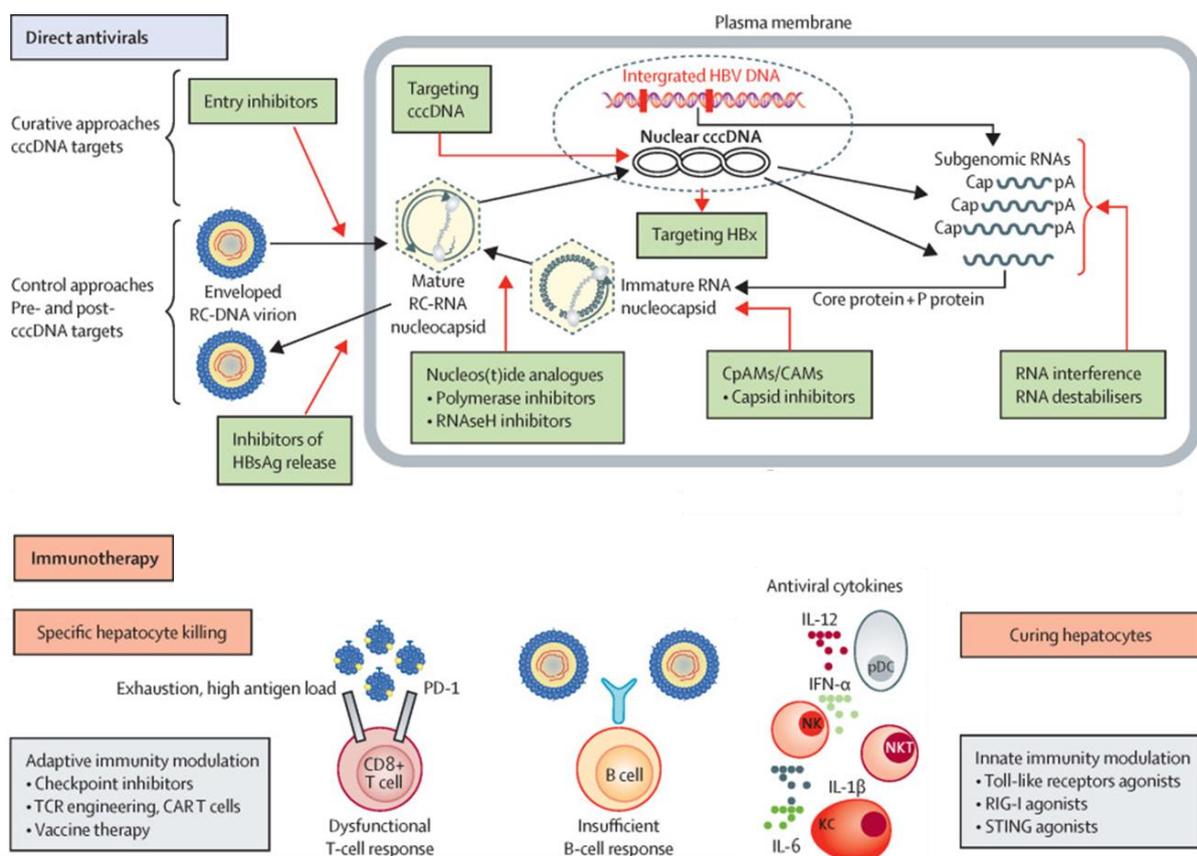


Figure 20 : Cibles thérapeutiques actuelles et futures du VHB. Les cibles antivirales sont dans la partie haute et les cibles immunologiques dans la partie basse de la figure (Revill et al., 2019).

Thérapies antivirales : La maintenance du pool d'ADNccc sous traitement peut être due à la réplication virale qui n'est pas totalement supprimée bien que la production de nouveaux virions soit suffisamment faible pour être indétectable avec les technologies de dépistage utilisées. Les particules virales nouvellement produites peuvent alors participer au renouvellement du pool d'ADNccc en infectant d'autres hépatocytes ou en étant recyclées directement à l'intérieur de la cellule. Les thérapies antivirales qui sont actuellement à l'étude suivent les stratégies suivantes :

- Inhiber l'entrée du virus dans les hépatocytes en empêchant l'interaction entre le récepteur NTCP et les antigènes HBs (phase IIb)
- Inactiver directement l'ADNccc en le clivant avec des nucléases ou en induisant des mutations ou délétions avec des endonucléases de type « CRISPR-Cas9 » (phase préclinique)
- Inhiber la transcription de l'ADNccc en restaurant l'expression du complexe Smc5/6 (via l'inhibition d'HBx – phase I) ou en modifiant la structure chromatinienne de l'ADNccc (via la modification des interactions entre l'ADNccc et les histones – phase I)
- Dégrader les transcrits viraux à l'aide d'ARN interférants (phase II), d'oligonucléotides antisens (ASOs) (phase II) ou de molécules déstabilisant les ARNs (phase I), pour limiter la production d'AgHBs synthétisés à la fois par l'ADNccc et par l'ADN viral intégré
- Inhiber la transcription inverse avec des nucléos(t)ides analogues plus efficaces et plus rapides (phase III) ou avec des inhibiteurs de l'activité RNAseH
- Inhiber l'assemblage de la capsid (phase II)
- Inhiber la production d'AgHBs avec des polymères d'acides nucléiques (NAPs) (Phase II)

Immunothérapies : Les patients contrôlant l'infection par le VHB ont une activation coordonnée de leur immunité humorale et cellulaire contre le virus, et les deux types d'immunités sont a priori nécessaires. Etant donné que le VHB a la particularité d'échapper à la reconnaissance de l'immunité innée, réinstaurer cette immunité pourrait potentiellement permettre le contrôle de l'infection. Cela est observé dans les co-infections avec les virus de l'Hépatite C ou Hépatite D : ces virus entraînent, en déclenchant une réponse immunitaire innée, une diminution de la réplication du VHB.

Par ailleurs, les approches d'immunothérapies semblent pour l'instant peu efficaces pour stimuler l'immunité contre le VHB et cela peut s'expliquer par le fait que les cellules T spécifiques au VHB existent en faible quantité et ont un phénotype d'épuisement. Néanmoins, plusieurs thérapies sont actuellement étudiées pour remédier à la défaillance immunitaire et suivent les stratégies suivantes :

- Activer la réponse immunitaire innée à l'aide de cytokines antivirales, d'agonistes des récepteurs Toll-Like (TLR) (TLR7 et TLR8 – Phase I) agissant dans les cellules immunitaires, ou d'agonistes de la protéine RIG-I agissant dans les hépatocytes
- Activer la réponse immunitaire adaptative à l'aide d'anticorps bloquant l'interaction PD1-PDL1 ou de vaccins thérapeutiques (Phase I)
- Utiliser des cellules T spécifiques au VHB modifiées (« engineered »)
- Utiliser des récepteurs aux cellules T (TCR) solubles pouvant rediriger les cellules T endogènes vers les hépatocytes infectés

L'environnement intra-hépatique des patients avec une infection B chronique (présence d'un infiltrat inflammatoire, fibrose, cirrhose) ainsi que la persistance d'antigènes viraux dans leur sérum et dans leur foie sur une durée longue, peuvent modifier leur profil immunologique et ainsi altérer l'efficacité de certains traitements. L'enjeu actuel repose donc principalement sur la mise au point de stratégies de combinaisons thérapeutiques adaptés aux patients afin d'améliorer la prise en charge.

C. Influence des facteurs viraux sur l'infection : variabilité du génome et co-infections

Comme cela a été mentionné précédemment, l'histoire naturelle de l'infection par le VHB est déterminée non seulement par les caractéristiques de l'individu infecté et de son mode de vie (sexe, âge, consommation d'alcool, syndrome métabolique...) mais également par des facteurs viraux liés au VHB ou à d'autres virus. En effet, les conséquences cliniques peuvent différer en fonction du génome du VHB à l'origine de l'infection et être impactées par la présence d'autres virus dans l'organisme en cas de co-infections par le Virus de l'Hépatite C (VHC), le Virus de l'Hépatite D (VHD) ou encore le Virus de l'Immunodéficience Humaine (VIH) (McMahon, 2009).

Par ailleurs, la diversité des génomes du VHB peut être divisée en deux catégories : une variabilité génotypique qui résulte de l'évolution progressive du génome en l'absence de pression sélective, et une variabilité phénotypique issue de l'adaptation du virus à des pressions sélectives (réponse immunitaire de l'hôte, vaccination, et traitements antiviraux). Les variants génotypiques sont sélectionnés s'il améliorent les capacités globales de réplication virale (la « fitness » du virus), alors que les variants phénotypiques ne peuvent être conservés que s'ils permettent au virus de résister face à des pressions antivirales précises (Kay and Zoulim, 2007).

1) Variabilité génotypique

La variabilité du VHB est liée au taux d'erreur associé à l'étape de transcription inverse effectuée par la polymérase virale, et par la très grande quantité de virions produits quotidiennement chez un individu infecté (jusqu'à 10^{12} particules virales par jour). La polymérase n'ayant pas d'activité 3'-5' exonucléase permettant la correction de ses erreurs, le taux de mutation du génome viral est donc estimé entre 10^{-3} et 10^{-6} par site et par an, ce qui est considérablement plus élevé que chez les autres virus à ADN (Kramvis, 2014; Lazarevic, 2014). Ces mutations ne conduisent cependant pas nécessairement à des clones pouvant se répliquer. En effet, les différents ORFs du génome viral étant superposés, une mutation sur un gène viral n'entraînant pas de conséquences directes sur la protéine qu'il code peut néanmoins altérer considérablement le gène chevauchant ou impacter des séquences régulatrices ou structurelles du génome. Au final, le VHB ne peut donc tolérer qu'une faible diversité et l'émergence de mutants reste plus lente que chez des virus à ARN ayant le même taux de mutation comme le VIH ou le VHC (Kay and Zoulim, 2007).

La première classification des souches du VHB était basée sur la définition de neuf sérotypes à partir de séquences présentes dans les déterminants antigéniques de la protéine AgHBs. Le développement du séquençage de nouvelle génération a permis d'effectuer une nouvelle classification basée sur la comparaison des séquences nucléotidiques entières des différentes souches virales. Huit principaux génotypes identifiés de A à H sont officiellement répertoriés dans les classifications taxonomiques et deux autres génotypes I et J ont été décrits plus récemment mais semblent plus

minoritaires. Ces dix génotypes ont ensuite été subdivisés en 38 sous-génotypes. Génotypes et sous-génotypes sont définis comme ayant plus de 8% et 4% de divergence respectivement dans leur séquence nucléotidique. Certains de ces sous-génotypes sont des virus recombinants issus d'une co-infection entre plusieurs génotypes du VHB. Les jonctions entre les séquences des différents génotypes se situent principalement dans la région préCore/Core (Tong and Revill, 2016). La longueur des génomes des différents génotypes varie entre 3182 et 3248 nucléotides en fonction d'insertions ou délétions présentes dans les régions préS1 et C (Figure 21). Il existe désormais plusieurs bases de données disponibles en ligne répertoriant les génotypes du VHB décrits dans la littérature et permettant des analyses phylogénétiques (McNaughton et al., 2020).

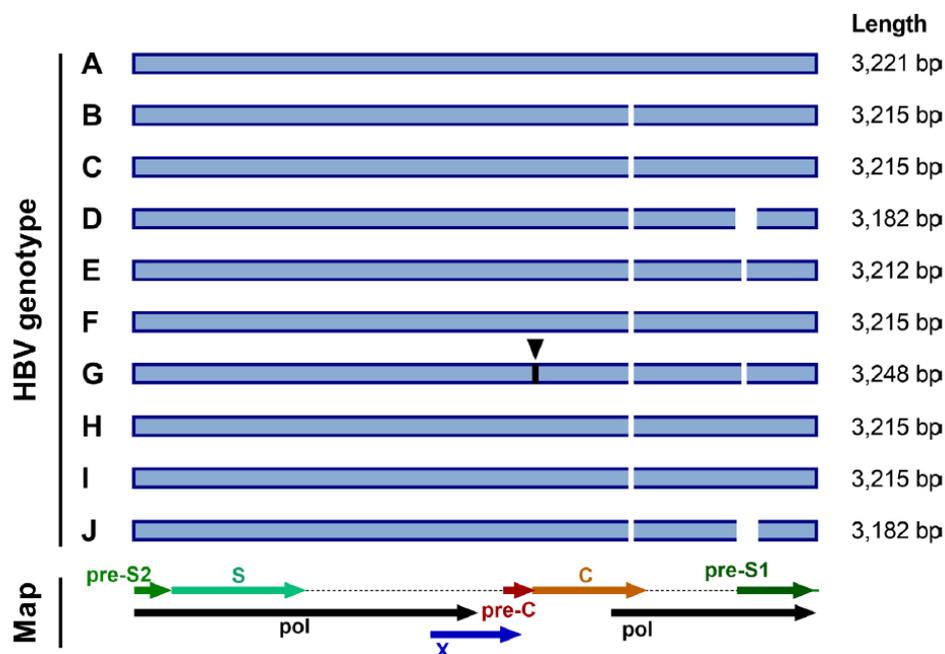


Figure 21 : Différences entre les séquences nucléotidiques des différents génotypes du VHB. (McNaughton et al., 2020).

La corrélation entre les génotypes et la répartition géographique des patients infectés a longuement été caractérisée (Figure 22). Les génotypes B et C sont majoritairement présents en Asie, le génotype E en Afrique sub-saharienne, les génotypes A et D en Europe et Amérique du Nord et les génotypes F et H en Amérique du Sud. Le génotype G a comme particularité d'être retrouvé principalement chez des hommes ayant des relations sexuelles avec d'autres hommes et d'être associé à une co-infection avec le Virus de l'Immunodéficience Humaine (VIH) (Tong and Revill, 2016).

mutations apparaissent par la suite, lorsque le système immunitaire est actif et commence à éliminer les cellules infectées. Les immunités humorales et cellulaires sont alors une pression de sélection importante qui participe fortement à l'émergence de ces mutants. Les mutations sélectionnées répondent ainsi à deux nécessités : réduire ou supprimer la production d'AgHBe et de particules subvirales afin de modifier les épitopes antigéniques des protéines virales et de permettre un échappement immunitaire, ou augmenter les capacités de réplication virale pour compenser la destruction des virions par le système immunitaire (Tong and Revill, 2016).

Les conséquences des mutations varient en fonction des régions du génome :

- Mutations de la région préCore/Core pour supprimer l'expression d'AgHBe

Un nombre croissant de patients conservent une réplication virale malgré l'absence d'AgHBe dans leur sérum. Cela s'explique par la présence de mutations situées dans la région préC ou dans le promoteur de cette région. Plusieurs sites de mutations ont été décrits dans cette partie du génome, mais les altérations les plus fréquentes sont la double mutation A1762T/G1764A et la mutation G1896A (Tong and Revill, 2016).

Les mutations A1762T et G1764A sont situées dans la région promotrice de l'ORF préC/C (dans le Promoteur Basal du Core (PBC)) et entraînent une forte diminution de l'expression de l'AgHBe ainsi qu'une augmentation de la réplication virale. Cette double mutation est associée à des atteintes hépatiques plus sévères chez les patients infectés, et favorise la progression de la maladie chronique vers une fibrose avancée et une cirrhose ainsi que le développement d'un CHC (Lazarevic, 2014).

La mutation G1896A introduit un codon Stop dans la région préCore et inhibe donc totalement la synthèse de l'AgHBe. Son impact sur le développement de la maladie hépatique demeure cependant controversé. Certaines études montrent une association avec des lésions hépatiques sévères, d'autres relient au contraire cette mutation avec une meilleure évolution clinique en l'absence de mutations du PBC. Ces différences pourraient être liées à l'ethnicité des patients (Ducancelle et al., 2016).

Les altérations du PBC et du préCore peuvent être associées entre elles, les mutations du PBC apparaissant a priori en premier, avant l'émergence des mutations du préCore pendant la séroconversion de l'AgHBe. Cette combinaison favorise par ailleurs la survenue d'hépatites fulminantes lors la phase aigüe de l'infection (Lazarevic, 2014).

- Mutations de la région préS/S pour altérer l'expression des AgHBs et échapper à la neutralisation des anticorps anti-HBs

Des substitutions dans la région immunogène de l'ORF S peuvent avoir d'importantes conséquences sur la mise en place de l'immunité contre l'infection et peuvent notamment favoriser un échappement immunitaire après vaccination ou après traitement par immunoglobulines anti-HBV. La mutation la plus décrite est la substitution G145R qui permet un échappement vaccinal. Un autre effet des mutations de la région S est que les AgHBs produits peuvent ne plus être détectés par les tests diagnostiques utilisés en clinique, et certains patients ont donc un statut AgHBs négatif tout en ayant de l'ADN viral détectable dans le foie et/ou dans le sérum. Ces mutations constituent une des explications proposées pour les infections occultes, définies comme la présence d'ADN viral répliatif en absence d'AgHBs (Lazarevic, 2014).

Par ailleurs, de larges délétions ont été fréquemment détectées dans la région préS dans des tissus de CHC. Ces délétions entraîneraient une rétention des protéines d'enveloppe, déclenchant un stress oxydatif et des dommages à l'ADN, ce qui favoriserait donc la prolifération et transformation tumorale (Tong and Revill, 2016).

- Mutations de la région P pour résister aux traitements antiviraux

Comme cela a déjà été mentionné, les traitements antiviraux majoritairement utilisés à l'heure actuelle sont des analogues nucléos(t)idiques (N) qui inhibent la polymérase virale pendant l'étape de transcription inverse. Ces traitements devant être pris pendant des durées longues, ils favorisent l'émergence de variants de résistance situés au niveau du site catalytique de la transcriptase inverse. Les mutations primaires confèrent directement la résistance et des mutations secondaires ou compensatoires augmentent la capacité de répliation des mutants (Lazarevic, 2014). La lamiduvine, l'adéfovir et le telbivudine ont des barrières génétiques faibles et nécessitent souvent un changement de traitement (Lampertico et al., 2017). La lamivudine a le moins bon profil de résistance : des variants se développent dans 23% et 80% des patients après 12 mois et 5 ans de traitement respectivement (Tong and Revill, 2016). L'entécavir et le ténofovir ont une barrière génétique plus forte et les mutations de résistance créées par ces traitements sont pour l'instant des événements rares : des mutations ont été décrites chez 1,2% des patients après 5 ans pour l'entécavir, et aucune mutation n'a été rapportée après 4 ans pour le ténofovir (Lazarevic, 2014) (Table 1).

HBV variant	Lamivudine	Telbivudine	Entecavir	Adefovir	Tenofovir
Wild-type	S	S	S	S	S
M204V	R	S	I	I	S
M204I	R	R	I	I	S
L180M + M204V	R	R	I	I	S
A181T/V	I	I	S	R	I
N236T	S	S	S	R	I
L180M + M204V/I ± I169T ± V173L ± M250V	R	R	R	S	S
L180M + M204V/I ± T184G ± S202I/G	R	R	R	S	S

Table 1 : Principaux variants de résistance du VHB face aux antiviraux actuels. Le profil des variants est indiqué par une lettre pour chaque traitement : S (sensible), I (susceptibilité intermédiaire/réduite), ou R (résistant) (Lampertico et al., 2017).

Ces mutations de la région P peuvent également impacter la région S superposée et donc favoriser un échappement immunitaire chez ces patients (Tong and Reville, 2016).

- Mutations de la région X dont l'impact n'est encore précisément déterminé

Des mutations et des délétions de la région X ont été fréquemment décrites dans les CHC, et les formes intégrées du VHB sont en général tronquées à l'extrémité 3' de l'ORF X. Ces altérations joueraient un rôle important dans le développement tumoral (Lazarevic, 2014), mais cela sera évoqué dans la partie III.B.3.

3) Co-infections virales

Des co-infections par le VHC, le VHD et le VIH sont fréquemment observées chez les patients infectés par le VHB. Tous ces virus utilisant les mêmes voies de transmission, deux infections peuvent parfois avoir lieu simultanément. Cependant, ces co-infections apparaissent généralement en deux temps, lorsqu'un patient ayant une infection chronique est infecté par un deuxième virus, déclenchant alors une « superinfection ». On parle d'effet dominant d'un virus lorsqu'il réplique plus activement et supprime la réplication du virus non-dominant. Les co-infections posent souvent des complications par rapport aux traitements administrés (Mavilia and Wu, 2018).

- Co-infection VHB/VHC

La co-infection VHB/VHC est principalement due à une superinfection du VHC chez des patients ayant une infection B chronique. Elle a une prévalence mondiale qui varie entre 1 et 15% mais qui serait probablement sous-estimée à cause de la présence

d'infections B occultes (OBI) qui ne sont pas diagnostiquées. En effet, des études estiment que jusqu'à 50% des patients ayant une infection C chronique aurait une OBI (Konstantinou and Deutsch, 2015; Mavilia and Wu, 2018).

Bien que ces deux virus aient tous les deux un tropisme hépatique, leurs cycles de répliquions sont très différents. Alors que le VHB est un virus à ADN répliquant dans le noyau, le VHC est un virus à ARN répliquant exclusivement dans le cytoplasme. Des études *in vitro* ont prouvé qu'ils peuvent donc répliquer simultanément dans une même cellule sans interférer. Cependant, les études cliniques montrent que la répliquion virale du VHB est fréquemment supprimée par le VHC (chez presque 50% des patients, contre environ 20% de cas où les deux virus répliquent activement et seulement 14% de cas où le VHB est dominant sur le VHC), même si le mécanisme exact n'est pas bien compris. Il existe probablement une compétition entre les deux virus pour utiliser la machinerie cellulaire de répliquion, et plusieurs théories ont été proposées pour expliquer la dominance du VHC sur le VHB : la protéine de capsid du VHC se liant à la polymérase virale du VHB empêcherait l'initiation de sa répliquion dans le noyau, l'expression du microARN-122 qui stimule la répliquion du VHC inhiberait simultanément celle du VHB et enfin, l'infection par le VHC entraînerait la production d'interférons qui ont un effet antiviral sur le VHB (Mavilia and Wu, 2018).

La majorité des études montrent un taux plus important de maladies hépatiques sévères (fibroses avancées, cirrhoses, décompensations hépatique) chez les patients co-infectés VHB/VHC que chez les patients mono-infectés par l'un ou de ces virus. Le risque de survenue d'un CHC est également plus élevé chez les patients co-infectés, le VHB étant dans ce cas a priori initiateur du processus néoplasique et le VHC accentuant et accélérant le développement tumoral (Konstantinou and Deutsch, 2015).

Pour ces co-infections, les recommandations de traitement dépendent du virus dominant dont la répliquion est active. Cependant, traiter les patients pour supprimer la répliquion de ce virus dominant peut entraîner la réactivation du virus non-dominant. Ainsi la réactivation du VHB est souvent observée chez des patients traités pour le VHC, même si cela arrive plus rarement avec les nouveaux traitements antiviraux à action directe (autour de 10%) et même si cela ne déclenche pas toujours de signes cliniques. Une surveillance de la répliquion du VHB doit dans tous les cas être mise en place sous thérapie anti-VHC (Mavilia and Wu, 2018).

- Co-infection VHB/VHD

5% des personnes infectés par le VHB sont co-infectés par le VHD, mais ce chiffre demeure incertain car les patients n'étant pas diagnostiqués pour le VHD dans la majorité des pays, il est probablement plus élevé (World Health Organization, 2017).

Le VHD est un petit virus à ARN dont le génome simple-brin de 1,7kb est contenu dans une nucléocapside formée par des protéines nommées antigènes HD (AgHD). Cette capside est elle-même entourée d'une enveloppe lipidique contenant les trois types de protéines de surface du VHB (Figure 23). Le VHD est donc un virus incomplet qui nécessite les AgHBs pour être pathogène. Une fois entré dans les hépatocytes de la même façon que les virions du VHB, le génome du VHD est libéré et transloqué vers le noyau où il utilise l'ARN polymérase II pour se répliquer. Cette réplication ne crée pas de formes ADN intermédiaires, mais est basée sur un mécanisme circulaire double (« Rolling-circle »). Une première étape génère un anti-génome complémentaire au génome du VHD et une deuxième étape utilise cet anti-génome pour reformer le génome initial qui sera enveloppé dans les nouveaux virions. Un ARN anti-génomique permet par ailleurs la synthèse des deux formes de l'AgHD : les protéines S-AgHD nécessaires à la réplication virale et les protéines L-AgHD qui permettent l'assemblage des virions en se liant aux protéines AgHBs (Koh et al., 2019).

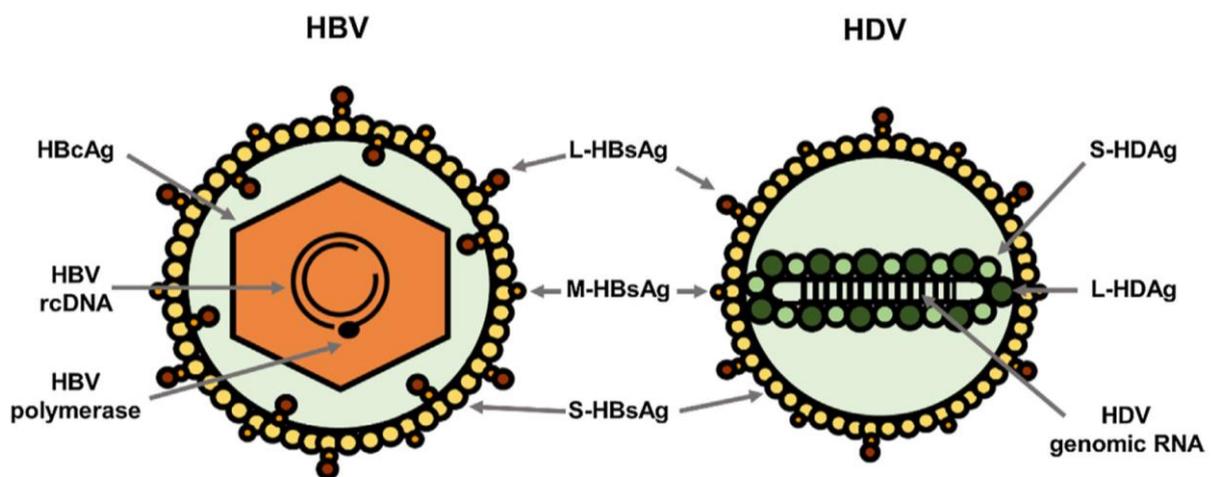


Figure 23 : Structure des virions du VHB et du VHD (Mentha et al., 2019).

L'infection par le VHD entraîne généralement une diminution de la réplication virale du VHB, même si la dominance virale semble fluctuer dans le temps chez les patients.

Les mécanismes proposés pour expliquer cette dominance du VHD reposent sur les AgHD, qui induisent une régulation épigénétique de l'ADNccc du VHB réduisant la transcription de l'ARNpg, qui répriment les séquences « Enhancer » du VHB et enfin qui diminuent la stabilité des ARNs viraux du VHB. Une autre possibilité est que l'assemblage des virions du VHD soit favorisé dans les hépatocytes où le VHB ne réplique pas mais où des AgHBs sont néanmoins produits à partir des formes intégrées du virus. Enfin, comme le VHC, le VHD induit une forte réponse immunitaire produisant des interférons ayant un effet antiviral sur le VHB (Mentha et al., 2019).

L'infection par le VHD est décrite aujourd'hui comme la forme la plus sévère et la plus rapide d'hépatite virale, bien qu'elle soit dépendante de la présence du VHB. La co-infection simultanée des deux virus déclenche une infection aiguë qui se résout dans 90-95% des cas même si le risque d'insuffisance hépatique est plus élevé que chez des patients infectés par le VHB seul. En revanche, la superinfection par le VHD d'un patient ayant une infection B chronique entraîne une double infection chronique dans plus de 90% des cas. Une progression vers la cirrhose est observée chez 10-15% de ces patients après 2 ans et chez plus de 80% des patients après 30 ans. Les observations cliniques récentes tendent à montrer que la co-infection augmente également le risque de survenue d'un CHC par rapport à une infection par le VHB seul mais le mécanisme moléculaire précis n'est pas connu (Mentha et al., 2019).

Il n'existe pas à l'heure actuelle de traitement spécifique contre le VHD. Les traitements par analogues nucléos(t)idiques n'ayant pas montré d'amélioration clinique chez les patients co-infectés VHB/VHD, les recommandations consistent à administrer un traitement par IFN- α pégylé, afin réduire la progression de la fibrose et le risque de développement d'un CHC. Ce traitement a néanmoins beaucoup d'effets indésirables et ne montre qu'une efficacité limitée. Cependant, de nouvelles thérapies à l'étude semblent prometteuses, comme notamment les inhibiteurs d'entrée utilisés également pour les infections par le VHB seul, ou des inhibiteurs spécifiques de l'assemblage ou de la sécrétion des virions du VHD (Koh et al., 2019).

- Co-infection VHB/VIH

En 2015, 2,7 millions de personnes avaient une co-infection VHB/VIH, soit 7,4% des patients infectés par le VIH, et 1% de ceux infectés par le VHB. 71% de ces personnes co-infectées vivent en Afrique sub-saharienne (World Health Organization, 2017).

Le VIH est un virus à ARN qui infecte les cellules du système immunitaire et principalement les cellules T CD4+. L'infection par le VIH se caractérise donc par une chute du nombre de ces cellules et par un déclin progressif de la réponse immunitaire : le Syndrome de l'Immunodéficience Acquisée (SIDA) (Zulfiqar et al., 2017).

L'infection par le VIH altérant considérablement l'immunité innée et adaptative, elle affecte l'histoire naturelle de l'infection par le VHB. Dans un premier temps, elle augmente le risque de développement d'une maladie chronique après la phase aiguë de l'infection par le VHB : alors qu'il est inférieur à 5% chez les adultes non infectés par le VIH, ce risque est de 25% en cas de co-infection. Concernant l'évolution de l'infection, on observe chez les patients co-infectés VHB/VIH une quantité d'ADN viral dans le sérum plus élevée, notamment chez ceux ayant de faibles taux de cellules T CD4+, ainsi que des taux de séroconversions spontanées des AgHBe et AgHBs plus faibles que chez les patients infectés par le VHB seul. De plus, si ces patients co-infectés ont déjà effectué la séroconversion de l'AgHBs, ils ont un risque plus élevé de réactivation du VHB suite à une perte des anticorps anti-HBs, alors que cet événement est très rare dans la population générale. Enfin, la co-infection accélère les dommages hépatiques provoqués par le VHB et la progression de la fibrose, et entraîne donc plus fréquemment le développement d'un CHC (Ganesan et al., 2019; Phung, 2014).

Il semble paradoxal que les dommages hépatiques dus à une réaction immunitaire contre le VHB soient exacerbés chez des patients rendus immunodéficients par le VIH. Plusieurs hypothèses ont ainsi été proposées : le VIH entraînerait l'émergence de variants spécifiques du VHB associés à une maladie hépatique plus agressive, ou alors c'est la modulation par le VIH de la réponse immunitaire anti-HBV qui pourrait altérer l'environnement cytokinique du foie et donc y créer des dommages, ou enfin, c'est l'hyperactivation immunitaire qui s'installe suite à la translocation bactérienne favorisée par l'infection par le VIH qui aggraverait la maladie hépatique (Thio, 2009). Les traitements antirétroviraux actuels suppriment efficacement la charge virale du VIH. Certains montrent aussi une activité antivirale contre le VHB, comme le ténofovir qui a en plus une forte barrière génétique comme cela a déjà été abordé. Les patients traités gardent néanmoins un risque plus élevé de développer une cirrhose et un CHC. Enfin, des études récentes suggèrent qu'une proportion importante de patients infectés par le VIH ont une infection B occulte, mais il n'existe pas de recommandations précises de traitements pour ces patients pour l'instant (Sarmati and Malagnino, 2019).

III. LA CARCINOGENESE INDUITE PAR LE VHB

Plus de 50% des CHC dans le monde sont liés à une infection par le VHB, et cette proportion atteint 85% dans les régions où le virus est endémique comme en Chine et en Afrique sub-saharienne. Ces nombres sont probablement aussi sous-estimés car certains CHC peuvent se développer dans le cadre d'infections occultes du VHB qui ne sont pas toujours identifiées. Le risque de développer un CHC est 15 à 20 fois plus élevé chez les personnes ayant une infection B chronique que pour la population non infectée. Ce risque augmente avec la progression de la maladie hépatique et parmi les patients cirrhotiques, l'incidence de développer un CHC sur 5 ans est de 15% dans les régions de forte endémicité et de 10% dans le reste du monde. Mais quelle que soit la gravité de la maladie hépatique sous-jacente, le risque de CHC chez les personnes infectées par le VHB est plus élevé dans les zones où le virus est endémique que dans les pays occidentaux, peut-être en raison d'une acquisition plus précoce de l'infection par le virus, d'une durée plus longue de la maladie ou d'une exposition à des toxines environnementales (Fattovich et al., 2004; Petruzzello, 2018).

De nombreux facteurs favorisant l'infection par le VHB ou la carcinogénèse ont déjà été évoqués dans les deux parties précédentes. L'objectif de cette partie est donc de présenter plus spécifiquement les mécanismes de carcinogénèse induits par le virus et de donner une description génomique plus précise des CHC liés au VHB.

A. Oncovirus et carcinogénèse viro-induite

Selon les chiffres de l'OMS, 15,4% des cancers dans le monde sont dus à des infections et 9,9% plus spécifiquement à des virus. Onze pathogènes ont été classifiés comme des agents carcinogéniques pour l'humain par le Centre International de Recherche sur le Cancer (CIRC). Après la bactérie intestinale *Helicobacter pylori*, ce sont quatre virus qui causent le plus grand nombre de cancers : le papillomavirus humain (VPH), le Virus de l'Hépatite B (VHB), le Virus de l'Hépatite C (VHC) et le virus Epstein-Barr (VEB). Un virus est défini comme carcinogénique s'il possède les caractéristiques suivantes (Borozan et al., 2020) :

- Présence et persistance d'ADN viral dans les biopsies tumorales
- Stimulation de la prolifération induite par les gènes viraux dans les modèles expérimentaux cellulaires ou animaux
- Corrélation du phénotype malin avec l'expression continue de gènes viraux ou avec la modification des gènes de l'hôte
- Evidences épidémiologiques que l'infection virale représente un risque majeur de développement d'un cancer

Les cancers ayant une origine infectieuse permettent en principe de mettre en place des stratégies de prévention plus facilement que pour les autres types de cancers. Il suffirait en effet de protéger l'organisme contre les agents pathogènes identifiés ou d'entraver la persistance des infections pour prévenir l'apparition de ces cancers. Les vaccins contre le VHB et le VPH ont en effet ainsi permis une diminution considérable du nombre de CHC et de cancers du col de l'utérus respectivement, dans certains pays ayant mis en place des programmes de vaccination à grande échelle. Il existe cependant aujourd'hui de fortes disparités dans l'incidence de ces cancers entre les pays développés et les pays en développement soulignant l'importance de trouver des solutions pour appliquer ces mesures préventives à l'échelle mondiale et de continuer à mettre au point de nouveaux outils pour le contrôle des maladies infectieuses et des cancers associés (De Flora and La Maestra, 2015).

1) Généralités sur les oncovirus

L'existence d'une origine infectieuse à la carcinogénèse a été établie au début des années 1900 pour les cancers animaux et en 1964 pour les cancers humains quand Epstein et Barr ont découvert des particules virales dans des lymphomes de Burkitt. Aujourd'hui, six virus sont inclus dans le groupe 1 du classement du CIRC répertoriant les agents carcinogéniques pour l'humain : les virus VPH, VHB, VHC et VEB (aussi appelé virus de l'Herpès Humain 4 (HHV-4)) précédemment mentionnés, mais également le virus de l'herpès associé au sarcome de Kaposi (aussi appelé virus de l'Herpès Humain 8 (HHV-8)) et le virus du lymphome T humain de type 1 (HTLV-1). Le Virus de l'Immunodéficience Humaine (HIV) se trouve également dans le groupe 1 de cette classification mais il n'induit la carcinogénèse que de manière indirecte en étant

un cofacteur, notamment dans les infections des virus Herpès HHV-4 et HHV-8. Enfin, trois polyomavirus se trouvent aussi dans la liste du CIRC : le polyomavirus à cellules de Merkel qui est « probablement carcinogène », et les virus JC et BK qui sont « possiblement carcinogènes » (Bouvard et al., 2009; Gaglia and Munger, 2018). Ces six virus sont présentés en Table 2 avec les types de cancer associés.

Virus	Genome	Cancer Type	Main mechanisms	Viral oncoproteins
Hepatitis B virus (HBV)	Partially double-stranded DNA hepadnavirus with retroviral features	Hepatocellular carcinoma	Chronic inflammation, oncoprotein, insertional mutagenesis	HBx (HBs)
Hepatitis C virus (HCV)	Positive-strand, single-stranded RNA flavivirus	Hepatocellular carcinoma	Chronic inflammation, oncoprotein	Core, NS3/4, NS5
Human papillomavirus (HPV)	Circular, double-stranded DNA virus	Cervical, vaginal, vulvar, penile, anal, oral carcinomas, bladder cancer	Oncoprotein, insertional mutagenesis, DNA damage	E6, E7 (E5)
Human T lymphotropic virus type 1 (HTLV-1)	Positive-strand, single-stranded RNA retrovirus	Adult T cell leukemia/lymphoma, cutaneous T-cell lymphoma	Oncoprotein, DNA damage	Tax, HBZ
Epstein-Barr virus (EBV or HHV-4)	Double-stranded DNA herpesvirus	Hodgkin's lymphoma, Burkitt's lymphoma, gastric cancer, nasopharyngeal carcinoma	Oncoprotein, DNA damage	EBNA1, EBNA3s, LMP1, LMP2
Kaposi's sarcoma-associated herpesvirus (HHV-8)	Double-stranded DNA herpesvirus	Kaposi's sarcoma, primary effusion lymphoma, multicentric Castleman's disease	Oncoprotein, DNA damage	LANA, Kaposin B, v-cyclin, vFLIP, vGPCR, K1, K15

Table 2 : Caractéristiques des six virus classés carcinogéniques pour l'humain par le CIRC (Adaptée de Chen et al., 2014; Gaglia and Munger, 2018).

Ces infections virales ne semblent cependant pas une condition suffisante pour initier la carcinogénèse, car l'incidence des cancers qu'ils provoquent est toujours très inférieure à la prévalence du virus dans la population. L'incidence du VPH par exemple est de 23% mais seulement 0,3%-1,2% des patientes infectées vont développer un cancer du col de l'utérus. Une autre illustration est le paradoxe important qui réside dans l'occurrence très rare des lymphomes de Burkitt, un cancer induit par le VEB qui infecte 95% de la population mondiale. L'infection par un virus oncogène n'implique donc pas nécessairement la survenue d'un cancer mais les personnes infectées peuvent néanmoins être considérées comme une population à risque. Les facteurs de risques communs au développement de ces cancers viro-induits sont la charge virale et la durée de l'infection (Chen et al., 2014). Par ailleurs, les études épidémiologiques montrent que ces cancers se développent plus fréquemment chez des personnes dont le système immunitaire est affaibli ou qui vivent dans de mauvaises conditions sanitaires et avec un accès limité aux soins (Gaglia and Munger, 2018).

Bien qu'ils aient des tropismes et des cycles de réplication très différents, ces virus ont comme point commun de tous mettre en place une infection persistante chez leur hôte. En effet, les cancers se développent généralement des années après l'infection qui est même souvent asymptomatique au début. Cette persistance virale nécessite de mettre en échec les réponses immunitaires innées et adaptatives pendant une durée longue et d'avoir un génome viral qui peut survivre longtemps ou se renouveler au cours du temps. Les virus Herpès sont néanmoins les seuls à avoir un mécanisme de latence, les autres virus se répliquent de façon continue dans les cellules infectées mais sans y générer d'effets cytopathiques (Gaglia and Munger, 2018). Les stratégies de réplication et de persistance du virus impliquent alors le détournement de voies de signalisation de la cellule-hôte (Mesri et al., 2014) :

- Pour créer des conditions favorables à sa réplication, le virus stimule le cycle cellulaire, reprogramme le métabolisme de la cellule et induit l'angiogenèse
- Pour assurer une réplication virale efficace et rapide, le virus recrute ou contraire inhibe les mécanismes cellulaires de réparation de dommages à l'ADN
- Pour maximiser la production de virions infectieux, le virus empêche la destruction des cellules infectées en prévenant l'apoptose et en mettant en place un échappement immunitaire
- Pour permettre la persistance de l'infection dans les cellules, le virus garantit la survie cellulaire et le maintien de la prolifération

Ces mécanismes moléculaires déployés par les oncovirus impliquent ainsi l'acquisition de caractéristiques oncogéniques. Mais l'objectif de ces stratégies consiste dans tous les cas à favoriser la persistance de l'infection et non à déclencher la transformation tumorale. La carcinogénèse viro-induite n'est donc a priori qu'une conséquence rare et accidentelle de ces infections (Gaglia and Munger, 2018; Mesri et al., 2014).

2) Mécanismes de carcinogénèse viro-induits

La carcinogénèse viro-induite repose sur trois mécanismes principaux, communs à des oncovirus infectant des cellules différentes (Table 2). Il est intéressant de comprendre les voies dérégulées dans les différents types de cancers viro-induits pour en déduire les spécificités des CHC liés au VHB qui seront détaillées par la suite.

- Expression de protéines virales oncogéniques

Les dérégulations dans les réponses cellulaires peuvent être directement causées par l'activité de protéines virales. C'est un mécanisme commun aux six oncovirus du groupe 1 de la classification du CIRC (Gaglia and Munger, 2018) (Table 2).

L'exemple le plus évident est l'action des oncoprotéines E6 et E7 du VPH. Ce virus utilise la machinerie cellulaire de la réplication de l'ADN génomique pour pouvoir répliquer son propre ADN, il a donc besoin de maintenir la cellule infectée dans un état prolifératif. La protéine virale E7 entraîne la dégradation de la protéine Rb, qui est un suppresseur de tumeur régulant le cycle cellulaire, ce qui permet de forcer l'entrée des cellules en phase S et donc la prolifération des cellules différenciées. Mais cette dérégulation déclenche l'activation de la protéine p53, un autre suppresseur de tumeur, dont le but est d'induire l'apoptose pour éviter une prolifération aberrante et non-contrôlée. La protéine E6 du virus intervient alors pour induire la dégradation de p53 et permettre ainsi la survie des cellules infectées. Cette protéine virale augmente aussi l'activité de la télomérase pour garantir la maintenance des télomères malgré la répétition des divisions cellulaires. Les deux oncoprotéines E6 et E7 ont donc une action combinée sur la régulation du cycle cellulaire qui permet l'immortalisation des cellules infectées. La protéine virale E5 joue également un rôle dans ce mécanisme en coopérant avec E6 et E7. Le virus facilite ainsi l'émergence d'une population clonale de cellules avec un fort avantage prolifératif et ayant des conditions favorables pour la transformation et la progression tumorales. Ces protéines virales participent par ailleurs à la cancérogénèse en interagissant avec d'autres voies cellulaires, en induisant une instabilité génomique et en contribuant à l'échappement immunitaire des cellules infectées (Harden and Munger, 2017; Moody and Laimins, 2010).

Un autre exemple est le virus HTLV-1 qui encode pour deux oncoprotéines intervenant à des moments différents du processus tumoral. D'un côté, la protéine Tax favorise la transformation maligne en stimulant la prolifération des cellules T, en garantissant leur survie via l'inhibition de l'apoptose et la réactivation de la télomérase, et en générant une instabilité génomique, mais elle n'est que peu exprimée une fois la tumeur formée. De l'autre côté, la protéine HBZ semble alors jouer un rôle dans la maintenance tumorale en modulant l'environnement intracellulaire des cellules infectées et en prévenant la sénescence cellulaire (Bangham and Ratner, 2015).

De nombreuses oncoprotéines des virus Herpès (VEB et HHV-8) augmentent l'expression de protéines et microARN cellulaires oncogéniques et activent notamment la voie de signalisation NF- κ B qui induit la prolifération des cellules B ou des cellules endothéliales à l'origine de ces tumeurs (Gaglia and Munger, 2018).

En ce qui concerne le VHC, bien que ces observations ne soient basées pour l'instant que sur des études en surexpression *in vitro*, certaines protéines semblent avoir des propriétés oncogéniques via l'activation de voies de signalisation cellulaires. Des études ont montré que la protéine de capsid (Core) peut notamment déréguler le cycle cellulaire, induire des changements transformationnels dans les hépatocytes et augmenter l'activité de la télomérase. De plus, des protéines non-structurales du VHC (NS3/4, NS5) actionnent différents mécanismes pour inhiber l'apoptose, accélérer la prolifération, et induire une instabilité génomique (Shirvani-Dastgerdi et al., 2016).

- Etablissement d'une inflammation chronique

La réponse inflammatoire est un élément essentiel de la défense antivirale mise en place par l'organisme pour contrer une infection. Elle permet l'induction de gènes indispensables impliqués dans l'activité antivirale, le recrutement de cellules immunitaires et la survie ou la mort cellulaire. Cependant, si l'infection persiste, l'inflammation peut devenir excessive et entraîner alors des dommages importants, notamment via un effet mutagène sur l'ADN (Read and Douglas, 2014).

Ce mécanisme d'inflammation chronique est donc un mécanisme de carcinogénèse indirect médié par les cellules immunitaires. Il joue notamment un rôle primordial dans les infections par les virus au tropisme hépatique : le VHB et le VHC. En effet, comme cela a été mentionné dans la partie I, la répétition rapide de cycles de dommages et de régénération cellulaire causés par ces virus dans le foie entraîne une fibrose hépatique de plus en plus sévère pouvant aller jusqu'à la cirrhose. Cela s'accompagne d'une accumulation de mutations oncogéniques et de modifications épigénétiques pouvant induire le développement d'un CHC (Shirvani-Dastgerdi et al., 2016). Dans les hépatites chroniques, la majorité des cytokines pro-inflammatoires est produite par l'infiltrat lymphocytaire recruté dans le foie. Ces lymphocytes ayant une faible spécificité contre les virus, ils ne parviennent pas à résoudre l'infection mais leur fort niveau d'activation participe à la création d'un milieu inflammatoire via la production des cytokines IL-1 β , TNF- α et IL-6, qui sont également sécrétées par les hépatocytes

infectés. Ces cytokines sont essentielles pour activer la prolifération et la survie des hépatocytes et donc permettre la régénération du foie, via notamment l'activation des voies NF- κ B et STAT3. Par ailleurs, ces infections déclenchent un stress du réticulum endoplasmique générant une augmentation de la quantité intracellulaire d'espèces réactives à l'oxygène (ERO), ce qui renforce l'instabilité génomique et participe à l'accumulation d'anomalies et de mutations génétiques, et joue ainsi un rôle dans l'initiation de la cancérogénèse (Fallot et al., 2012; Read and Douglas, 2014).

- Altérations du génome de la cellule hôte : instabilité, mutations, intégrations...

L'instabilité génomique peut être générée par divers facteurs : l'expression d'oncogènes viraux ou cellulaires, des changements dans le microenvironnement, une hyperprolifération forcée des cellules, ou encore la présence d'importantes quantités d'ERO ou d'autres espèces réactives toxiques. Ce mécanisme a été décrit comme pouvant être un élément-clé de la carcinogénèse, et il est souvent déclenché ou accéléré par l'action des oncoprotéines virales déjà citées (Gaglia and Munger, 2018).

Dans les infections par le virus HTLV-1 par exemple, la protéine virale Tax induit la survenue de grands réarrangements chromosomiques dans les cellules infectées en favorisant la multinucléation, les ponts de chromatine et l'aneuploïdie, entraînant ainsi le développement de leucémies à partir de cellules aux génomes instables. Tax génère par ailleurs la création de cassures double-brin pendant la réplication de l'ADN de la cellule hôte et empêche simultanément la réparation par recombinaison homologue. Ces cassures sont donc réparées par jonction d'extrémités non homologues, un mécanisme de réparation moins fiable qui génère fréquemment des mutations et des délétions. L'accumulation de ces erreurs de réparation, associée à l'augmentation de la prolifération des cellules T, augmente également considérablement le risque de transformation maligne (Bangham and Ratner, 2015; Baydoun et al., 2012).

Les virus Herpès et le VPH créent également directement des cassures double-brin dans l'ADN mais dans le but de déclencher le recrutement des ADN polymérase cellulaires et de nucléotides que les virus pourront utiliser pour leur propre réplication. De la même façon que pour HTLV-1, l'augmentation du nombre de dommages à l'ADN génomique peut cependant générer des aberrations chromosomiques importantes et notamment des translocations chromosomiques issues de cycles BSB (Breakage-Fusion-Bridge) pouvant initier la cancérogénèse (Gaglia and Munger, 2018).

Des mutations oncogéniques peuvent également être générées par les enzymes de la famille des cytidine-déaminases appelées APOBEC3 (apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3) qui font partie des mécanismes de réponse aux infections. Le but de ces enzymes est de convertir les nucléotides cytosines des séquences virales en uraciles afin de les inactiver. Cette activité antivirale peut néanmoins causer également des mutations dans l'ADN génomique de la cellule hôte qui peuvent ensuite initier un développement tumoral. En particulier, un grand nombre de cancers associés au VPH présentent une signature mutationnelle spécifique due à l'expression d'APOBEC3 (Borozan et al., 2020; Roberts et al., 2013).

Un dernier mécanisme important de la cancérogénèse viro-induite est l'intégration du génome viral dans le génome de la cellule hôte. Parmi les six virus oncogènes du groupe 1, seuls les VPH et le VHB ont cette caractéristique. Dans les tumeurs liées au papillomavirus, les séquences virales intégrées ne contiennent généralement que les gènes codant pour les oncoprotéines E6 et E7. Ces deux protéines peuvent donc être synthétisées en absence de la protéine virale E2 qui réprime normalement leur expression. De plus, les ARNs issus des formes intégrées ont été décrits comme plus stables que ceux issus des formes épisomales. L'intégration virale du VPH permet donc une forte surexpression des oncoprotéines du virus, mais elle induit également des réarrangements chromosomiques importants du génome de la cellule hôte (amplifications focales, délétions, translocations) (Harden and Munger, 2017). Une étude suggère que ces altérations sont notamment générées lors de la réplication de l'ADN à partir des origines de réplication présentes dans les séquences intégrées selon un modèle de « boucle » (« looping ») (Akagi et al., 2014). Par ailleurs, le laboratoire a récemment identifié la présence d'intégrations du virus AAV2 (Adeno-Associated Virus 2) dans 2% des CHC. Ces insertions situées à proximité d'oncogènes entraînent leur surexpression et favorisent ainsi la cancérogénèse (La Bella et al., 2020; Nault et al., 2015). Cela sera discuté plus en détails dans la partie Résultats.

Mécanismes de carcinogénèse induits par le VHB :

Les trois grands mécanismes de carcinogénèse viro-induits présentés peuvent être observés dans les CHC liés au VHB (Figure 23). En premier lieu, l'inflammation est un mécanisme indirect qui joue un rôle important dans l'hépatocarcinogénèse dans le cadre de nombreuses étiologies et notamment des hépatites virales. Cependant, alors

que les CHC liés au VHC se développent toujours sur des foies ayant un certain degré de fibrose et d'inflammation, un pourcentage important de CHC liés au VHB (10 à 30%) surviennent sans dommages hépatiques sous-jacents, soulignant ainsi l'existence de mécanismes oncogéniques directs forts induits par le VHB (Fattovich et al., 2004; Levrero and Zucman-Rossi, 2016). Un rôle oncogénique a ainsi principalement été attribué à la protéine HBx, qui a fait l'objet de nombreuses études et qui semble interagir avec un grand nombre de voies cellulaires favorisant le développement tumoral. Le deuxième mécanisme de carcinogénèse direct impliqué dans les CHC liés au VHB est l'intégration de séquences virales dans l'ADN génomique des hépatocytes infectés, pouvant altérer directement la structure ou l'expression des gènes adjacents. De plus, comme pour le VPH, ces insertions sont associées à des réarrangements chromosomiques et favoriseraient donc une instabilité génétique dans ces tumeurs qui s'additionnerait aux effets de l'activité de la protéine HBx. Enfin, les protéines d'enveloppe du VHB peuvent aussi stimuler la transformation tumorale si elles sont synthétisées en excès ou sous une forme tronquée, ce qui peut notamment être le cas lorsqu'elles sont produites à partir de formes intégrées de l'ADN viral. Afin de détailler plus précisément ces mécanismes oncogéniques, le rôle de la protéine HBx est présenté dans la partie suivante (III.A.3) et les intégrations virales dans la partie III.B.

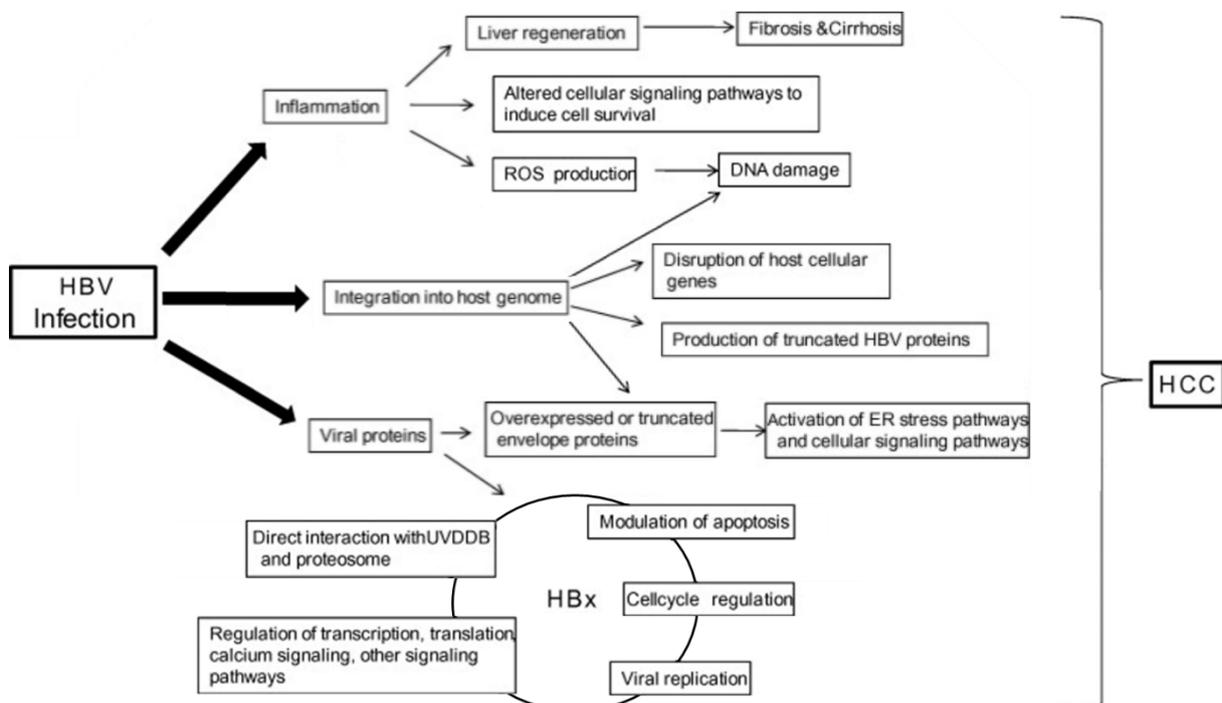


Figure 24 : Les différents mécanismes impliqués dans la carcinogénèse liée au VHB (Adaptée de Bouchard and Navas-Martin, 2011).

3) Rôle de la protéine HBx

La protéine HBx contient 154 acides aminés et elle est encodée par le plus petit ORF du génome viral du VHB. La majorité de la séquence protéique est essentielle aux fonctions de la protéine, mais la partie N-terminale semble être un élément régulateur négatif car sa délétion entraîne une augmentation des fonctions transcriptionnelles d'HBx. Cette protéine a fait l'objet de très nombreuses études et les résultats obtenus varient considérablement en fonction des modèles cellulaires et des conditions expérimentales utilisés. Une grande partie des travaux ont été effectués avec une surexpression très forte de la protéine HBx ou en l'absence des autres protéines virales, et ne reflètent donc pas nécessairement le contexte infectieux. Les fonctions de cette protéine sont donc encore très discutées mais peuvent être globalement réparties en deux groupes : le rôle d'HBx dans la réplication virale et son implication dans l'hépto-carcinogénèse (Bouchard and Navas-Martin, 2011) (Figure 25).

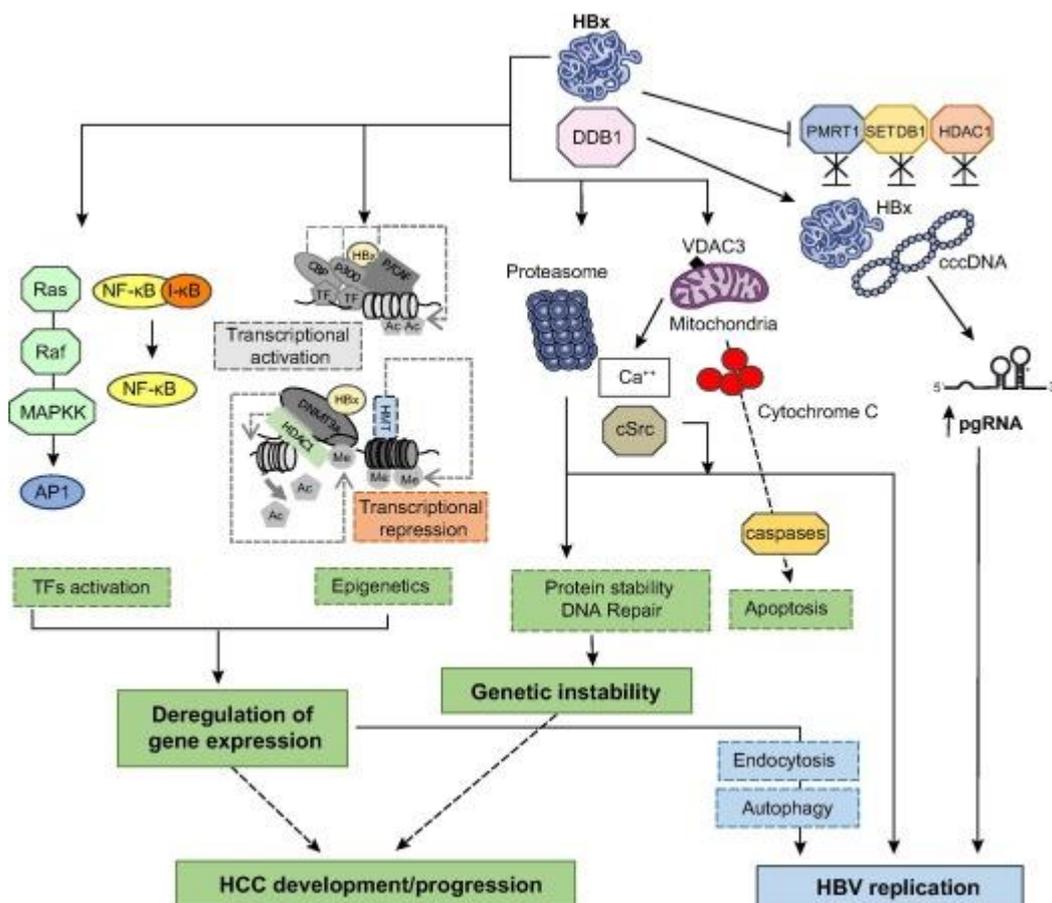


Figure 25 : Interactions de la protéine HBx et conséquences sur la réplication virale et sur l'hépto-carcinogénèse (Levrero and Zucman-Rossi, 2016).

La localisation de la protéine HBx est également débattue et semblerait au final être à la fois cytoplasmique et nucléaire. Lorsqu'elle est faiblement exprimée, HBx est majoritairement présente dans le noyau des hépatocytes et c'est l'augmentation de la quantité de protéines qui entraîne son accumulation dans le cytoplasme. Cette fraction cytoplasmique semble cependant être dégradée beaucoup plus rapidement que la fraction nucléaire, qui serait protégée par l'interaction avec la protéine cellulaire DDB1 (DNA Damage-Binding protein 1) (Benhenda et al., 2009). Dans le noyau, HBx interagit avec l'ADNccc et la machinerie cellulaire afin d'activer sa transcription. Dans le cytoplasme, HBx active divers promoteurs et « Enhancer » viraux et cellulaires pour favoriser la réplication virale, et altère notamment des facteurs affectant la survie, le métabolisme, la prolifération et la transcription (Slagle and Bouchard, 2018).

Rôle dans la réplication virale et la persistance de l'infection

Il est désormais admis que la protéine HBx est nécessaire à l'initiation et à la maintenance de la réplication virale. Bien qu'elle ne joue pas un rôle dans les étapes précoces de l'infection et notamment lors de l'entrée du virus dans les hépatocytes, elle est essentielle pour assurer la transcription de l'ADNccc. Une étude basée sur un modèle d'infection naturelle (avec des hépatocytes humains primaires et avec la lignée cellulaire HepaRG) montre en effet que l'ADNccc est inactif en absence d'HBx. Il est cependant intéressant de noter qu'HBx n'est pas impliquée dans la transcription de l'ADN viral linéaire pouvant notamment être présent sous forme intégrée, et qu'elle n'affecte pas la quantité d'ADNccc intracellulaire (Lucifora et al., 2011).

Pour stimuler la réplication virale, HBx est recrutée dans le noyau au niveau du mini-chromosome formé par l'ADNccc et induit alors des modifications épigénétiques des histones qui font partie de cette structure, tout en bloquant l'activité inhibitrice qu'ont des protéines cellulaires sur la transcription virale du VHB (Levrero and Zucman-Rossi, 2016). HBx active en parallèle la transcription de gènes impliqués dans les voies de signalisation de l'endocytose et l'autophagie, deux mécanismes dont dépend la réplication virale. De plus, l'interaction entre HBx et la protéine cellulaire DDB1 permet la dégradation du complexe de maintenance de la structure des chromosomes Smc5/Smc6, qui se lie à l'ADNccc et inhibe sa transcription. HBx supprime ainsi l'activité répressive du complexe sur l'ADNccc et permet le bon déroulement de la réplication virale (Decorsière et al., 2016).

Rôle dans la carcinogénèse hépatique

Le premier indice suggérant l'implication de la protéine HBx dans la carcinogénèse est l'absence du gène codant pour cette protéine chez les *avihépadnavirus*, les virus de la famille des *hépadnavirus* infectant les oiseaux, qui sont les seuls animaux infectés par des *hépadnavirus* à ne pas développer de tumeur hépatique. Différents modèles cellulaires et animaux ont été mis en place depuis cette observation, mais le rôle direct joué par HBx dans la transformation maligne demeure controversé. La majorité des études ont néanmoins acté que la protéine peut agir comme un cofacteur favorisant le développement tumoral. En effet, bien que la présence d'HBx seule ne montre pas d'effet oncogénique dans plusieurs modèles murins, elle peut coopérer avec d'autres agents carcinogènes. Mais HBx étant une protéine multifonctionnelle qui altère de nombreux processus cellulaires tels que la transcription des gènes, la transduction du signal intracellulaire, la prolifération, l'apoptose et la réparation de l'ADN, elle peut contribuer à la cancérogénèse via plusieurs de ses activités et il est difficile d'identifier précisément les mécanismes mis en jeu (Benhenda et al., 2009; Kew, 2011).

De façon non-exhaustive, HBx dérégule notamment les mécanismes suivants :

- Régulation de la transcription : HBx active ou réprime la transcription de nombreux gènes cellulaires via la modulation de modifications épigénétiques et l'interaction avec des facteurs de transcription, des enzymes modificatrices de la chromatine, ou des composants de la machinerie transcriptionnelle cellulaire (Levrero and Zucman-Rossi, 2016). En particulier, l'expression d'HBx dérégule les gènes codant pour les DNA méthyltransférases (DNMT) générant une hyperméthylation spécifique de gènes suppresseurs de tumeurs comme *CDKN2A* et une hypométhylation globale (Park et al., 2007; Zhu et al., 2010).
- Régulation de la sénescence : HBx contribue à la réactivation de la télomérase pour éviter la sénescence des cellules en induisant une surexpression du gène *TERT*. Par ailleurs, la sénescence médiée par les protéines p53 et Rb est inhibée par HBx qui diminue l'expression, d'une part des gènes *ASPP1* et *ASPP2* qui sont des activateurs de p53 et d'autre part, du gène *CDKN2A* par méthylation de son promoteur qui permet l'inactivation de Rb (Levrero and Zucman-Rossi, 2016). Enfin, des études montrent une interaction entre HBx et p53 entraînant une séquestration de p53 dans le cytoplasme et donc la répression de toutes ces activités transcriptionnelles (Kew, 2011).

- Régulation de l'apoptose : HBx possède à la fois des propriétés pro-apoptotiques et anti-apoptotiques, en fonction de son niveau d'expression et du contexte cellulaire. Dans les cellules infectées, HBx altère les voies de signalisation du calcium et les mitochondries, notamment via son interaction avec la protéine mitochondriale VDAC3 (Voltage-Dependent Anion-selective Channel protein 3). Cela induit une accumulation d'ions calcium Ca^{2+} dans les mitochondries, engendrant une augmentation des niveaux d'ERO et pouvant déclencher l'apoptose de la cellule. Cependant, une très forte expression d'HBx entraîne une activation de la voie NF- κ B et bloque ainsi l'apoptose médiée par les voies TNF- α , FAS, p53 et TGF- β , permettant aux hépatocytes infectés de survivre aux dommages générés par le système immunitaire (Levrero and Zucman-Rossi, 2016). De plus, la répression des activités transcriptionnelles de p53 par HBx inhibe en particulier son rôle pro-apoptotique (Kew, 2011).
- Réparation de l'ADN : HBx interagissant avec des protéines impliquées dans la réparation des dommages à l'ADN telles que DDB1 et p53, elle interfère avec certains mécanismes comme notamment la réparation par excision de nucléotides. L'expression d'HBx ne semble cependant pas augmenter le taux de mutations spontanées mais pourrait favoriser la cancérogénèse en stimulant la prolifération de cellules ayant un ADN endommagé (Benhenda et al., 2009).
- Régulation du cycle cellulaire : HBx stimule en effet la progression du cycle cellulaire en activant des voies de signalisation telles que les voies des kinases MAP-Kinase, JNK ou Src. Mais des résultats contradictoires ont été présentés et cette régulation semble en réalité dépendre de l'état de différenciation des cellules. Par ailleurs, bien que le mécanisme moléculaire soit encore inconnu, il est avéré qu'HBx induit une dérégulation du processus mitotique et entraîne l'apparition d'importants défauts génomiques (duplication de centromères, formation de fuseaux multipolaires, ségrégation de chromosomes ou encore multinucléation des cellules), participant ainsi à l'induction d'une instabilité chromosomique dans les cellules (Benhenda et al., 2009; Fallot et al., 2012).

Compte tenu de la diversité des activités d'HBx qui ont été reportées, son rôle dans la cancérogénèse n'est pas encore bien établi. De plus, une étude récente du laboratoire a observé la présence plus fréquente de mutations inactivatrices du gène *HBx* dans les CHC par rapport aux tissus non-tumoraux, suggérant l'existence d'une pression de

sélection spécifique durant la cancérogénèse qui favorise la prolifération des cellules dans lesquelles HBx est inactive. Ainsi, malgré les potentielles fonctions oncogéniques de la protéine qui ont été proposées, le gène *HBx* possède plutôt un profil de gène suppresseur de tumeur et c'est donc l'inactivation d'HBx qui jouerait potentiellement un rôle dans le développement tumoral hépatique (Amaddeo et al., 2015).

B. Intégrations du VHB et conséquences oncogéniques

L'identification d'une séquence du VHB intégrée dans l'ADN cellulaire d'un CHC humain a été décrite pour la première fois en 1980 (Brechot et al., 1980). De nombreuses études ont ensuite été réalisées pour investiguer plus en détails ces événements d'intégrations et ont permis de mettre en évidence de premiers gènes altérés par les insertions. Ces résultats étaient cependant limités par le faible nombre d'échantillons analysés, et par la technique utilisée qui identifiait les intégrations uniquement à proximité des séquences Alu. Le développement du séquençage haut-débit a alors permis depuis 2012 de cribler des séries de CHC bien plus importantes et de donner une description plus complète des intégrations (Sung et al., 2012).

La plupart des études a été effectuée sur des échantillons tumoraux de patients infectés par le VHB avec un statut AgHBs positif, et des événements d'insertions sont observés dans la quasi-totalité des tumeurs (80-90%) (Benhenda et al., 2009; Bréchet et al., 2000). Une publication plus récente montre que ce pourcentage est également très élevé (75%) parmi des échantillons tumoraux de patients ayant une infection B occulte (OBI) (Saitta et al., 2015). Ces intégrations sont donc une caractéristique très particulière des CHC liés au VHB mais les mécanismes précis sont encore peu compris alors qu'ils représentent des enjeux importants pour le traitement des patients. En effet, le risque de CHC lié à l'infection par le VHB persiste après la « guérison fonctionnelle » des patients (Yip et al., 2017). Bien que ce risque soit fortement associé à l'inflammation antivirale chronique, d'autres facteurs peuvent être à l'origine de la carcinogenèse, comme l'instabilité chromosomique due à des lésions génotoxiques antérieures, l'expression résiduelle de protéines virales oncogènes ou la présence de modifications génétiques pré-néoplasiques. Ces éléments peuvent tous être la conséquence du processus d'intégration virale du VHB (Budzinska et al., 2018a).

1) Place de l'intégration dans l'histoire naturelle de l'infection

Le mécanisme d'intégration virale observé pour le VHB étant commun à l'ensemble des virus de la famille des *hépadnavirus*, des études *in vivo* ont d'abord été effectuées chez la marmotte et le canard. Des événements d'insertions du WHV et du DHBV ont ainsi été observés quelques jours après infection à une fréquence d'un événement parmi 10^3 à 10^4 cellules. D'un côté, la cartographie précise sur le génome du DHBV des extrémités des séquences insérées a permis de mettre en évidence que ces séquences virales étaient préférentiellement issues de formes linéaires du virus (Yang and Summers, 1999). D'un autre côté, des études sur des marmottes ayant développé une infection aiguë du WHV ont montré qu'au moins 95% des hépatocytes sont infectés par le virus et que, bien que la majorité de ces cellules soit rapidement éliminée par la réponse antivirale du système immunitaire, la proportion d'hépatocytes contenant de l'ADN intégré (0,01-0,1%) demeure stable. Cela suggère que les hépatocytes nouvellement générés sont en réalité issus du pool de cellules initialement infectées et des événements d'intégrations persistent donc dans l'ADN des cellules du foie après résolution de l'infection (Summers et al., 2003). Chez les marmottes ayant une infection chronique du WHV, le taux de cellules contenant de l'ADN viral intégré est plus élevé (1-2%). La plupart de ces hépatocytes est vraisemblablement issue de l'expansion clonale de cellules ayant un avantage sélectif particulier (leur permettant d'échapper à la réponse immune cytotoxique ou de proliférer plus rapidement) qui a eu lieu au cours de l'infection chronique, et qui n'est pas nécessairement liée à la présence d'un environnement cirrhotique (Mason et al., 2005).

Plus récemment, une modélisation de l'infection par le VHB *in vitro* utilisant des hépatocytes primaires et des lignées cellulaires a permis de confirmer ces résultats pour le VHB. Des intégrations ont en effet été observées à une fréquence d'un événement parmi 10^3 cellules dans les 5 jours suivant l'infection (Tu et al., 2018). Des études effectuées aux différents stades de l'infection sur des patients infectés par le VHB ont alors analysé les intégrations présentes et ont utilisé ces données pour évaluer l'expansion clonale des hépatocytes pendant l'infection chronique. Des insertions virales ont été observés très précocement, chez des patients ne présentant pas d'atteintes hépatiques d'un point de vue histologique, et même chez des patients ayant eu une infection aiguë, confirmant ce qui a été observé dans les modèles *in vivo* et *in vitro* (Kimbi, 2005; Mason et al., 2016). Le nombre d'événements d'intégration

augmente ensuite avec la progression de la maladie hépatique, et on observe donc plus d'insertions virales dans les tissus de patients ayant une hépatite B chronique que dans ceux de patients ayant eu une hépatite B aiguë, et encore plus dans les CHC et tissus non-tumoraux adjacents (Yang et al., 2018).

Pendant le processus de séroconversion de l'AgHBe, le nombre d'hépatocytes infectés diminue lentement, soulignant l'action persistante de la réponse immunitaire antivirale. La pression de sélection exercée par cette réaction immunitaire induit donc l'expansion clonale de cellules pouvant échapper à la réponse cytotoxique dans le foie de patients non-cirrhotiques. On observe en effet des clones de taille plus importante dans le foie des patients négatif pour l'AgHBe. Cet avantage sélectif pourrait potentiellement être dû à la disparition de l'infection active dans certaines cellules qui contiendraient de l'ADN viral intégré mais non répliatif et qui ne seraient donc plus des cibles à éliminer. Mais les intégrations contenues dans ces clones ne semblent pas être directement à l'origine de leur expansion clonale via l'altération de gènes cellulaires en *cis* étant donné qu'aucun site récurrent d'insertion n'a été identifié. La possibilité que les insertions jouent un rôle via l'induction d'une instabilité chromosomique ou via la surexpression de protéines virales ne peut cependant pas être exclue (Magdalena A. Budzinska et al., 2018b; Mason et al., 2016, 2010; T. Tu et al., 2015).

Par ailleurs, l'observation d'insertions clonales plus fréquentes dans un contexte cirrhotique peut s'expliquer par les mécanismes de régénération hépatique induits par la cirrhose et par la présence éventuelle de cellules dysplasiques possédant déjà un avantage prolifératif (Mason et al., 2010). Et ces événements d'intégrations clonales sont donc aussi présents dans les CHC issus de la transformation maligne d'une cellule qui contenait initialement une insertion virale. Une étude récente de séquençage de cellule unique montre en effet l'uniformité des intégrations virales parmi des cellules tumorales distantes spatialement au sein d'un nodule, soulignant de nouveau le caractère précoce des insertions mais également leur très grande stabilité au cours du développement tumoral (Duan et al., 2018).

L'expansion clonale d'hépatocytes contenant des intégrations virales est donc observée aux différents stades de l'infection chronique, mais le rôle direct joué par l'insertion dans l'initiation tumorale reste largement débattu. La taille des clones contenant de l'ADN intégré a tendance à augmenter avec l'âge, l'activité inflammatoire,

la progression de la fibrose et de l'infection, mais ils sont principalement constitués d'hépatocytes d'apparence normale, et leur taille ou fréquence ne diffèrent pas entre les patients avec et sans CHC coexistant (Levrero and Zucman-Rossi, 2016). Etant donné la fréquence des intégrations et dans l'hypothèse où elles seraient réparties uniformément dans le génome humain, des prédictions estiment cependant que chaque région de 1kb est altérée par une séquence virale dans au moins un hépatocyte. La probabilité qu'une de ces insertions contribue donc à un moment donné à la transformation maligne est donc très élevée (Mason et al., 2016).

Le modèle privilégié actuellement est schématisé sur la figure 26 ci-dessous :

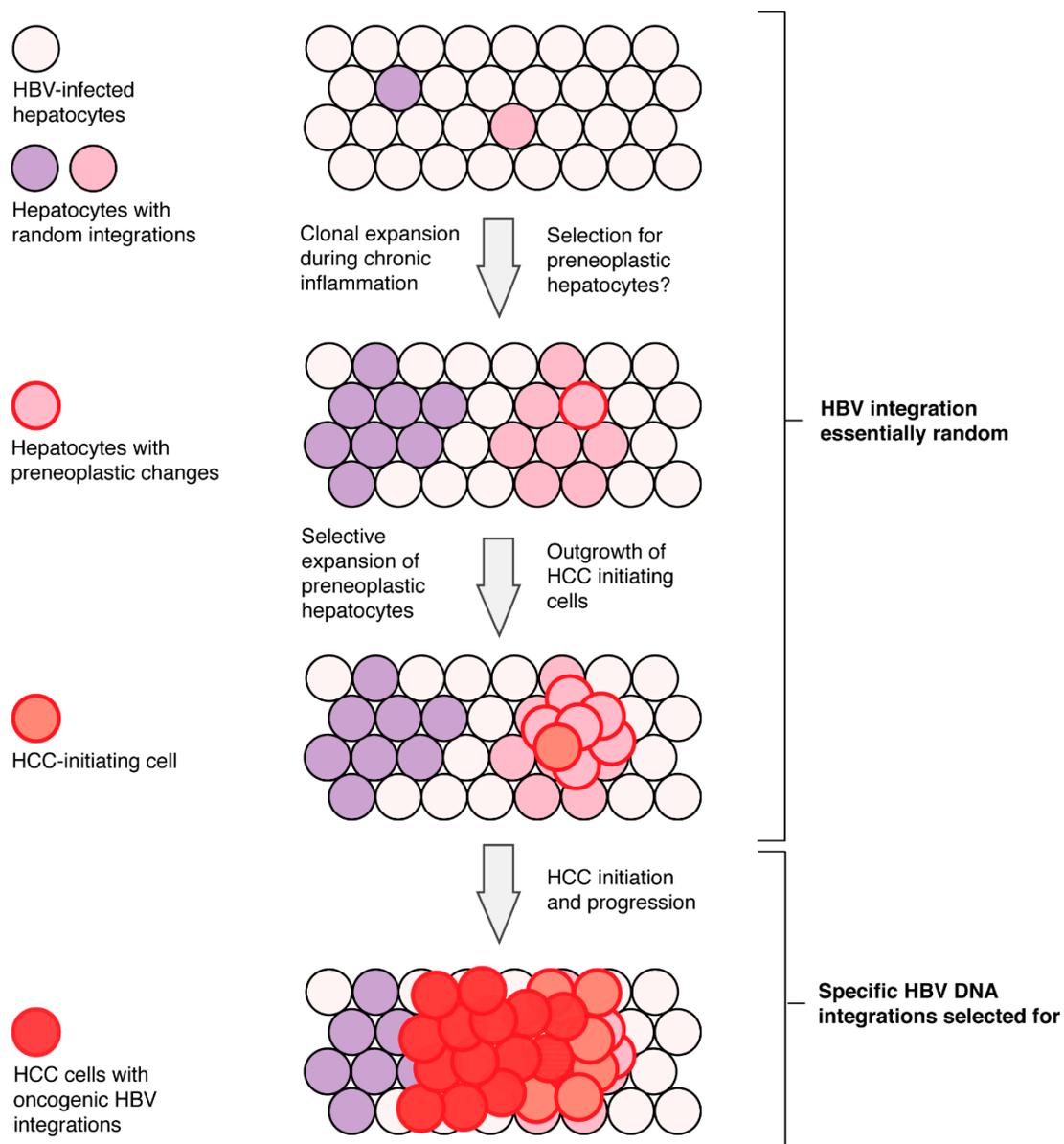


Figure 26 : Modèle du processus d'intégration de l'ADN du VHB au cours de l'hépatite chronique (Magdalena A. Budzinska et al., 2018a).

Ce modèle peut être résumé de la façon suivante : le virus s'intègre de façon aléatoire dans le génome des hépatocytes infectés. Ces hépatocytes sont alors soumis à une pression de sélection provoquée par la réponse immunitaire et l'inflammation chronique. Les cellules ayant un avantage sélectif (lié à des mutations génétiques ou des modifications épigénétiques) leur permettant d'être plus sensible aux facteurs de croissance, d'échapper à la réaction immunitaire ou de résister aux signaux pro-apoptotiques ou pro-sénescence, vont alors subir une expansion clonale. Des cellules pré-néoplasiques peuvent ainsi être sélectionnées et initier la cancérogénèse. Bien que la majorité des insertions virales contenues dans ces cellules soient donc des altérations « passager » très précoces, certaines intégrations oncogéniques vont favoriser la progression tumorale des cellules transformées et accélérer ainsi le développement d'un CHC (Magdalena A. Budzinska et al., 2018a, 2018b).

2) Caractéristiques moléculaires des intégrations

L'intégration du génome des *hépadnavirus* dans celui des cellules infectées provient principalement d'une recombinaison anormale de la forme linéaire double brin ADN_{dsI} de l'ADN viral. Comme mentionné dans la partie II.A.4, cette forme est présente dans 10% des virions et elle est issue d'une altération du processus de transcription inverse formant normalement l'ADN_{rc}. L'intégration est un mécanisme indépendant du cycle de réplication du virus et qui ne dépend pas de l'expression des protéines virales (Seeger and Mason, 2015). D'autres formes d'ADN viral et notamment les intermédiaires de réplication ou les formes épissées pourraient aussi potentiellement servir de support pour l'intégration (Tu et al., 2017). Par ailleurs, dans les modèles d'infection de courte durée d'hépatocytes primaires et lignées cellulaires, l'inhibition de l'entrée des virions dans les hépatocytes via le récepteur NTCP bloque l'intégration alors que la suppression de la réplication virale n'affecte pas ce processus, suggérant ainsi que la majorité des séquences intégrées provient de l'ADN viral contenus dans les virions primo-entrants (Tu et al., 2018). La présence d'une infection active du VHB dans les tissus favoriserait donc de nouvelles intégrations, ce qui est également montré par la détection d'un nombre significativement plus élevé d'insertions dans les tissus non-tumoraux contenant une grande quantité d'ADN_{ccc} réplcatif. Cela est aussi observé dans les tumeurs mais sans être pour autant significatif (Yoo et al., 2017).

Plusieurs publications suggèrent que le VHB s'intègre au niveau de cassures double-brin de l'ADN génomique. Une étude *in vitro* utilisant un modèle d'infection du DHBV dans une lignée cellulaire montre en effet que ce virus s'intègre préférentiellement aux sites de cassures induits de manière artificielle par une enzyme de restriction recombinante. Dans ce modèle, les cassures sont réparées par le mécanisme de jonction d'extrémités non homologues (NHEJ – Non-Homologous End Joining), ce qui induit parfois l'insertion d'une séquence virale linéaire (Bill and Summers, 2004). Pour le VHB, la majorité des intégrations semble également être générée par le mécanisme de réparation NHEJ, étant donné que les séquences virales intégrées ne partagent qu'une faible homologie avec les régions où elles s'insèrent. Cependant, les analyses *in vitro* et *in vivo* montrent que 30% des jonctions présentent une homologie de 2 à 6 nucléotides, ce qui est bien supérieur au taux de 10% qui était attendu si la localisation des insertions était totalement indépendante des séquences. Cela suggère alors l'implication du mécanisme de réparation de jonction d'extrémités médiée par micro-homologie (MMEJ – Microhomology-Mediated End Joining). Ces deux mécanismes peuvent avoir lieu tout au long du cycle cellulaire et le taux d'insertions n'est donc pas impacté par les éventuelles divisions cellulaires (Tu et al., 2018). Les études effectuées avec du séquençage haut-débit confirment l'implication du MMEJ dans une partie des événements d'intégration du VHB (Yan et al., 2015; Yoo et al., 2017; Zhao et al., 2016).

Sites d'insertions dans le génome humain

La majorité des études d'intégrations du VHB a cherché à décrire la distribution des sites d'insertions dans le génome humain dans les tumeurs (T) et dans les tissus non-tumoraux (NT). Cependant, la détection des intégrations parmi les échantillons est très variable en fonction de la sensibilité de la technique utilisée et des analyses effectuées sur le traitement des données. Une description des différentes études d'intégrations basées sur le séquençage haut-débit est présentée dans la Table 3.

Les intégrations dans les tissus non-tumoraux semblent suivre une distribution aléatoire, sans enrichissement d'une région particulière du génome (Tu et al., 2017). Par rapport aux intégrations dans les tumeurs, elles sont cependant davantage situées à proximité de gènes fortement exprimés dans le foie normal et qui sont donc dans un état transcriptionnel actif, suggérant que le VHB s'intègre plus facilement quand l'ADN est dans une configuration de chromatine ouverte (Furuta et al., 2018; Li et al., 2014).

Publication	NGS utilisé	Nombre de patients HBV+/HBV-	HBV genotypes	Nombre et pourcentage d'échantillons avec intégrations NT/T	Nombre total de jonctions NT/T	Nombre moyen de jonctions par échantillon NT/T	Principaux gènes codants ayant des intégrations récurrentes	Remarques
Sung et al., 2012	Génome entier	81 / 7	Inconnu (Chine)	27 (33%) / 76 (94%)	55 / 344	2 / 4,5	TERT (91T), MLL4 (9T), FN1 (5N), CCNE1 (4T), ROCK1 (3T), SENP5 (2T)	Enrichissement dans les promoteurs et exons dans les T et dans les introns dans les NT. Co-localisation entre sites d'intégrations et CNA dans les T
Fujimoto et al., 2012	Génome entier	11 / 25	C (100%)	0 (0%) / 8 (73%)	0 / 23	0 / 2,9	TERT (4T)	Seulement 9 insertions considérées comme clonales, toutes dans les T, et 4 à moins de 10kb d'un CNA
Jiang et al., 2012	Génome entier	3 / 1	B (33%) / C (33%) / D (33%)	3 (100%) / 3 (100%) en ADN et idem en ARN	107 / 148 en ADN / 87 / 27 en ARN	35,7 / 49,3 en ADN / 29 / 9 en ARN	TERT (6T), FN1 (4N+1T)	Enrichissement des intégrations dans les gènes codants et sur le chromosome 17
Ding et al., 2012	Séquençage ciblé	40 / 0	Inconnu (Chine)	32 (65%) / 26 (80%)	254 / 42	7,9 / 1,6	TERT (6T), GAT2A6 (2T), ADAR2 (2T)	Enrichissement dans les promoteurs et sur le chromosome 10 dans les T
Toh et al., 2013	Séquençage ciblé	48 / 0	A (6%) / B (60%) / C (35%)	13 (30%) / 22 (51%)	40 / 57	3,1 / 2,6	TERT (2T)	Enrichissement dans les promoteurs et sur le chromosome 10 dans les T
Jhunjunwala et al., 2014	Génome entier et transcriptome	8 / 4 (dont 4 de Jiang et al., 2012)	Inconnu	5 (63%) / 8 (100%) en ADN / 6 (75%) / 8 (100%) en ARN	54 / 92 en ADN / 459 / 140 en ARN	10,8 / 11,5 en ADN / 76,5 / 22 en ARN	TERT (17T), MLL4 (6T)	Enrichissement sur les chromosomes 1p, 8p, 10q, et aux centromères, co-localisation avec des CNA
Tokki et al., 2014	Séquençage ciblé	68 / 130	Inconnu (Japon)	ND / 65 (96%)	ND / 628	ND / 9,7	TERT (17T), MLL4 (6T)	
Shirashi et al., 2014	Génome entier et transcriptome	22 / 0 (dont 4 de Fujimoto et al., 2012)	Inconnu (Japon)	17 (77%) / 17 (77%)	161 / 43 en ARN	9,5 / 2,5 en ARN	TERT (2T), MLL4 (5T), FN1 (7N)	
Dong et al., 2015	Transcriptome	44T-4N / 6T-1N	B (23%) / C (77%)	2 (50%) / 16 (36%) en ARN	6 / 27 en ARN	3 / 1,7 en ARN	MLL4 (8T)	
Yan et al., 2015	Séquençage ciblé	60 / 0 (que les tumeurs)	B (47%) / C (53%)	ND / 50 (83%)	ND / 156	ND / 3,1	TERT (11T), MYC (7T), FN1 (4T), STAT1 (4T), ALB (4T), ST18 (4T), MLL4 (4T)	Nombre d'intégrations similaire entre CHC précoces et tardifs. Enrichissement dans les régions répétées, et sur le chromosome 8 pour les CHC précoces
Zhao et al., 2016	Séquençage ciblé	426 / 0	B (13%) / C (75%)	160 (38%) / 328 (77%)	739 / 3486	4,6 / 10,6	TERT (101T), MLL4 (31T), FN1 (17N), CCNA2 (8T), CCNE1 (7T)	Enrichissement dans les îlots CpG, dans les sites fragiles rares et dans les télomères dans les T, ainsi que sur les chromosomes 5, 16, 17 et 19
Fujimoto et al., 2016	Génome entier	85 / 184 (dont 27 de Fujimoto et al., 2012)	B (4%) / C (94%) / D (1%)	ND / 59 (69%)	ND / 205	ND / 3,5	TERT (9T), MLL4 (2T), MYO7A (2T)	Association entre le nombre d'intégrations et le nombre de variants structuraux
Kawanishi et al., 2016	Séquençage ciblé	27 / 77 (que les tumeurs)	B (4%) / C (96%)	ND / 25 (93%)	ND / 78	ND / 3,1	TERT (14T), MLL4 (4T), SOX5 (2T)	Parmi 21 patients avec OBI, 3 ont des intégrations (14%) avec 16 événements dont un dans TERT
Chiu et al., 2016	Transcriptome	16 / 0	B (19%) / C (81%)	13 (81%) / 11 (69%)	413 / 94	31,8 / 8,5	ALB (9N+1T), FN1 (6N), LXN (5N)	Enrichissement des jonctions virus-humain aux sites dépoussés
Yoo et al., 2017	Transcriptome	21 / 0	Inconnu	17 (81%) / 18 (86%)	424 / 133	24,9 / 7,4	FN1 (7N), PRKCE (5N), ALB (4N+2T), MLL4 (4T), TF (3T+2N)	Enrichissement des intégrations dans les introns dans les T et dans les promoteurs et introns dans les N ainsi que sur les chromosomes 1, 2 et 4 dans les N
Ally et al., 2017	Transcriptome	44 / 152	Inconnu	7 (16%) / 37 (84%)	116 / 226	16,6 / 6,1	MLL4 (5T), TERT (2T)	50% of integrations within genes
Yang et al., 2018	Séquençage ciblé	54 / 0	Inconnu	53 (98%) / 53 (98%)	4466 / 2870	84,3 / 54,2	FN1 (17N), TERT (28T), MLL4 (23T), SNTG1 (23T), PDXN1 (23T)	Moins d'intégrations observées dans des tissus de patients ayant une hépatite B chronique ou aigüe
Furuta et al., 2018	Séquençage ciblé	40 / 4 (dont 4 de Fujimoto et al., 2016)	Inconnu	40 (100%) / 35 (89%)	1424 / 258	35,6 / 7,4	FN1 (15N), TERT (8T), MLL4 (6T)	Enrichissement sur les chromosomes 2, 4, 12, 19 dans les NT et les chromosomes 5, 13, 16, 19 dans les T
Li et al., 2019	Séquençage ciblé	101 /	B (60%) / C (41%)	ND / 94 (93%)	ND / 510	ND / 5,4	TERT (28T), MLL4 (14T), CCNE1 (6T), CCNA2 (3T)	Enrichissement dans les promoteurs dans les T
Sze et al., 2020	Séquençage ciblé	95 / 0	Inconnu	ND / 85 (89%)	Inconnu	Inconnu	TERT (34T), MLL4 (11T), EYS (4T), MIR4457 (4T), CCNE1 (3T), CLPTML (3T), CCNA2 (2T), ROCK1 (2T)	
Zapalka et al., 2020	Génome entier	61 / 269 (données TCGA et ICGC)	Inconnu	ND / 36 (59%)	ND / 109	ND / 3	TERT (6T), MLL4 (4T), SEMA6D (2T)	
Li et al., 2020	Séquençage ciblé	50 / 0	B (64%) / C (24%)	ND / 44 (88%)	ND / 240	ND / 5,5	TERT (16T), MLL4 (5T), CCNA2 (3T)	

Table 3 : Etudes d'intégrations du VHB à partir de séquençage haut-débit.
T, tissu tumoral et NT, tissu non-tumoral

Dans les tumeurs, plusieurs structures ont été détectées comme étant particulièrement enrichies en sites d'intégrations virales :

- Les télomères (Zhao et al., 2016) et des longs ARNs non-codants impliqués dans la maintenance des télomères (Yang et al., 2017).
- Les îlots CpG (Zhao et al., 2016), mais l'enrichissement spécifique dans les CHC par rapport aux tissus non-tumoraux est débattu (Budzinska et al., 2018b)
- Les régions répétées telles que les Long Interspersed Nuclear Elements (LINEs) et Short Interspersed Nuclear Elements (SINEs) et les microsatellites (Ding et al., 2012; Yan et al., 2015)
- Les promoteurs et exons de gènes codants, alors que les intégrations dans les tissus non tumoraux sont plutôt enrichies dans les introns. Mais il n'y a pas d'enrichissement spécifique autour des sites de liaison des facteurs de transcription ou des sites d'initiation de la transcription. Pour les tissus tumoraux et non-tumoraux, la majorité des insertions se situe cependant dans des régions intergéniques (Sung et al., 2012; Toh et al., 2013; Zhao et al., 2016).
- Les régions fragiles et les régions denses où une augmentation du stress réplicatif peut induire une structure secondaire plus souple de l'ADN et des cassures plus fréquentes (Li et al., 2014; Yang et al., 2017; Zhao et al., 2016).

La plupart des régions mentionnées comme étant enrichies en sites d'intégrations dans les tumeurs correspond à des sites riches en GC (télomères, CpG, motifs répétées, gènes codants), mais l'hypothèse que le contexte GC favoriserait les insertions du VHB n'a pas été formellement testé (Budzinska et al., 2018a).

Points de jonctions dans l'ADN viral et séquences intégrées

Etant donné que l'ADN_{sl} est le support majoritaire de l'intégration, les points de jonctions sur le génome viral sont principalement regroupés vers les extrémités de cette forme (autour de la position 1800) (Li et al., 2019; Sung et al., 2012; Zhao et al., 2016) et la structure des séquences intégrées est nécessairement altérée. Seul l'ORF préS/S peut être intégré intact et rester sous l'action de son promoteur viral. Ce n'est en effet pas le cas des ORF P et préC/C qui, bien que pouvant être présents en entier dans les séquences intégrées, sont forcément séparés de leur promoteur. Il semblerait néanmoins que ce dernier peut être remplacé par un promoteur cellulaire car des transcrits codant pour les AgHBe et AgHBc ont déjà été décrits dans des CHC.

L'ARNpg ne peut pas être synthétisé et il ne peut donc pas y avoir de réplication virale à partir des séquences intégrées. L'ORF X est fréquemment tronqué à son extrémité finale lors de l'intégration et l'absence de codon stop peut donc générer la synthèse de protéines de fusion virus-hôte (S. Jiang et al., 2012; Tu et al., 2017) (Figure 27).

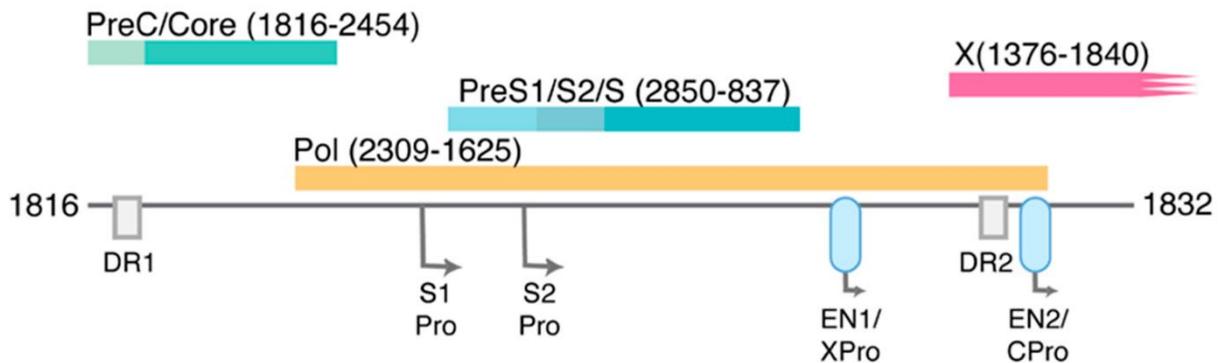


Figure 27 : Position des ORF dans la forme ADNdsI du VHB (Tu et al., 2017). Les promoteurs et Enhancer viraux sont indiqués par des flèches sur la séquence virale.

Enfin, des réarrangements complexes sont parfois observés dans les séquences intégrées, mais seulement dans une proportion limitée (autour de 10%) et pas spécifiquement dans les tissus tumoraux. Le fait que ces réarrangements aient lieu avant l'intégration (dans le cas d'insertions de génomes épissés ou déjà altérés) ou après l'intégration (en conséquence d'une instabilité chromosomique) est cependant encore inconnu (S. Jiang et al., 2012; Tu et al., 2017).

3) Conséquences oncogéniques

Etant donné que les événements d'intégration ont lieu dans les quelques jours suivant l'infection, des dizaines d'années peuvent s'écouler entre l'insertion virale et le développement d'un CHC. La fonction réelle des intégrations du VHB dans l'hépatocarcinogénèse n'est donc pas encore tout à fait claire, mais trois mécanismes ont été proposés et vont être détaillés ci-dessous.

- Mutagenèse insertionnelle

Dans les échantillons non-tumoraux, un gène semble néanmoins être ciblé de manière récurrente par les intégrations du VHB dans les différentes séries, il s'agit du gène

FN1 (Fibronectine 1). Ce gène code pour la protéine Fibronectine 1 qui fait partie de la matrice extracellulaire et joue un rôle majeur dans l'adhésion cellulaire, la motilité, la cicatrisation des tissus et le développement embryonnaire (Ding et al., 2012). Ce gène est très exprimé dans le foie, et son niveau d'expression augmente avec la fibrogenèse. Ces intégrations semblent d'ailleurs se produire préférentiellement chez des patients ayant un stade de fibrose avancé (Yoo et al., 2017). Des altérations de l'expression de FN1, de sa dégradation ou de sa structure ont été associées précédemment au cancer et à la fibrose, mais la conséquence de ces insertions dans le tissu non-tumoral reste indéterminée. L'analyse de l'expression de *FN1* donne en effet des résultats contradictoires en fonction des échantillons et des études. Les conséquences de ces insertions pourraient donc potentiellement dépendre du contexte ou de la position de l'intégration dans le gène. Le ratio d'expression entre une tumeur et le tissu non-tumoral adjacent est très variable entre les échantillons, même parmi ceux qui n'ont pas d'insertion virale dans *FN1*. Dans tous les cas, aucune différence significative d'expression n'est observée entre les échantillons avec ou sans intégration dans ce gène (Ding et al., 2012; Sung et al., 2012). Des transcrits de fusion HBV-FN1 ont par ailleurs été décrits dans des analyses de transcriptomes mais leur rôle fonctionnel n'a pas été investigué (Furuta et al., 2018; Shiraishi et al., 2014).

Dans les tumeurs, certaines intégrations induisent un changement important dans l'expression du gène. Une activation locale de la transcription est en effet observée au niveau des sites d'intégration pour la majorité des insertions situées dans des promoteurs, exons ou introns, et dans une plus faible proportion pour les insertions dans les régions intergéniques (Ally et al., 2017a; Z. Jiang et al., 2012; Toh et al., 2013). Trois gènes sont notamment identifiés de façon récurrente dans l'ensemble des études : *TERT*, *MLL4* et *CCNE1*, ils sont altérés par des intégrations dans 40 à 50% des tumeurs positives pour le VHB et induisent généralement une surexpression du gène concerné (Li et al., 2019; Sung et al., 2012).

TERT est le gène le plus fréquemment altéré par les intégrations du VHB qui se situent principalement dans le promoteur. Ces insertions sont mutuellement exclusives des autres altérations et notamment des mutations du promoteur qui ont été décrites dans la partie I.C.1 (Fujimoto et al., 2016; Sze et al., 2020; Totoki et al., 2014). Par ailleurs, la surexpression du gène induite par la séquence intégrée est d'autant plus élevée que le site d'intégration est proche du site d'initiation de la transcription, suggérant que la

séquence virale agit alors comme un « Enhancer » (Toh et al., 2013). Deux études récentes ont en effet montré à l'aide de modèles cellulaires que la région « Enhancer I » de la séquence intégrée peut induire l'activation de *TERT* via le recrutement de facteurs de transcription. La première publication a mis en évidence que la séquence virale recrute le récepteur des androgènes qui active la transcription de *TERT* ou celui des œstrogènes qui la réprime, cette régulation étant dépendante du facteur HNF4 α . Le récepteur des androgènes augmente également la transcription du gène lorsque le promoteur contient la mutation -124G>A, via la liaison entre le facteur GABPA et le promoteur muté. Ce mécanisme apporte une explication au fait que les intégrations ou mutations du promoteur de *TERT* sont retrouvées plus fréquemment et entraînent une surexpression du gène plus importante chez les hommes que chez les femmes (Li et al., 2019). La deuxième publication suggère que l'activation de *TERT* est liée au recrutement du facteur de transcription ELF4 (E74-like ETS transcription factor 4), avec un degré différent en fonction de l'orientation de la séquence virale intégrée (Sze et al., 2020). Par ailleurs, des transcrits de fusion HBx-*TERT* ou HBs-*TERT* peuvent être générés si l'insertion se situe à l'intérieur du gène. L'ensemble de ces transcrits débutent dans la séquence virale intégrée et se poursuivent dans la séquence de *TERT* jusqu'au codon stop. Si l'intégration est intronique, des sites d'épissage dans la séquence du VHB permette généralement la synthèse de ces transcrits avec ou sans rétention d'une partie de l'intron (Furuta et al., 2018; Shiraishi et al., 2014). Enfin, les patients ayant une intégration du VHB dans *TERT* ont une survie moyenne plus courte par rapport aux autres patients ayant des CHC liés au VHB (Zhao et al., 2016).

Pour ***MLL4***, les intégrations sont situées entre les exons 2 et 6. Une surexpression est observée lorsque les insertions se situent dans les exons, probablement induite par le promoteur viral d'HBx. Les transcrits générés peuvent alors contenir la séquence virale entière sans épissage. Ces transcrits HBV-*MLL4* modifiant cependant le cadre de lecture, ils entraînent probablement une perte de fonction du gène. Par ailleurs, les intégrations dans les introns induisent une expression aberrante des régions introniques et la formation de fréquents variants épissés du gène (Dong et al., 2015; Furuta et al., 2018; Shiraishi et al., 2014; Sung et al., 2012). La synthèse de protéines de fusion HBx-*MLL4* stables a également été décrite (Saigo et al., 2008).

Enfin, le laboratoire a récemment caractérisé les conséquences des insertions virales dans ***CCNE1***, et cela sera développé dans la partie Résultats (Bayard et al., 2018).

D'autres études se sont concentrées sur les conséquences fonctionnelles d'insertions dans des gènes et ont montré l'existence de transcrits de fusion récurrents. Ainsi, un transcrit particulier a été identifié comme générant la synthèse d'une protéine HBV-CCNA2 dans 12,5% des CHC d'une série de 48 échantillons. Cette protéine de fusion ne possédant pas le domaine N-terminal de la protéine CCNA2 normale qui permet sa dégradation, son accumulation engendre une progression aberrante du cycle cellulaire et une instabilité génomique (Chiu et al., 2016). Un autre exemple récent est l'identification d'un transcrit HBx-LINE1 dans 23,3% des CHC d'une série de 90 échantillons. Ce transcrit favorise le développement de tumeurs *in vivo*, et augmente la prolifération tumorale en activant la voie Wnt/ β -caténine *in vitro*. Enfin, sa détection dans les CHC a été associée à une plus faible survie des patients (Lau et al., 2014). Ces résultats n'ont cependant pas été confirmés dans d'autres séries.

- Expression des protéines virales à partir de l'ADN intégré

La présence de séquences du VHB dans le génome des hépatocytes permet la persistance de l'expression de certaines protéines virales et notamment des protéines d'enveloppe puisque l'ORF préS/S peut rester sous l'action de son promoteur. Un indice suggérant que les séquences intégrées contribuent au moins partiellement à la sécrétion d'AgHBs repose sur le fait que la quantité de transcrits de fusion identifiés dans le foie est corrélée au niveau d'AgHBs détecté dans le sérum (Yoo et al., 2017). De plus, la quantité d'ARN total détectée dans les tumeurs n'étant en revanche pas corrélée à la quantité d'ADNccc, la majorité des transcrits proviennent a priori de l'ADN intégré (Halgand et al., 2018). Il a d'ailleurs été prouvé chez des chimpanzés avec infection chronique du VHB ayant un statut AgHBe négatif que la majorité des AgHBs synthétisés étaient bien issus des séquences virales intégrées et non de l'ADNccc (Wooddell et al., 2017). Enfin, dans une expérience *in vitro*, des lignées cellulaires n'ayant ni réplique active ni production de virions détectable mais contenant des insertions virales, ont permis la réplique du VHD qui nécessite les protéines d'enveloppe du VHB (décrit dans la partie II.C.3). Cette étude a ainsi démontré qu'une séquence intégrée du VHB peut produire des protéines L-HBs et S-HBs fonctionnelles permettant la formation de virions infectieux du VHD (Freitas et al., 2014).

Cependant, des délétions de la région préS ont été identifiées dans les CHC et suppriment la partie C-terminale du domaine préS1 ou N-terminale du domaine préS2.

In vitro, l'expression de protéines d'enveloppe altérées induit un stress du réticulum endoplasmique, un stress oxydatif et des dommages à l'ADN, aboutissant à la progression du cycle cellulaire et à la transformation maligne. Ces formes mutantes peuvent également donner un avantage prolifératif à la cellule et donc stimuler l'expansion des hépatocytes (Su et al., 2014; Tong and Reville, 2016).

Par ailleurs, l'ORF X est fréquemment altéré dans les séquences intégrées, et étant donné la structure de l'ADN_{sl} il est notamment tronqué à son extrémité C-terminale (Figure 27). Une séquence tronquée d'HBx est en effet identifiée dans au moins la moitié des CHC pour lequel ce gène est détectable (Sze et al., 2020) et cela aboutit à la synthèse d'un transcrit tronqué ou d'un transcrit de fusion. Ces modifications sont observées dans les tumeurs et dans le tissu non-tumoral mais dans des proportions différentes. Une étude décrit des transcrits de fusion dans 45,5% des T et 55,9% des NT mais des transcrits tronqués dans 33,3% des T et 11,9% des NT. Dans les transcrits altérés, les régions permettant l'interaction avec DDB1 et l'activité de transactivation sont généralement conservés, mais pas celle permettant l'interaction avec p53 (Toh et al., 2013). Les transcrits tronqués sont parfois détectés à un pourcentage très élevé dans les CHC (dans 46% ou 79,3% des cas). La surexpression d'une protéine HBx tronquée à son extrémité C-terminale a été associée dans des modèles *in vitro* et *in vivo* à des propriétés de cellules progénitrices, à la transformation maligne et l'inhibition de l'apoptose et à l'invasion tumorale (Ma et al., 2008; Ng et al., 2016; Sze et al., 2013). Ces phénotypes peuvent cependant être dus à une surexpression de la protéine bien supérieure au niveau d'expression qu'elle a sous l'activité de son promoteur natif. De plus, comme cela a été mentionné dans la partie III.A.3, le rôle joué par la protéine HBx normale dans la cancérogénèse n'étant lui-même pas encore bien établi, il est difficile de conclure sur le potentiel oncogénique de la protéine tronquée sans d'autres modèles d'infection (Slagle et al., 2015).

De plus, bien que les intégrations ne semblent pas avoir de rôle dans la réplication du virus, le fait qu'elles aient lieu précocement laisse penser qu'elles pourraient être impliquées dans la modulation de la réplication ou dans la persistance de l'infection. L'ADN intégré constituant un réservoir stable et persistant de l'expression des protéines virales les formes intégrées du VHB pourraient en effet favoriser la transcription de l'ADN_{ccc} via une surexpression d'HBx et moduler la réponse immunitaire via une surexpression des AgHBs. (Tu et al., 2017).

- Intégrations et instabilité chromosomique

L'enrichissement des insertions à proximité des sites fragiles, des régions répétées, des îlots CpG et des télomères, dans les tumeurs et non dans les tissus non-tumoraux suggère que les hépatocytes portant des altérations dans ces régions sont sélectionnés pendant la cancérogénèse et que ces intégrations favorisent donc le développement du CHC en induisant une instabilité génomique ou en modifiant les mécanismes de maintenance de la stabilité des chromosomes (Zhao et al., 2016).

De plus, une partie significative des sites d'intégrations se situent à proximité de changements du nombre de copies de l'ADN génomique (CNA) (Sung et al., 2012; Totoki et al., 2014; Zapatka et al., 2020). Une étude plus précise d'une de ces intégrations a rapporté que la cassure de l'ADN induisant une large perte de copie d'un chromosome coïncide précisément avec le site de l'intégration virale. Cette perte chromosomique entraîne une diminution de l'expression des gènes situés dans la région perdue, et notamment des gènes codant pour des caspases qui jouent un rôle central dans les mécanismes d'apoptose et pourrait donc favoriser le développement tumoral (Z. Jiang et al., 2012). Le nombre d'intégrations est par ailleurs associé au nombre de variants structuraux dans une tumeur donnée (Fujimoto et al., 2016), et deux hypothèses ont été proposées pour expliquer cette association. Premièrement, l'instabilité chromosomique induite par un facteur extérieur peut favoriser le mécanisme d'intégration du virus. En effet, les insertions du VHB se produisant préférentiellement à des sites de cassures double-brins, des échantillons ayant un plus grand nombre de variants structuraux ont nécessairement plus d'opportunités d'acquérir des insertions virales (Fujimoto et al., 2016). Deuxièmement, l'intégration virale peut être « driver » et à l'origine de l'instabilité génomique induisant des variants structuraux. C'est notamment le cas lorsque l'intégration altère les gènes *CCNA2* et *CCNE1* comme cela sera discuté dans la partie Résultats (Bayard et al., 2018).

C. Spécificités des CHC liés au VHB

La vaccination universelle contre le VHB à la naissance est l'élément de prévention principal qui devrait permettre une diminution considérable de l'incidence du CHC liés au VHB dans les années qui viennent. Cela n'aura cependant que peu d'impact à court

terme étant donné que la plupart des tumeurs se produisent après 50 ans. Pour les personnes ayant déjà une infection chronique du VHB, la prévention réside alors dans la prise d'un traitement antiviral efficace qui réduit le risque de développement tumoral bien qu'il ne l'élimine pas. A l'heure actuelle, la caractérisation des CHC liés au VHB a donc pour enjeu de permettre une stratification précise des risques, d'aider à ajuster les thérapies antivirales, de définir une stratégie de surveillance, de déterminer le traitement adapté contre le CHC s'il survient, et enfin d'anticiper une éventuelle récurrence (Voulgaris et al., 2020; Xu et al., 2017).

L'objectif de cette dernière partie est de donner un aperçu global des facteurs de risques et des spécificités cliniques et génomiques qui sont caractéristiques des CHC liés au VHB ainsi que des conséquences que cela implique pour le suivi et le traitement des patients. Il est important de souligner que ce virus peut également entraîner des altérations épigénétiques pouvant jouer un rôle précoce dans l'hépatocarcinogénèse. Le VHB peut notamment induire une méthylation anormale de l'ADN dérégulant l'expression d'oncogènes ou de gènes suppresseurs de tumeurs, et provoquer des modifications post-transcriptionnelles des histones modulant le profil d'expression global des hépatocytes (Levrero and Zucman-Rossi, 2016; Tian and Ou, 2015). Cet aspect épigénétique ne sera cependant pas développé dans cette thèse.

1) Bilan des facteurs de risque et scores prédictifs

La plupart des facteurs de risque ont déjà été mentionnés dans les parties précédentes mais l'ensemble est résumé dans la Table 4.

Viral Factors	Host factors	Environmental factors
<ul style="list-style-type: none"> - HBV viral load - HBeAg positivity - HBV genotype (C>B, D>A) - HBV mutants (BCP mutations, preS deletions) 	<ul style="list-style-type: none"> - Male sex - Advanced age - Liver cirrhosis - Ethnicity (African/Asian) - Predisposing SNP (<i>STAT4</i>) - Family history 	<ul style="list-style-type: none"> - Co-infection with HCV, HDV, HIV - Alcoholic or non-alcoholic liver disease - Cigarette smoking - Aflatoxin or acid aristolochic exposure

Table 4 : Liste des facteurs de risque associés au développement d'un CHC chez les patients infectés par le VHB (Fallot et al., 2012; Kao et al., 2010).

- Premièrement, parmi les facteurs viraux, l'existence de génotypes et de mutants associés à un risque plus élevé de développer un CHC a été discutée dans la partie II.C. Mais un autre facteur de risque important est la charge virale, représentative d'une réplication intense et persistante du VHB. En effet, le niveau d'ADN viral présent dans le sérum semble prédire fortement la progression de la maladie hépatique et le développement d'un CHC, de façon indépendante de la positivité de l'AgHBe, du niveau d'ALT et de la présence d'une cirrhose. La charge virale est ainsi corrélée à l'incidence du CHC de manière dose-dépendante et peut augmenter de 10 fois le risque pour les patients. Les patients ayant conservé l'expression de l'AgHBe ont également un risque plus important par rapport à ceux ayant eu une séroconversion. La quantité importante d'ADN viral et la positivité de l'AgHBe reflète directement la réplication active du virus qui augmente notamment la fréquence des intégrations virales, les atteintes hépatiques amenant vers la cirrhose, l'instabilité chromosomique, les dommages à l'ADN, l'inflammation et le stress oxydatif (Xu et al., 2017).

- Ensuite, parmi les caractéristiques propres à l'individu, le risque de survenue d'un CHC est plus élevé chez les patients âgés (ou ayant eu une durée d'infection importante), de sexe masculin et cirrhotiques. La différence d'incidence du CHC entre les hommes et les femmes est une caractéristique générale des CHC mais elle est encore plus marquée parmi les CHC liés aux VHB. Des études ont ainsi montré que la séquence « Enhancer I » du VHB recrute des facteurs de transcription permettant aux androgènes d'activer la transcription du virus et aux œstrogènes de l'inhiber. Ce mécanisme pourrait donc expliquer l'existence d'un risque plus élevé de développer un CHC chez les hommes (Li et al., 2019). Par ailleurs, les personnes d'origine africaine et asiatique sont aussi plus susceptibles de développer un CHC liés au VHB bien qu'il ne soit pas encore clair si la différence avec les patients d'origine caucasiennes est due à des facteurs génétiques ou à un mode de transmission différent (Xu et al., 2017). En ce qui concerne les facteurs génétiques, plusieurs SNPs ont été associés à un risque plus élevé de développer un CHC parmi les patients infectés par le VHB mais comme cela a été décrit dans la partie I.B.1, seul le SNP *STAT4* rs7574865 a pour l'instant été validé (Levrero and Zucman-Rossi, 2016; J. Yang et al., 2019b). Enfin, l'histoire familiale est importante car les parents au premier degré de patients ayant un CHC lié au VHB ont un risque accru de développer également une tumeur hépatique (Kao et al., 2010).

- En dernier lieu, parmi les facteurs environnementaux, les co-infections avec le VHC, le VHD et le VIH sont associés à un risque plus élevé de développer un CHC (comme décrit dans la partie II.C.3). La plupart des facteurs étiologiques connus agissent comme des cofacteurs dans le développement tumoral. C'est notamment le cas de la consommation d'alcool qui joue un rôle synergique avec l'infection par le VHB et des syndromes métaboliques (diabète et obésité) qui est un facteur de risque indépendant. Ces deux derniers facteurs augmentent de plus de 2 fois et 7 fois respectivement le risque de survenue d'un CHC. Le tabac et le risque de CHC sont également associés de façon quantitative (Levrero and Zucman-Rossi, 2016; Xu et al., 2017). Enfin, deux derniers facteurs de risque importants liés à l'infection par le VHB sont l'exposition à l'Aflatoxine B1 (AFB1) et à l'Acide Aristolochique (AA).

Les incidences les plus élevées de CHC et les patients atteints les plus jeunes sont retrouvés en Chine et en Afrique sub-saharienne, soit dans les régions où le VHB est fortement endémique et où le taux d'exposition à l'AFB1 est très élevé. De nombreuses études ont donc cherché à identifier une association entre ces deux facteurs de risque. Des résultats contradictoires ont été publiés montrant la présence ou l'absence d'une synergie mais sont réconciliés par la proposition selon laquelle l'AFB1 est un facteur de risque indépendant lorsqu'on y est très fortement exposé mais qui nécessite le VHB comme cofacteur lorsque l'exposition est faible ou modérée (Kew, 2003). D'un point de vue moléculaire, l'AFB1 est oxydé à l'intérieur des cellules par les enzymes CYP450 pour former notamment un métabolite particulièrement instable qui réagit avec l'ADN et produit des adduits sur les résidus guanines. La majorité de ces adduits sont réparés par le mécanisme d'excision de nucléotide mais les lésions de l'ADN restantes peuvent générer des mutations, et majoritairement des transversions G-T. La substitution G>C sur le gène *TP53* induisant la mutation R249S sur la protéine est identifiée dans la quasi-totalité des CHC dans les régions où l'AFB1 est présente, et est donc considérée comme une caractéristique de l'exposition à cette toxine. Sa forte incidence résulterait à la fois d'une affinité des métabolites de l'AFB1 pour ce site spécifique et d'une sélection biologique de cette mutation au cours de la cancérogénèse via l'inactivation de la voie p53 qui contrôle le cycle cellulaire. Plusieurs mécanismes ont été suggérés pour expliquer comment le VHB pourrait augmenter l'effet mutagène de l'AFB1 : le virus ou l'infection chronique augmenterait directement l'activité des enzymes CYP450, le stress oxydatif généré par l'infection chronique contribuerait également à

la génération de mutations, la pression de sélection liée à l'infection favoriserait la prolifération clonale des cellules contenant ces mutations, et enfin la protéine HBx dérégulant la réparation de l'ADN par excision de nucléotide pourrait empêcher la réparation des adduits à l'ADN (Gouas et al., 2009). Par ailleurs, une caractérisation génomique récente des CHC liés à l'AFB1 a identifié des mutations fréquentes du gène *ADGRB1* (adhesion G protein-coupled receptor B1) en plus de celles décrites dans *TP53*. Ces CHC contiennent davantage de néo-antigènes et possèdent une forte infiltration lymphocytaire et une expression élevée de PD-L1 (Zhang et al., 2017).

Le processus mutationnel lié à l'acide aristolochique a été bien moins étudié mais de la même façon que pour l'AFB1, c'est l'interaction entre des métabolites de cet acide et l'ADN qui génère la formation d'adduits sur des résidus adénine et entraîne ainsi l'apparition de mutations, et majoritairement des transversions A-T. Une étude épidémiologique récente menée à Taïwan a évalué que, l'AA ayant été prescrit comme traitement contre l'hépatite B chronique jusqu'en 2003, 59,4% des patients infectés par le VHB y ont été exposés. Ces travaux ont aussi mis en évidence une association dose-dépendante entre l'exposition à l'AA et le développement du CHC chez les patients infectés par le VHB. Cela suggère donc qu'il y aurait une coopération entre le VHB et l'acide aristolochique favorisant la cancérogénèse comme cela a été précédemment démontré pour l'AFB1 (Chen et al., 2018; Nault and Letouzé, 2019).

Scores prédictifs

Compte tenu de la forte association existant entre l'infection chronique par le VHB et la survenue d'un CHC, une surveillance des patients atteints d'une infection chronique est généralement recommandée mais elle est néanmoins modulée en fonction d'une estimation de leur risque présumé de développer un CHC et des analyses coût-efficacité du dépistage propres à chaque contexte. Pour faciliter la mise en place de cette surveillance personnalisée, plusieurs études récentes ont donc cherché à développer et valider des scores prédictifs du risque de survenue d'un CHC parmi ces patients. A partir d'analyses univariées et multivariées, le poids des différents facteurs de risque est déterminé dans une cohorte de dérivation, permettant ainsi de construire un score, dont la précision est ensuite évaluée dans une cohorte indépendante de validation. Cependant, étant donné l'hétérogénéité des patients atteints d'une infection B chronique et la multiplication des scores prédictifs proposés, il est souvent compliqué

d'en choisir un pour la pratique clinique. La totalité des scores ont été créés à partir de cohortes de patients asiatiques à l'exception du score PAGE-B (Voulgaris et al., 2020).

Ce score PAGE-B a été proposé en 2017 pour évaluer le risque de développement d'un CHC parmi des patients d'origine caucasienne ayant une infection chronique du VHB et traités avec des antiviraux oraux (entécavir/ténofovir). Il est très simple à utiliser en routine en clinique car il est uniquement basé sur l'âge, le sexe et le niveau de plaquettes des patients. L'ajout d'une information sur la cirrhose dans le score n'améliore pas sensiblement sa prévisibilité étant donné que le niveau de plaquettes est probablement déjà représentatif de la sévérité de la maladie hépatique. Les patients appartenant au groupe à faible risque (score inférieur à 10) qui représentent entre 11 et 25% des patients ne nécessitent pas a priori d'avoir une surveillance pour un CHC, à l'exception des patients cirrhotiques (4% de ce groupe). Ce score étant basé sur des paramètres qui ne sont généralement pas affectés par les traitements antiviraux, il pourrait être potentiellement utilisé pour des patients non-traités (Lampertico et al., 2017; Papatheodoridis et al., 2016). De plus, une étude de 2018 a validé ce score dans une cohorte asiatique et a proposé une version modifiée (score mPAGE-B) possédant une meilleure performance prédictive grâce à l'ajout des données sur le niveau d'albumine dans le sérum des patients (Kim et al., 2018). Les scores PAGE-B et mPAGE-B semblent donc représenter des scores simples qui peuvent être utilisés en pratique clinique pour identifier les patients atteints d'infection B chronique, non cirrhotiques et traités avec des antiviraux oraux, qui nécessitent une surveillance du CHC, tous les patients cirrhotiques devant avoir une surveillance dans tous les cas (Galle et al., 2018; Voulgaris et al., 2020). Concernant les risques de récurrence des CHC liés au VHB, différents facteurs de risque ont également été proposés. Le risque de récurrence précoce issue d'une dissémination de la tumeur primaire se développant dans les 2 ans après la chirurgie, est associé à des caractéristiques de la tumeur (invasion micro-vasculaire, marge chirurgicale positive, niveau d'AFP élevé). En revanche, le risque de récurrence tardive via le développement d'une tumeur *de novo* est associé à l'état inflammatoire du foie et à la quantité d'ADN viral (Hoshida, 2009; Wu et al., 2009).

Les données génomiques ne sont jusqu'à présent que très peu utilisées en clinique pour la surveillance et le suivi des CHC liés au VHB. Ces tumeurs ont pourtant des caractéristiques particulières qui vont être présentées dans la partie suivante.

2) Caractéristiques génomiques

Une étude du laboratoire a comparé une série de 86 CHC liés au VHB à 90 CHC dus à d'autres étiologies (40% liés à l'alcool, 23% au VHC, 7% à une hémochromatose et 4% à une NASH). D'un point de vue clinique, les CHC liés au VHB se développent chez des patients significativement plus jeunes qui sont en général d'origine africaine ou asiatique. Ils ont des niveaux d'AFP plus élevés et les tumeurs sont généralement plus petites et moins différenciées (Amaddeo et al., 2015).

Caractérisation virale :

La comparaison entre les quantités d'ADN viral total et d'ADNccc présentes dans le tissu tumoral et dans le tissu non-tumoral adjacent a donné des résultats opposés en fonction des études. Ces divergences peuvent être dues à la région du génome viral ciblé par la technique de détection ainsi qu'aux profils virologiques et au degré d'atteintes hépatiques des patients étudiés. Il existe dans tous les cas une coexistence de formes intégrées et de formes épisomales dans les tissus hépatiques tumoraux et non-tumoraux (Amaddeo et al., 2015; Bai et al., 2013; Halgand et al., 2018).

Une étude récente effectuée sur des patients traités ayant une virémie faible ou indétectable montre que l'ADNccc ne représente qu'un faible pourcentage de l'ADN total chez ces patients. La présence d'ADNccc dans les tumeurs suggère néanmoins qu'une réplication virale peut exister malgré le fait que les formes intégrées qui sont majoritaires ne permettent pas la transcription de l'ARNpg. En effet, de l'ADNccc et de l'ARNpg sont détectés dans 40% et 66% respectivement des CHC de cette étude, tout en étant à des niveaux bien plus faibles que dans le foie non-tumoral. Les tumeurs sont donc des tissus moins favorables à la réplication virale mais l'identification de souches virales différentes entre des CHC et le foie non-tumoral adjacent a prouvé qu'une réplication pouvait bien avoir lieu dans la tumeur. Par ailleurs, la présence d'ARNpg ou d'ADNccc est associée à une surexpression du NTCP, le récepteur d'entrée du VHB et à une diminution de l'expression de gènes impliqués dans le cycle cellulaire et la réparation des dommages à l'ADN. De plus, l'ADNccc contenu dans ces CHC code généralement pour des variants d'HBx ayant conservé leur activité transcriptionnelle et leur capacité à interagir avec le complexe Smc5/6. Au final, il a été observé que ces tumeurs permettant la réplication virale constituent un sous-

groupe spécifique de CHC plus différenciés et moins invasifs, associé à un meilleur pronostic. La réplication virale dans les CHC nécessiterait donc une faible prolifération combinée une forte différenciation cellulaire (Halgand et al., 2018; Rivière et al., 2019).

Concernant la séquence virale, les tumeurs sont caractérisées par la présence d'altérations fréquentes du gène *HBx* (mutations inactivatrices ou délétions de l'extrémité C-terminale), de délétions de la région préS, et de mutations de la région PréCore/Core, comme cela a déjà été mentionné dans les parties II.C.2 et III.B.3. Une étude rapporte que 39,5% des CHC contiennent des séquences virales qui ont été altérées par des délétions, inversions ou duplications, et qu'une grande partie de ces altérations se situent dans la région également enrichie pour les jonctions d'intégrations. Il n'est cependant pas précisé si ces séquences sont issues des formes intégrées ou épisomales du virus présentes dans les CHC (Toh et al., 2013).

Altérations génomiques

Une comparaison du profil mutationnel de CHC liés au VHB avec celui de CHC liés à d'autres étiologies a permis de déterminer, parmi les voies de signalisation qui sont fréquemment dérégulées dans les CHC (présentées dans la partie I.C), celles qui sont caractéristiques de l'infection par le VHB. En particulier, la voie p53 est très souvent altérée dans les CHC liés au VHB et notamment via des mutations inactivatrices des gènes *TP53* (dans 41% des cas contre 16% dans les CHC liés à d'autres étiologies) et *IRF2* (dans 7% des cas et uniquement dans les CHC liés au VHB), ces deux mutations étant mutuellement exclusives. Dans ces tumeurs liées au VHB, le spectre de mutations de *TP53* a comme particularité d'avoir une forte fréquence de mutations sur le codon 249 (R249S), correspondant à la mutation provoquée par l'exposition à l'Aflatoxine B1. Cette mutation est par ailleurs associée à la présence d'une mutation inactivatrice d'*HBx* et elle est majoritairement observée dans des tumeurs peu différenciées qui se sont développées sur foie sain. Cela suggère donc une coopération entre la mutation R249S et l'inactivation d'*HBx* pouvant favoriser une cancérogénèse sans contexte cirrhotique. De plus, les mutations de *TP53* sont associées à une survie plus faible parmi les patients ayant un CHC lié au VHB mais pas parmi ceux ayant un CHC lié à une autre étiologie, suggérant une valeur pronostic de ces mutations uniquement dans ce groupe spécifique de CHC particulièrement agressif. En revanche, les mutations des gènes *NFE2L2* et *CTNNB1* étant moins fréquentes dans

les CHC liés au VHB, les voies du stress oxydatif et Wnt/ β -caténine semblent moins souvent activées dans ces tumeurs (Amaddeo et al., 2015; Schulze et al., 2015).

Par ailleurs, une caractéristique génomique évidente des CHC liés au VHB est la présence d'intégrations fréquentes dans des gènes impliqués dans la carcinogénèse qui a été détaillé dans la partie précédente. Le promoteur de *TERT* est le site d'insertion le plus récurrent permettant la réactivation de la télomérase, suivi des gènes *MLL4* et *CCNE1* impliqués dans la régulation épigénétique et dans le cycle cellulaire respectivement (Levrero and Zucman-Rossi, 2016). Enfin, un nombre plus élevé d'aberrations chromosomiques est observé dans les CHC liés au VHB. Cela est probablement lié à la capacité du virus d'induire une instabilité génomique via les intégrations virales et via l'activité d'HBx ou à la fréquence des mutations des gènes *TP53* et *IRF2* liés à la maintenance de la stabilité des chromosomes (Guichard et al., 2012; Levrero and Zucman-Rossi, 2016).

Profil transcriptomique

Dans la classification moléculaire présentée dans la partie I.D, les CHC liés au VHB sont retrouvées dans l'ensemble des sous-groupes, soulignant ainsi leur diversité génomique. Cependant, ils appartiennent majoritairement à la classe proliférative (G1-G3), contrairement aux CHC liés au VHC ou à l'alcool. Il s'agit de CHC de grande taille et peu différenciés avec de fréquentes altérations inactivatrices d'*AXIN1*, *IRF2* (G1-G2) et *TP53* (G2-G3) et du gène *HBx*. Les CHC des groupes G1-G2 sont enrichis chez des patients jeunes de sexe féminin et caractérisés par une forte expression de gènes associés à des caractéristiques progénitrices (*EpCAM*, *AFP*, *KRT19*). Les tumeurs du groupe G3 sont des CHC agressifs avec invasion vasculaire (Amaddeo et al., 2015).

Bien que minoritaires, des CHC liés au VHB ont néanmoins un profil non-prolifératif et plus différencié (G4-G6). Ils se développent généralement chez des patients plus âgés ayant un facteur de risque additionnel (infection par le VHC, ALD ou NASH). Ces CHC pourraient donc potentiellement avoir suivi un développement tumoral différent, au cours duquel les mécanismes de carcinogénèse directs du VHB n'ont eu qu'un rôle mineur par rapport au cofacteur et par rapport à l'accumulation normale de mutations somatiques liée au vieillissement. Les CHC des groupes G5-G6 se caractérisent plus spécifiquement par des mutations activatrices de *CTNNB1* et une activation de la voie Wnt- β -caténine (Amaddeo et al., 2015; Levrero and Zucman-Rossi, 2016).

Les caractéristiques cliniques et génétiques des CHC liés au VHB sont présentées en fonction de la classification transcriptomique G1-G6 dans la Figure 28 ci-dessous.

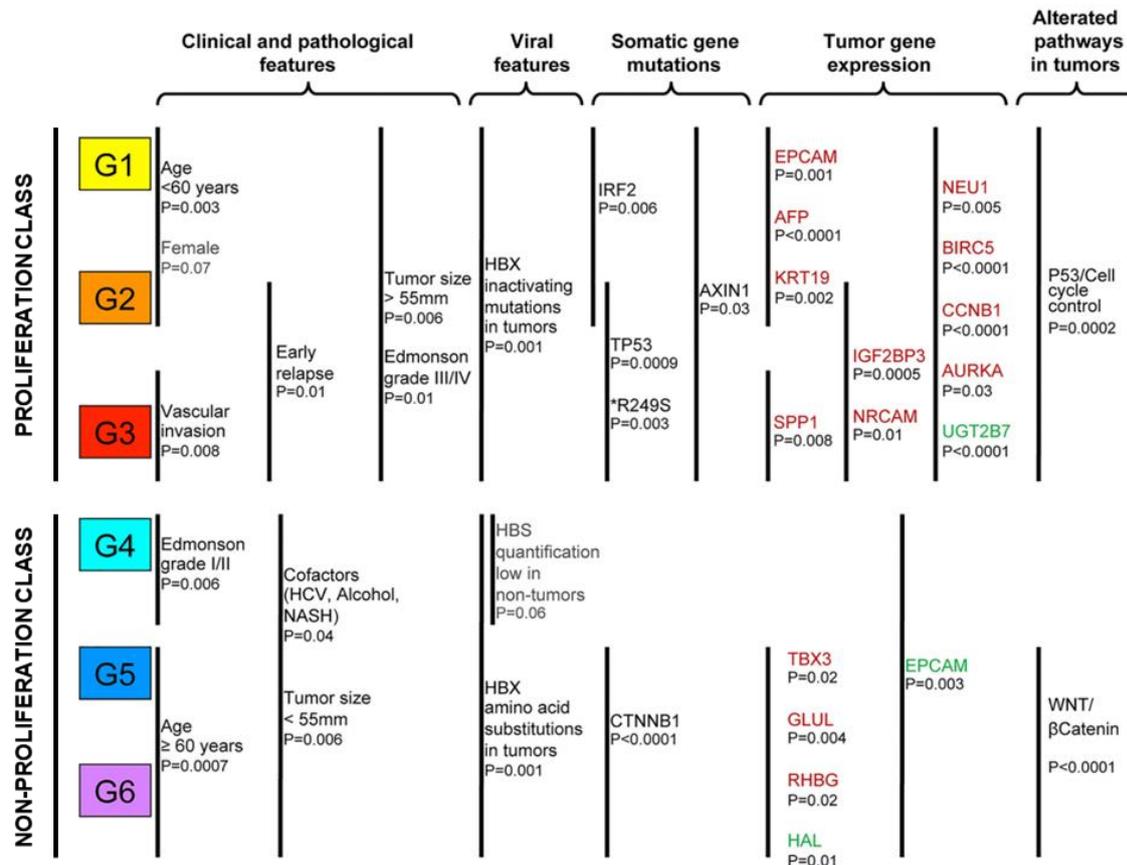


Figure 28 : Représentation schématique des différents sous-groupes de CHC liés au VHB définis par analyse transcriptomique. Les gènes écrits en rouge sont sous-exprimés et les gènes écrits en vert surexprimés (Amaddeo et al., 2015).

3) Enjeux thérapeutiques : infections occultes et intégrations

Deux points indépendants peuvent être considérés comme des enjeux thérapeutiques qu'il serait intéressant d'investiguer pour améliorer le suivi des patients infectés par le VHB et réduire l'incidence des CHC et d'éventuelles récurrences : caractériser les infections occultes (OBI) et utiliser les données des intégrations virales en clinique.

- Les infections occultes

La « guérison fonctionnelle » repose sur la perte de l'AgHBs, principalement due à une séroconversion et donc associée à la présence d'anticorps anti-HBs, et à l'absence de détection d'ADN du VHB dans le sang. L'absence d'AgHBs reflète en effet une forte suppression de la réplication virale et de l'expression des protéines virales. Elle peut

arriver de façon spontanée grâce aux mécanismes immunitaires de l'hôte, quelques mois après l'infection lors la résolution d'une hépatite B aigüe ou après plusieurs années d'une infection chronique (chez 1 à 3% des patients avec infection chronique) mais c'est surtout l'objectif visé par les thérapies antivirales actuelles (Lampertico et al., 2017). Parmi les patients traités, cela reste cependant un événement rare. Les traitements permettent au moins une suppression complète de la réplication virale après 3 ans de traitement dans 70-80% des cas. Cette suppression est associée à un risque plus faible de développer un CHC ou un autre événement hépatique grave (complications cirrhotiques, transplantation hépatique, mortalité liée au foie). En comparaison, la perte de l'AgHBs n'impacte que le risque de survenue d'un CHC en le réduisant davantage. La diminution des dommages hépatiques chez les patients traités semble donc principalement liée à l'inhibition de la réplication virale, qui réduit l'inflammation, permet la régression de la fibrose et empêche ainsi la progression vers la cirrhose. En revanche, la perte de l'AgHBs semble être nécessaire pour diminuer le risque de survenue d'un CHC (Yip et al., 2019). Ce risque n'est cependant jamais totalement éliminé puisque le taux d'incidence du CHC après la perte de l'AgHBs est de 0,55% par an. Ce risque est par ailleurs plus faible si la séroconversion se produit lorsque le patient est jeune et/ou qu'il n'a pas de fibrose avancée (Kim et al., 2015). On parle de « guérison fonctionnelle » et non de « guérison totale » car, bien que la réplication virale soit supprimée, de l'ADN viral peut persister dans le foie des patients, sous forme d'ADNccc ou de séquences intégrées, et le virus n'est donc pas éradiqué.

En revanche, les patients n'ayant pas d'AgHBs détectable dans le sang mais chez lesquels est détecté de l'ADN du VHB dans le sang et/ou de l'ADNccc réplicatif dans le foie sont décrits comme ayant une infection B occulte (OBI) (Raimondo et al., 2019, 2008). La présence de séquences virales intégrées seules n'est donc pas considéré comme une infection occulte. Par ailleurs, la perte de l'AgHBs s'accompagne généralement de l'apparition d'anticorps anti-HBs et du maintien des anticorps anti-HBc. Ces deux types d'anticorps peuvent donc être utilisés pour dépister les OBI, mais ils ne sont pas détectés chez 1 à 20% des patients ayant une OBI et pourraient donc avoir disparu progressivement au cours de l'infection ou même ne jamais avoir été présents (infection occulte primaire). Ainsi, étant donné la difficulté de dépistage, la prévalence de ces infections occultes n'est pas précisément connue dans la population générale, mais elle est plus élevée parmi les patients ayant une maladie chronique du

foie, et/ou ayant une co-infection par le VHC (15%-33%) ou par le VIH (10%-45%). Une OBI serait présente chez 40 à 75% des patients ayant un CHC sans expression d'AgHBs. Ces infections constituent un enjeu thérapeutique important car elles possèdent a priori les mêmes propriétés pro-oncogéniques qui sous-tendent le développement du CHC dans les infections « ouvertes » où l'AgHBs est positif. En effet, chez un patient ayant une OBI, l'ADN du VHB persiste dans les hépatocytes sous formes épisomales et intégrées mais le virus maintient son activité de réplication, de transcription et de synthèse des protéines virales bien qu'à des niveaux très bas, et l'infection peut générer en parallèle une nécro-inflammation faible mais continue pouvant favoriser une progression vers la cirrhose. L'OBI peut donc participer au développement tumoral via les mécanismes directs et indirects déjà décrits, mais le fait qu'elle soit réellement un facteur de risque n'a pas encore été confirmé (Pollicino and Saitta, 2014; Raimondo et al., 2019). Les intégrations virales sont observées dans 75% des CHC de patients ayant une OBI et sont comparables à celles détectées dans les CHC liés à une infection ouverte en termes de fréquence des événements, de sites d'insertion dans le génome humain et de séquences intégrées. Elles peuvent donc jouer un rôle majeur dans la carcinogénèse liée aux OBI (Saitta et al., 2015).

- Les intégrations virales

La caractérisation des intégrations virales constitue un deuxième enjeu thérapeutique important pour les CHC liés au VHB car c'est un événement très précoce de la carcinogénèse. Le nombre d'intégrations dans les CHC est inversement corrélé avec l'âge des patients et il est associé avec leur niveau d'AFP mais également avec leur survie, les patients dont la tumeur contient un grand nombre d'intégrations ayant une durée de survie moyenne plus courte (Sung et al., 2012). Une étude suggère par ailleurs que non seulement le nombre d'intégrations mais également les sites d'insertions peuvent influencer sur le pronostic des patients et notamment sur la survenue d'une récurrence après traitement chirurgical du CHC. Ces travaux ont identifiés parmi les gènes ciblés par les intégrations un enrichissement en gènes suppresseurs de tumeurs dans les tissus non-tumoraux de patients ayant eu une récurrence tumorale, mettant ainsi en avant que ces tissus peuvent contenir des informations permettant potentiellement de prédire une éventuelle récurrence (Yoo et al., 2017). Avec le même objectif, une étude récente a analysé le plasma collecté avant chirurgie et deux mois après chirurgie ainsi que le tissu tumoral de patients ayant eu un CHC lié au VHB. A

partir de l'intégration principale identifiée dans un CHC, un set de sondes spécifique a été créé pour chaque patient dont la tumeur contenait au moins une insertion, et a permis de détecter par Droplet Digital PCR des séquences d'ADN virus-hôte dans le plasma dans 98% des cas. La comparaison des plasmas collectés a alors montré que 22,7% des patients avaient une signature identique avant et après chirurgie, révélant la présence de cellules tumorales résiduelles. Parmi ces patients, 90% ont développé une récurrence dans l'année contre 38,2% des patients pour lesquels l'intégration n'était plus détectable dans le plasma après chirurgie. Et pour 77,3% des patients, le plasma collecté lors du diagnostic de la récurrence avait la même signature que le plasma collecté initialement, indiquant que ce deuxième CHC était issu du même clone tumoral que la tumeur primaire. Ainsi, la présence de séquences d'ADN virus-hôte dans le plasma est un facteur de risque indépendant prédictif d'une récurrence. En résumé, la détection dans le plasma d'un patient de séquences caractéristiques d'une intégration virale présente dans sa tumeur pourrait être utilisée comme biomarqueur circulant permettant une surveillance personnalisée ainsi que l'identification d'une récurrence plus rapidement que par imagerie et plus spécifiquement que par détection des mutations somatiques de l'ADN tumoral circulant dans le plasma (Li et al., 2020).

Enfin, l'analyse des intégrations virales pourrait aussi potentiellement permettre de mettre au point des immunothérapies personnalisées pour les CHC lié au VHB. Un traitement prometteur de ces CHC consiste à modifier des cellules T pour qu'elles ciblent spécifiquement les AgHBs et éliminent les cellules qui les expriment. Cette thérapie ne peut cependant pas être utilisée chez la majorité des patients traités chez qui la réplication virale a été supprimée et dont les cellules tumorales ne synthétisent donc plus d'AgHBs normaux. Cependant, une étude a montré que les séquences intégrées présentes dans les CHC induisent la synthèse d'antigènes altérés pouvant être reconnus et activés par des cellules T spécifiquement modifiées. Ainsi, en analysant les transcrits générés par les séquences du VHB intégrées dans un CHC, il est possible de modifier des cellules T en conséquence et de mettre en place cette immunothérapie personnalisée. Cela n'a pour l'instant été testé que sur deux patients et nécessite donc une validation dans une plus grande série, mais si les résultats encourageants sont confirmés, la caractérisation des séquences intégrées pourrait alors être utilisée pour traiter un grand nombre de patients ayant une sérologie AgHBs négative et donc notamment ceux ayant une OBI (Tan et al., 2019).

CONTEXTE DU PROJET DE RECHERCHE ET OBJECTIFS

Les infections virales représentent le facteur de risque principal des carcinomes hépatocellulaires dans le monde et notamment le virus de l'Hépatite B. Les CHC liés à ce virus ont comme spécificité par rapport à ceux liés au virus de l'Hépatite C de se développer dans 30% à 50% des cas en absence de cirrhose, ce qui suggère que le VHB possède des propriétés oncogéniques propres qui pourrait être directement à l'origine de la carcinogénèse. Cependant le rôle du virus dans l'initiation tumorale reste largement débattu et les mécanismes moléculaires qui favorisent le développement tumoral des années après l'infection ne sont pas encore bien compris.

En ayant constitué une grande série de carcinomes hépatocellulaires, le laboratoire a pour objectif de mieux caractériser les altérations génomiques déjà décrites et d'identifier des biomarqueurs et de nouvelles cibles thérapeutiques afin d'améliorer la prise en charge et le suivi des patients. Les CHC liés au VHB ont fait l'objet d'une étude spécifique du laboratoire publiée en 2015 qui a comparé 86 tumeurs positives pour le VHB à 90 tumeurs liées à d'autres étiologies (Amaddeo et al., 2015). Dans ce projet, un séquençage ciblé des gènes *HBs* et *HBx* du virus et de neuf gènes du génome humain a permis de mettre en évidence l'hétérogénéité de ces CHC au niveau génomique, et l'analyse de l'expression de 37 gènes de les positionner dans la classification transcriptomique G1-G6 (Boyault et al., 2007). En parallèle de ces travaux, la découverte d'une séquence intégrée de l'Adénovirus Associé 2 (AAV2) dans le promoteur de *TERT* a incité le laboratoire à réaliser une capture virale sur près de 200 échantillons tumoraux et non-tumoraux dans le but d'identifier des insertions récurrentes de ce virus (Nault et al., 2015)

Le premier objectif de mon projet de thèse était donc d'approfondir la caractérisation des CHC liés au VHB en utilisant des données de séquençage de nouvelle génération, et notamment en mettant au point au laboratoire une capture virale spécifique au VHB afin d'analyser une grande série de CHC. Cette technique a pour but de pouvoir identifier précisément les intégrations du virus dans les tissus hépatiques, qui peuvent avoir un rôle oncogénique majeur, et qui n'avaient pas été décrites auparavant dans la collection de tumeurs du laboratoire. J'ai donc réalisé une analyse intégrée de données moléculaires de tissus hépatiques tumoraux et non-tumoraux, issues de la

capture virale, de séquençage d'exome ou génome et de séquençage de transcriptome (RNAseq), en lien avec les caractéristiques cliniques des patients.

Un deuxième objectif était d'identifier de nouvelles altérations génomiques dans des tumeurs développées en absence de cirrhose, sans facteur étiologique connu ou dans le cadre d'une infection par le VHB. J'ai ainsi participé à la caractérisation des insertions virales et de leurs conséquences fonctionnelles dans un sous-groupe spécifique de CHC agressifs de patients non-cirrhotiques qui représente 7% des CHC. Ces tumeurs possèdent des altérations des gènes *CCNA2* et *CCNE1* et une signature de réarrangements chromosomiques bien spécifique induit par un stress réplicatif.

Enfin, un dernier objectif de mon projet était d'investiguer la présence d'autres virus dans les tumeurs hépatiques et en particulier de décrire plus en détail l'infection par le virus AAV2 pour clarifier ses conséquences sur la carcinogénèse. J'ai donc pris part à l'élaboration et la mise en place de nouveaux outils expérimentaux et bio-informatiques pour caractériser des insertions virales dans leur globalité. Le but était d'identifier les formes virales présentes dans les tissus et de reconstruire les différents génomes.

La collection de CHC et tissus hépatiques non-tumoraux disponible au laboratoire et l'analyse intégrée des données génomiques et transcriptomiques qui en sont issues représentent une opportunité unique de décrypter le mécanisme d'insertion virale et d'analyser les conséquences sur la carcinogénèse. Approfondir la caractérisation des CHC liés au VHB est un enjeu d'autant plus important qu'il n'existe pas à l'heure actuelle de traitement curatif pour cette infection. Les patients nécessitent donc un suivi et un traitement à vie et conservent sous traitement un risque élevé de développer un CHC (Coppola et al., 2016).

RESULTATS

I. Caractérisation des intégrations du Virus de l'Hépatite B dans les tissus hépatiques tumoraux et non-tumoraux

Le Virus de l'Hépatite B est le facteur de risque principal du carcinome hépatocellulaire et l'ADN viral possède la capacité de s'intégrer dans le génome humain bien que cela ne fasse pas partie de son cycle de réplication (Seeger and Mason, 2015). La carcinogénèse pouvant également être favorisée par les propriétés oncogéniques de la protéine virale HBx et par l'inflammation générée par l'infection chronique, le rôle joué par l'intégration du VHB dans l'initiation et/ou dans le développement tumoral est encore largement débattu à l'heure actuelle (Tu et al., 2015). De nombreux travaux ont ainsi cherché à décrire les intégrations du VHB et leurs conséquences oncogéniques, en utilisant le séquençage de nouvelle génération (Références dans la Table 3 de la partie Introduction). Cependant, les résultats présentés sont parfois contradictoires sur le nombre et la caractérisation des événements, et les études contiennent rarement des analyses intégrées des sites d'insertions, des formes virales présentes et de leurs séquences, des caractéristiques cliniques et moléculaires des patients et enfin des conséquences fonctionnelles de ces intégrations virales dans l'hépatocarcinogénèse.

En utilisant la série de CHC liés au VHB disponible au laboratoire, nous avons ainsi cherché à réaliser une étude approfondie de ces événements d'intégrations. La première étape a consisté à constituer une cohorte de 1128 patients et à cribler les tissus hépatiques de ces patients par PCR quantitative pour sélectionner ceux ayant de l'ADN du VHB dans le foie ou ayant un statut sérologique positif pour le VHB. J'ai ensuite pu mettre au point une technique de capture virale afin d'identifier précisément les insertions présentes dans une sélection de tissus contenant une grande quantité d'ADN viral. La particularité de cette série de 190 échantillons analysés par capture est le pourcentage important de tumeurs développées chez des patients d'origine européenne (54%) et africaine (30%). En effet la très grande majorité des études publiées a été réalisée sur des patients d'origine asiatique infectés par des souches virales de génotypes B et C, et aucune information n'est disponible sur les intégrations des autres génotypes viraux (Fujimoto et al., 2016; Li et al., 2019; Zhao et al., 2016).

Cette expérience a tout d'abord permis de mettre en évidence la présence d'un nombre plus élevé d'événements d'intégrations dans les tissus non-tumoraux par rapport aux tumeurs et notamment dans les tissus ayant une réplication active du virus chez des

patients jeunes et d'origine non-européenne ayant été infecté plus tôt dans leur vie. Le processus d'intégration viral a lieu préférentiellement dans des régions du génome fortement exprimées où la chromatine est ouverte et donc facilement accessible. Une expansion clonale est observée dans certains échantillons non-tumoraux et semble liée à un avantage sélectif de la cellule qui pourrait être indépendant des insertions virales qu'elle contient dans son ADN.

Dans les tumeurs en revanche, les intégrations virales sont à l'origine de la grande majorité de l'ADN viral détecté et elles sont fréquemment localisées à proximité de gènes connus comme ayant un rôle « driver » dans l'hépatocarcinogénèse. Le promoteur de *TERT* est le site d'insertion le plus récurrent et contient une intégration clonale du virus dans un tiers des CHC de notre série. Ces insertions induisent une surexpression du gène via la séquence « Enhancer 1 » du génome viral, qui est indépendante de l'orientation de la séquence intégrée mais qui est d'autant plus élevée que le site d'insertion est proche de l'ATG du gène. La présence d'insertions du VHB dans le promoteur de *TERT* est plus fréquente parmi les CHC liés au VHB appartenant à la classe transcriptomique des tumeurs non-prolifératives tout comme la présence d'une mutation du promoteur. Cela peut s'expliquer par le fait qu'un nombre important de CHC prolifératifs ne contiennent aucune altération du promoteur et constituent un groupe spécifique de tumeurs développés en absence de cirrhose qui expriment fortement des marqueurs progéniteurs. Ce groupe rassemble des CHC ayant des intégrations du VHB dans le gène *KMT2B (MLL4)*, des tumeurs de patients très jeunes d'origine africaine sans altération identifiée et des CHC ayant des mutations du gène *TP53* souvent associées à une exposition des patients à l'Aflatoxine B1.

La mutation ou délétion du gène *TP53* est la deuxième altération la plus fréquemment observée après celle du promoteur de *TERT* dans les CHC liés au VHB. Elle est associée à une importante instabilité chromosomique qui est observée à la fois via un plus grand nombre de réarrangements structuraux dans le génome humain et un plus grand nombre d'insertions du VHB pouvant être impliqués directement dans ces réarrangements. En effet, un changement du nombre de copie d'une région chromosomique est identifié au site d'intégration dans un tiers des événements clonaux détectés dans les tumeurs. Ces réarrangements peuvent être situés dans les régions centromériques ou télomériques et induire la délétion de régions contenant

des gènes suppresseurs de tumeurs comme *TP53* ou le gain de régions contenant des oncogènes comme *MYC* ou de la région du promoteur de *TERT*.

Ce projet a donc montré que les intégrations du VHB ont lieu précocement dans l'histoire naturelle de l'infection et sont favorisées par la présence d'ADN viral répliatif. Elles peuvent ne pas avoir de conséquences fonctionnelles mais rester présentes de façon stable dans des cellules suivant une expansion clonale induite par d'autres facteurs telle que la pression de sélection liée à l'inflammation chronique. Mais ces intégrations peuvent aussi avoir des conséquences oncogéniques directes en modifiant par mutagénèse insertionnelle la structure ou l'expression d'un gène, ou indirectes en induisant des pertes ou des gains de régions chromosomiques spécifiques et/ou en favorisant une instabilité chromosomique globale dans la cellule.

Article 1: Hepatitis B virus integrations promote local and distant oncogenic driver alterations in hepatocarcinogenesis

Camille Péneau, Sandrine Imbeaud, Tiziana La Bella, Théo Z. Hirsch, Stefano Caruso, Julien Calderaro, Valérie Paradis, Jean-Frédéric Blanc, Eric Letouzé, Jean-Charles Nault, Giuliana Amaddeo, Jessica Zucman-Rossi

Article in preparation.

Hepatitis B virus integrations promote local and distant oncogenic driver alterations in hepatocarcinogenesis

Camille Péneau^{1,2}, Sandrine Imbeaud^{1,2}, Tiziana La Bella^{1,2}, Théo Z. Hirsch^{1,2}, Stefano Caruso^{1,2}, Julien Calderaro³, Valérie Paradis^{4,5}, Jean-Frédéric Blanc^{6,7,8}, Eric Letouzé^{1,2}, Jean-Charles Nault^{1,2,9}, Giuliana Amaddeo¹⁰, Jessica Zucman-Rossi^{1,2,11}

1. Centre de Recherche des Cordeliers, Sorbonne Université, Inserm, Université de Paris, F-75006 Paris, France

2. Functional Genomics of Solid Tumors laboratory, équipe labellisée Ligue Nationale contre le Cancer, Labex OncoImmunology, F-75006, Paris, France

3. Service d'Anatomopathologie, Hôpital Henri Mondor, APHP, Institut Mondor de Recherche Biomédicale, F-94010 Créteil, France

4. Service de Pathologie, Hôpital Beaujon, APHP, F-92110 Clichy, France

5. Université Paris Diderot, CNRS, Centre de Recherche 27 sur l'Inflammation (CRI), Paris, F-75890, France

6. Service Hépato-Gastroentérologie et Oncologie Digestive, Hôpital Haut-Lévêque, CHU de Bordeaux, F-33000 Bordeaux, France

7. Service de Pathologie, Hôpital Pellegrin, CHU de Bordeaux, F-33076 Bordeaux, France

8. Université Bordeaux, Inserm, Research in Translational Oncology, BaRITOn, F-33076 Bordeaux, France

9. Service d'Hépatologie, Hôpital Jean Verdier, Hôpitaux Universitaires Paris-Seine-Saint-Denis, APHP, F-93140 Bondy, France

10. Service d'Hépato-Gastro-Entérologie, Hôpital Henri Mondor, APHP, Université Paris Est Créteil, Inserm U955, Institut Mondor de Recherche Biomédicale, F-94010 Créteil, France

11. Hôpital Européen Georges Pompidou, APHP, F-75015 Paris, France

Contact information :

Jessica Zucman-Rossi ; MD, PhD

Centre de Recherche des Cordeliers

Functional Genomics of Solid Tumors

15 rue de l'Ecole de Médecine, 75006 Paris, France

TEL : +33 6 01 07 78 75

Email : jessica.zucman-rossi@inserm.fr

ABSTRACT

Infection by Hepatitis B Virus (HBV) is the main risk factor for Hepatocellular Carcinoma (HCC) worldwide. HBV can directly drive carcinogenesis through integrations in human genome. This integrative analysis of a French cohort of 190 patients mainly coming from Europe and Africa provides a deep characterization of HBV integrations, in relation with viral and host genomics and clinical features. Clonal HBV integrations directly contribute to HCC development through alterations of cancer driver genes, either at distance when co-localizing with copy number alterations, either locally through viral Enhancers. Together, this analysis gives new insights on HBV-associated driver mechanisms involved in hepato-carcinogenesis.

INTRODUCTION

Despite the implementation of vaccination programs and the strong decrease of new infections among children, the percentage of people living with chronic Hepatitis B Virus (HBV) infection worldwide remained as high as 3.5% of the global population in 2015¹. Moreover, only 9% of infected people were diagnosed and only 8% of them were under treatment, which makes Hepatitis B a major public health threat now prioritized by the World Health Organization¹⁻³. The burden of HBV infection is closely related to the development in HBV-infected patients of Hepatocellular Carcinoma (HCC), the more frequent primary liver cancer and the fourth cause of cancer death worldwide⁴. Persistent HBV infection is actually responsible for more than 50% of all HCC cases worldwide and up to 85% in some areas where the infection is endemic⁵. Even patients receiving antiviral treatment who had maintained viral suppression remain at risk of developing a HCC⁶⁻⁹.

HBV-related HCC occur not only in the setting of cirrhosis but also in normal liver¹⁰, underlying that the virus has its own oncogenic properties besides the induction of chronic inflammation in the liver. HBV is a small 3.2kb DNA virus that could integrate in human DNA and promote cell transformation by insertional mutagenesis or through expression of viral oncoproteins such as the protein HBx¹¹⁻¹³. As HBV integrations occur early during infection¹⁴, it is a key factor reflected in the genetic landscape of HCC. Therefore, comparing the insertions occurring in normal hepatocytes and in tumor cells may enable to identify new genomic defects associated with tumor development.

The development of next generation sequencing has given in the last decade the opportunity to characterize more precisely the role of HBV insertion in HCC initiation and development¹⁵⁻²⁰. Cancer-related genes such as *TERT*, *CCNE1* and *KMT2B* have been identified as recurrently targeted by HBV insertions in HCC, with specific functional consequences and clinical outcomes^{17,21-23}. Thus, the presence of integrated HBV DNA in hepatocytes has been pointed out as a driver of hepato-carcinogenesis through the alteration of the expression or function of these genes¹³. But up to now, such identifications of HBV insertions in HCC have been mainly performed in Asian populations where HBV genotypes B and C are the most prevalent HBV strains²⁴. Furthermore, as a significant proportion of HBV-related HCC does not contain any integration in a known cancer-related gene, two other independent mechanisms involving integrated HBV DNA have been proposed: induction of chromosomal instability and persistent expression of altered HBV genes²⁵. However, the interactions between the three mechanisms mentioned and the precise role they might play individually or synergistically in HCC initiation and development remain to be fully elucidated.

Therefore, in this work, we developed a reliable pipeline of analysis to characterize HBV integrations in a French cohort of 190 patients mainly coming from Europe (49%) and Africa (33%). We identified different integration profiles and HBV genome rearrangements in 234 tumors and their corresponding non-tumor liver tissues, in relation with clinical features and viral characteristics: replicative forms, genotypes and variants. We also investigated the oncogenic consequences of recurrent HBV integrations on gene expression and chromosomal instability, involved in hepato-carcinogenesis occurring with or without underlying cirrhosis.

RESULTS

HBV integrations occur in open chromatin regions and relate to viral replication in the liver.

In 84% of non-tumor tissues (143 out of 170), we identified 6610 HBV integrations breakpoints at HBV/human junctions corresponding to unique (82%), subclonal (17%) or clonal (1%) events, involving similarly all HBV genotypes (**Fig 1a; Supplementary Fig. 1; Supplementary Table 1**). In the human genome, HBV integration breakpoints were enriched in active and open chromatin regions, and they were closed to early-replicated and highly expressed genes. Integrations were also more frequent in simple repeats, in particular in telomeric and sub-telomeric regions, but not in Alu motifs, fragile sites or LINE/SINE elements (**Supplementary Fig. 2a**). Reconstruction of 488 events of subclonal or clonal integrations in non-tumor tissues showed that they all contained a large part of the whole HBV sequence (median size=2912nt) and 42% of them were bordered by 1-to-8 nucleotide homology between the targeted human sequence and the integrated HBV sequence, suggesting the involvement of Microhomology-Mediated End Joining (**Supplementary Fig 2b-2c**).

In the non-tumor liver tissues, recurrent clonal/subclonal insertions were identified in 20 genes with the following characteristics: highly expressed (8/20, FPKM>70) or very large genes (8/20, >200kb), such as *FNI* (14 cases), *CPS1* (4 cases), *KCNT2* or *ADH1B* (3 cases), *ADH4* or *ADH6* or *ALB*... (2 cases) (**Fig. 1b**). In *FNI*, 86 HBV breakpoints were distributed all over the gene, mainly in introns (78/86) and without specific hotspot. In RNA-Seq, two major types of fusion transcripts were identified: (1) out-of frame HBx-FN1 transcripts in the same or in the opposite direction (11 among 32 reconstructed transcripts), and (2) in-frame HBs-FN1 transcripts generated by a cryptic splice site at position 458 in the S ORF (13 among 32; **Supplementary Fig. 3a**). Overall, the diversity of the fusion transcripts HBs-FN1 starting at different exons of *FNI* and the global lack of over-expression of the genes targeted by HBV integrations indicated that most of the clonal/subclonal insertions in non-tumor liver tissues were probably non-functional (**Supplementary Fig. 3b**). Interestingly, clonal insertions were observed in 11% of non-tumor samples, all in fibrotic livers (F2-F4) with a decreased inflammatory response and NF- κ B/TNF signaling at the RNA level, suggesting a local proliferation of non-neoplastic hepatocytes (**Supplementary Fig. 3c**).

In 170 analyzed non-tumor liver tissues, a high HBV copy number per cell was significantly associated with female gender, young age, African or Asian geographic origin, and positivity for HBsAg, HBeAg and HBV DNA in the serum. In contrast, presence of cofactors of chronic liver disease such as active HCV or HDV infection, alcohol consumption or metabolic syndrome was associated with lower copy number of HBV DNA. Only 19% of the samples contained replicative HBV DNA, they were enriched in Genotype A and showed a higher number of integrations, underlying that HBV replication and integration were linked processes (**Fig. 1c-1d; Supplementary Fig. 1**). In the same line, samples with high HBV copy number showed frequent mutations of the Basal Core Promoter (BCP; A1762T/G1764A) or the Reverse Transcriptase (RT) region: mutations known to favor viral replication (**Fig. 1c**).

Complex HBV integrations and viral genome rearrangements are frequent in HCC

In 88% of 177 analyzed HCC tumors, we identified HBV integrations in human genome totalizing 2199 breakpoints at HBV/human junctions, and 31% of them corresponded to clonal events (**Fig 1a; Fig. 2a-2b**). As in non-tumor tissues, the number of HBV integrations in HCC was highly correlated with the HBV copy number per cell (**Fig. 2c**). However, whereas tumors

showed a higher HBV copy number per cell, they harbored a lower number of integrations per sample than in their corresponding non-tumor tissues (12 versus 39, $p < 0.001$) (**Fig. 2d; Supplementary Fig. 4a-4b**). In both types of tissues, the most frequent hotspot of HBV integration was observed in the C-terminal region of the HBx gene around the DR1 sequence (1817-1836) corresponding to the ends of the double-strand linear (ds)DNA form of HBV (**Fig. 2e**). Yet, in tumors, only a part of the HBV genome was detected, corresponding to integrated HBV DNA with frequent truncation of the HBx gene (**Supplementary Fig. 5a**). In addition, HBV genomes in tumors contained a higher number of structural variants within the viral sequence (deletions, duplications, inversions; **Fig. 2f**) and only 3% of these structural variants in tumors were observed in the adjacent liver tissues (17/514). Finally, the localization of viral integration breakpoints in the human genome suggested more frequent chromosome rearrangements in tumors when compared with non-tumor tissues, suggesting a more complex integration process or post-integration structural modifications (**Supplementary Fig. 5b**).

HBV replication was less frequent in tumors (8%) than in non-tumors (19%, $p < 0.001$) (**Fig. 2g**). In addition, episomal forms in tumors accounted for a small proportion of HBV DNA, since most of HBV transcripts derived from integrated genomes with truncated HBx (**Supplementary Fig. 5c-5d**). Moreover, the majority of tumors with clonal HBV integrations had a positive mRNA expression level in the S region, comparable with tumors containing replicative HBV forms (**Supplementary Fig. 5e**). Finally, by analyzing the Variant Allele Frequency (VAF) of the HBV variants in the viral genome, the same major genotype was identified in the tumors and their corresponding non-tumor liver tissues in 138 out of 140 cases. Surprisingly, we observed a negative selection of HBeAg variants (A1762T and G1764A BCP variants, G1896A PC variant) and antiviral resistance-associated mutations (RT region) in the tumors, suggesting that emergence of variants in the HBV sequence aimed to improve the virus fitness but does not give specific advantage for tumor development (**Supplementary Fig. 5f**). Overall, these results suggest that HBV integrations in non-tumor and tumor liver tissues may reflect different viral dynamics and selection processes not directly correlated.

Human chromosome rearrangements were recurrently delimited by HBV integrations.

Among 522 clonal integrations identified from the capture, WES and/or WGS sequencing of 121 HCC, 183 events (35%) delineated precisely boundaries of chromosome copy number alterations (CNA) in the human genome, showing a direct relationship between the viral insertion event and a chromosome structural rearrangement (**Fig. 3a**). Precisely, three major types of CNA bordered by HBV integrations were observed in more than 10 samples: (1) large deletions of the chromosome 17p including the locus of the tumor-suppressor gene *TP53* (15 tumors) and frequently related to centromeric insertions (7/15), (2) focal gains of *TERT* on the chromosome 5p (14 tumors), and (3) large gains at chromosome 8q including the oncogene *MYC* (13 tumors) and related to centromeric insertions (4/13) (**Fig. 3b**; see examples #1733T and #1597T in **Fig. 3c-3d**).

In order to investigate more precisely how integrations and copy number alterations were associated, we re-constructed genomic sequences at the insertion breakpoints in three selected cases with long-read or Bionano sequencing. In #1151T, we identified 2 different HBV integrations on chromosome 8q that together with a whole duplication of the human genome resulted in a large 8 copies amplification of 57 Mb that included *MYC* oncogene. We reconstructed the two distinct integration events: (1) HBV integration resulted in an inter-chromosome t(8q21;17p11) translocation inducing a large 17p deletion, and (2) HBV integration located in the 8q centromeric region bordered a duplication-inverted-like event

potentially reflecting an isochromosome iso(8q) (**Supplementary Fig. 6a**). In #1597T (**Fig. 3d**), we identified a *TERT* amplification at 8 copies associated with multiple structural variants including two HBV integrations located within chromosome 1q (**Supplementary Fig. 6b**).

Interestingly, even if HCC with integration-associated rearrangements had significantly a higher number of both HBV integrations and CNA in their genome, these two features were not correlated (**Supplementary Fig. 7a**), meaning that only a subset of the viral integrations were associated with chromosome rearrangement, in particular integrations located at centromeric or telomeric regions. Interestingly, centromeric integrations were enriched in young patients with African origin (**Supplementary Fig. 7b**); telomeric HBV integrations usually occurred directly or in proximity to telomere repeats (see examples #4229T and #4268T in **Supplementary Fig. 7c**) and were enriched in tumors with HBV replication and a high number of CNA (**Supplementary Fig. 7a**). As shown in #1994 with Bionano sequencing, telomeric HBV integration can support the translocation of a large chromosome region at the extremity of another chromosome (**Supplementary Fig. 7d**). These results showed that functionally, integration-associated rearrangements frequently altered a cancer driver gene, either at distance when associated with recurrent *TP53* or *MYC* alterations, either locally with *TERT* alterations.

HBV genomes can integrate in cancer driver genes with *cis*-activating consequences

In tumors, HBV integrations followed a different distribution along the human genome when compared to non-tumors and HBV copy number/cell was associated with the number of clonal events (**Fig. 2a; Fig. 4a**). These clonal insertions were enriched around cancer driver genes, suggesting a strong functional selection of cells containing such events (**Supplementary Fig. 8a**). However, among the 248 genes harboring clonal integrations in their vicinity (+/-25kb), only three genes were recurrently found in more than two HCC: *TERT* (n=55), *CCNE1* (n=4), *KMT2B* (n=4). Five other genes had HBV integrations in two samples: *AHRR*, *CTNND2*, *NRG1*, *TRIM16L* and *ST18* (**Fig. 2a – Supplementary Table 1**).

All HBV integrations at the *TERT* locus were located in the promoter region and 5'UTR (**Fig. 4b - Supplementary Fig. 8b**). Most of the promoter integrations were in the same 5'>3' orientation and contained both viral Enhancer regions (39 out of 55). Increased mRNA *TERT* expression was higher for HBV integrations situated closer to the *TERT* ATG (less than 800bp) but was independent of the orientation of the inserted sequence (**Fig. 4b – Supplementary Fig. 8c**); they induced a higher over-expression than promoter mutations at the -124 and -146 hotspots and structural variants inducing copy number alterations (**Supplementary Fig. 8d**). Modeling *TERT* promoter integration identified in #3885T in luciferase reporter vector (**Supplementary Fig. 8e-8f**), we confirmed that *TERT* activation was caused by the two viral Enhancers with Enhancer 1 providing a stronger activation compared to Enhancer 2 (**Fig. 4c**). HBV insertions in *TERT* promoter were exclusive from mutations or structural variations of the same region, and more frequent in tumors with a higher HBV copy number per cell driven by more HBV clonal insertions and more HBV replication.

Three additional HBV integrations were identified in *CCNE1* using RNA-seq, they were located in the promoter region of the gene and induced a significant overexpression but no transcript alteration at the mRNA level (**Supplementary Fig. 9a-9b**). HBV insertions in *CCNE1* or *CCNA2* genes were part of a homogenous group of HCC (CCN-HCC) characterized by a signature of specific chromosome rearrangements (templated-insertion cycles) induced by stress replication, as we previously described²¹. In addition, four *KMT2B* integrations were identified between exons 3 to 6, which increased the mRNA expression and altered the

transcript structure through alternative splicing of intron retention (**Supplementary Fig. 9a-9c**). Overall, *TERT* promoter integrations were enriched in non-proliferative tumors ($p=0.007$; **Fig. 4d**). In contrast, *CCNE1*, *CCNA2* or *KMT2B* integrated tumors were more frequently identified in non-cirrhotic patients (5/8) with African or Asian origin (6/8), in tumors belonging to the proliferative class of HCC associated with G1 (*KMT2B*) or G3 (CCN-HCC) transcriptomic subgroup. Interestingly, CCN-HCC showed *TERT* promoter alterations whereas *KMT2B*-integrated HCC did not show *TERT* promoter or any other driver genes alterations, suggesting different processes of carcinogenesis (**Fig. 4d**).

Integrated analysis between clinical, HBV and human genomic alterations

Compared with 135 HBV-negative tumors previously sequenced in our team^{21,26}, the 130 HBV-positive HCC were enriched in younger patients with African or Asian origin, with cirrhosis, and frequent HDV infection or carcinogens exposure (**Fig. 5a, left**). Indeed, the two sporadic mutational signatures 22 and 24, characteristics of acid aristolochic (AA) and Aflatoxin B1 (AFB1) exposure respectively, were more frequent within HBV-positive patients. Transcriptomic and genomic analysis revealed an association between HBV infection and G3 transcriptomic class, more frequent *TP53* mutations and less *CTNNB1* mutations in tumors.

In addition, in the series of 177 HBV-positive patients analyzed in capture, 60 had an African origin; they were younger patients and had HBV-related HCC characterized by AFB1-mutational signature and a high expression of progenitor markers such as *EPCAM* (**Fig. 5a, right**). Interestingly, no specific viral feature identified in tumors were significantly associated with the geographic origin of patients, except for the HBV genotype. Among patients with cofactors, whereas AFB1 exposure was strongly associated with *TP53* R249S mutation, HBV/HDV-related HCC harbor less *TERT* promoter or any other driver genes alterations. Finally, 10 patients had a negative HBsAg serology, they were negative for HBV DNA in the serum but positive in the liver; their tumors showed low number of HBV copies per cell and of HBV integrations and these HCC were enriched in G1 transcriptomic group. Of note, two of these HCC had a clonal HBV integration in the *TERT* promoter (#4229T and #4265T; **Supplementary Table 1**).

Furthermore, tumors with a high number of HBV integrations (>10) were associated with a poor prognosis, independently from other features such as tumor size, microvascular invasion, differentiation status and transcriptomic groups (**Fig. 5b-5c; Supplementary Table 2**). This association between the high number of integrations in tumors and poor survival was independent from the other features of HBV infection (**Supplementary Table 2**). However, these patients were significantly younger and with tumors harboring a high proportion of HBV integrations affecting cancer driver genes such as *TERT* ($p=0.02$) through an insertion in the promoter, *TP53* ($p<0.001$) or *MYC* ($p=0.03$) through HBV-associated CNA (**Fig. 5a**).

DISCUSSION

This study presents an integrative analysis of HBV genomes in non-tumor and tumor liver tissues of a large cohort of patients mainly coming from Europe and Africa. On one hand, HBV integrations in non-tumor tissues reflect an active viral replication and the expansion of hepatocytes with a low inflammatory response. On the other hand, HBV integrated sequences in tumors highlight a functional selection of cells presenting HBV-associated structural rearrangements or insertional mutagenesis as driver alterations.

Our method of identification of HBV integrations based on viral capture coupled with WGS and WES on frozen tissues enabled to characterize integrations based on their clonality defining precisely the detected events as “clonal”, “sub-clonal” or “unique” with a quantitative method. In accordance with the clonal cell expansion during carcinogenesis and the selection of functional genomic alterations, more clonal HBV insertions were identified in tumors compared to the non-tumor liver tissues. Also, the lack of clear discrimination between clonal and non-clonal insertion events may explain the discrepancies on the number of HBV integrations identified in previous NGS-based studies^{15,16,23,27}. In the same line, we observed an important enrichment of HBV integrations in simple repeats, underlying the importance of filtering insertional events detected in viral capture to remove identical integrations. Therefore, we believe our study is based on a highly confident description of HBV integrations that can be used to provide a comprehensive view of the integration process in liver tissues.

In non-tumor liver, we confirmed that integration events occur more frequently in regions of the human genome in an open chromatin state^{23,28} resulting in part from Microhomology-Mediated End Joining^{16,29,30}. More importantly, viral replication is related to the number of integrations reflecting either an increased number of infected cells, either a multi-steps integration process within a single cell. Both hypotheses are probably occurring concurrently in the liver. A recent *in vitro* study¹⁴ has shown that the integration rate within the first week after infection is not altered by replication suppression, suggesting that the majority of integration events occurs at primo-infection. However, all NGS-based studies, including ours, identified the presence of several clonal integrations in some cellular clones. Therefore, several HBV integrations must accumulate at a low rate in a single cell, promoted by active replication.

Number of HBV integrations may increase over time within hepatocytes in normal liver and as *in vitro* and *in vivo* studies have shown that HBV integrations occur early after infection^{14,31,32}, and a majority of events could be considered as passenger. Nevertheless, the presence of clonal integrations reflect the expansion of hepatocytes and this process appears to be different in non-tumor tissues and in tumors. In non-tumor tissues, it might be explained by a selective pressure induced by chronic inflammation and the emergence of cirrhotic nodules in the liver but this process seems to be independent from integrations and viral replication as it has already been suggested^{33–35}. Indeed, in our study, clonal integrations in these tissues do not appear to have functional consequences but are located in regions recurrently targeted by integrations (large, highly expressed genes), and are not associated with viral replication. Yet, as non-tumor tissues harboring clonal integrations showed a downregulation of genes involved in inflammatory response, it suggests that some hepatocytes with a selective advantage to escape immune antiviral response might undergo active proliferation and obtain a protective profile against malignant transformation. However, the low number of non-tumor samples with clonal integrations limits this analysis.

Although the majority of HBV integrations are passenger events and do not have any functional consequences, some can be driver and promote HCC initiation in cirrhotic or non-cirrhotic livers. We described in this work several strong oncogenic alterations due to HBV integrations. While integrations had already been associated with copy number alterations in previous studies^{15,20,36}, we identified for the first time recurrent structural rearrangements associated with viral insertions in HCC. Through integrations often located around centromeric or telomeric regions, HBV is involved in large complex structural rearrangements frequently inducing gains of chr8q (including *MYC* locus) and chr5p (including *TERT* locus) and losses of chr17p (including *TP53* locus). In tumors harboring these integration-associated rearrangements, HBV integrations are then associated with alterations of cancer-driver genes located at distance. As HBV DNA integration occurs at double-strand breaks^{25,37}, the presence of HBV integrated sequences delimitating CNA may result from opportunistic mechanism. However, our data showed no correlation between the numbers of HBV integrations and of CNA in tumors, and highlighted the role of specific alterations due to integration-associated rearrangements in promoting selection of hepatocytes and HCC development.

HBV integrations may also drive carcinogenesis by altering the closest gene from the viral integrated sequence through insertional mutagenesis. Our study confirmed that the *TERT* promoter is the main HBV integration hotspot in HCC, as one third of HBV-related HCC in our series harbor a clonal integration at this hotspot. In these tumors, the viral sequence induced a strong activation of *TERT* due to the presence of the viral Enhancers as it had already been reported^{17,38}, at proximity of the ATG and in an orientation-independent manner. In addition, HBV integrations in *CCNA2* or *CCNE1* genes have already been investigated in a previous study of our lab²¹. HBV integrated sequence is either altering directly the structure of the gene (for *CCNA2*) or inducing a strong overexpression due to its Enhancer sequences (for *CCNE1*). Alterations of one of these two genes induce replicative stress and a signature of rearrangements directly promoting the development of a tumor belonging to a specific subgroup of HCC developed on non-cirrhotic livers (CCN-HCC). Finally, even if further investigation are needed to fully understand the functional consequences, tumors with HBV integrations in *KMT2B* are also mainly developing on non-cirrhotic liver with a high expression of progenitor markers, suggesting another strong direct mechanism induced by HBV integrated sequence.

Overall, this study underlines the complexity of the interplay between HBV integrations, HBV replication and chromosomal instability in hepato-carcinogenesis, in addition to other cofactors. On one side, Aflatoxin B1 exposure, mainly present in Africa, may initiate directly HBV-related HCC development through *TP53* R249S mutation³⁹. On the other side, HDV infection limits HBV replication⁴⁰ and viral integrations but accelerates chronic inflammation and fibrosis in young patients. Importantly, our series showed that the number of HBV integrations in tumors is the only marker within viral features associated with poor prognostic.

In conclusion, this study described a new mechanism of alterations induced by HBV integrations in HCC: structural rearrangements associated with HBV integrations may drive carcinogenesis by altering cancer-driver genes at distance, through different mechanisms than *cis*-activating insertional mutagenesis. It underlined the heterogeneity of HBV-related HCC and the importance of molecular characterization to identify new specific therapeutic opportunities.

MATERIAL AND METHODS

Patients and tissue samples

A series of 1170 hepatocellular carcinoma and 1070 non-tumour counterparts from 1128 patients was assembled and the study was approved by the local Ethics Committee (CCPPRB Paris Saint-Louis). Samples were collected in 12 academic hospitals in France (LICA-FR cohort) and frozen immediately at -80°C after resection or biopsy. DNA was systematically extracted, as RNA when the quantity of tissue was sufficient. The majority of tumor samples was from primary tumors except for 12 samples collected at relapse. Clinical features and serological data were gathered from each center and are summarized in **Supplementary Table 3**. A flowchart of the study inclusion at each step is provided in **Supplementary Fig. 10**.

Viral DNA screening

Genomic DNA were screened to detect the presence of HBV DNA by quantitative RT-PCR (qRT-PCR) on Fluidigm 96.96 dynamic arrays using the BioMark Real-Time PCR system with TaqMan probe sets designed with Primer3Plus software. In total eleven probes were designed to detect specifically five viral regions for all the eight main HBV genotypes (**Supplementary Table 4**). Results were analysed using the Fluidigm Real-Time PCR Analysis software (V.4.1.3) and reported to HMBS as reference gene. A series of plasmids with known concentration was used to perform a quantification of the results. The values obtained were tested for bimodal distribution using normalmixEM function of mixtools package in R⁴¹.

HBV viral capture

Viral capture was performed for a selection of 220 HCC and 177 non-tumor liver tissues from 177 patients with a known HBV etiology. 14 tumors and 8 non-tumor liver tissues from 13 HBV-negative patients were used as negative controls. Single-stranded biotinylated probes were designed with Roche NimbleGen (SeqCap EZ Designs, Roche NimbleGen Inc., Madison, WI, USA), to target 40 HBV genotypes and 4 human regions (TERT promoter, CCNE1, CCNA2 and KMT2B). The reference of sequences used is shown in **Supplementary Table 5**.

Samples containing 1 μg of DNA were used for DNA library preparation. DNA was sheared mechanically in fragments with an average length of 1kb, before following the SeqCap EZ HyperCap Workflow developed by Roche. We used KAPA Dual-Indexed Adapters for sample indexing and we multiplexed library samples by using one capture probe set for 12 to 42 samples. DNA samples were selected for multiplexing based on the HBV copy number per cell measured by qRT-PCR, in order to have similar viral loads between all library samples in one capture pool. A sample not containing HBV DNA was added in each pool as negative control. We performed a double capture with two successive hybridization steps to enrich twice for DNA targets and increase the specificity of capture, following an adapted protocol developed by Roche. Final libraries were quantified with KAPA Library Quantification Kit and controlled with a DNA High Sensitivity assay on Caliper LabChip GX. Two final libraries were merged to be sequenced using an Illumina MiSeq instrument with paired-end reads of 2×250 nt.

Analysis of integration events and clonality definition

Raw capture data obtained after sequencing were first aligned on the reference used (40 HBV sequences and 4 regions of HG19 genome) using Burrows-Wheeler Aligner (V.0.7.15)⁴². A

normalization factor was assessed based on the mean coverage of the targeted human regions and was corrected to remove capture biases. The main HBV genotype of each sample was determined as the sequence with the higher number of mapped reads. Read pairs with at least one read aligned on a HBV sequence were extracted using samtools (V.1.3)⁴³, and realigned to a custom reference genome including the HG19 reference and the HBV sequence of the main genotype identified for each sample (one single reference was chosen for each of the 8 genotypes A-H) (**Supplementary Table 5**).

HBV copy number per cell was assessed from the ratio (multiplied by 2) of the mean read coverage of the HBV genome and the normalization factor, and was compared with the measures obtained by qRT-PCR (**Supplementary Fig. 11a**). We identified the HBV/human chimeric regions with at least one chimeric read supporting the junction and one read pair with a read mapping on HBV and its mate mapping on HG19. Coverage of reads were computed with bedtools multicov utility⁴⁴ and the characteristics of junction breakpoints (position and sequence) were extracted from hard clipped reads with htlib package. Clonality of integration events was assessed from the ratio (multiplied by 2) of the coverage of reads at the junction breakpoint and the normalization factor. All viral insertions with a clonality greater than 0.03 were validated by visual inspection on IGV (Integrative Genomics Viewer). Similar integration events (same position on HG19 and on HBV and same sequence) found in two independent samples within one single capture (and with at least a 10-fold difference in clonality) were considered as contaminations intra-capture and the event with the lowest clonality was removed from the analysis. Similar integration events found in the non-tumor and tumor tissues of the same patients were inspected specifically and considered as contaminations based on the very few number of reads detected in the non-tumor tissues compared to the tumors. Integration events located in repeated regions of the human genome with the same breakpoint and same orientation in HBV genome were considered as one unique event. The 9183 integration events identified were classified in 3 categories using the k-means method: (1)“clonal integrations” observed in more than 48% of cells, (2)“unique integrations” observed in less than 3% of cells and (3)“subclonal integrations” with intermediate copy numbers (**Supplementary Fig. 11b**).

Integration events identified from viral capture were compared with those identified from WGS for 22 tumors and 20 non-tumor liver tissues (7 pairs tumor/non-tumor had already been published) (**Supplementary Fig. 11c-11d**). WGS was performed and analyzed as already described⁴⁵. The bioinformatics analysis of HBV integrations was similar for WGS as for capture except for the normalization factor that was corresponding to the mean sequencing coverage for WGS (30x for non-tumor tissues and 60x or 90x for tumors).

We annotated the integration breakpoints with the following genomic features: replication timing in HepG2 cell line (ENCODE⁴⁶), highly expressed (top 20%) genes in non-tumor liver from HBV-positive patients, common and rare fragile sites⁴⁷, repetitive sequences and CpG islands, and chromatin structure in adult liver (ROADMAP⁴⁸). We defined an insertion as classic if we identified two breakpoints within 25kb in opposite direction, as local inversion if we identified two breakpoints within 25kb in the same direction, as large rearrangements if we identified one breakpoint and as cluster of insertions if we identified more than two breakpoints.

Genomic DNA sequencing of driver genes

A selection of 780 tumor samples were sequenced with WGS (for 62 tumors with their adjacent liver tissues: 43 already published and 19 new cases), Whole Exome Sequencing (WES, for

203 tumors with their adjacent liver tissues: 234 already published and 69 new cases), or targeted sequencing of 15 to 31 genes using MiSeq (for 367 tumors) or Sanger (for 148 tumors). The genes targeted were TERT promoter, CTNNB1, TP53, ARID1A, ALB, APOB, AXIN1, MLL2, ARID2, ATM, BAP1, CDKN2A, NFE2L2, RPS6KA3, KEAP1, SETD2, MAP1B, CDKN1A, TSC1, TSC2, HNF1A, PIK3CA, IL6ST, GNAS, IRF2, KRAS, STAT3, JAK1, BRAF, HRAS, and NRAS. Protocols and primers of Sanger and MiSeq sequencing as well as the protocol and method of analysis of whole-exome sequencing have previously been described^{26,45,49}. In particular, all tumor samples without WGS were re-sequenced to search for TERT promoter mutation with Sanger or MiSeq sequencing as previously described⁵⁰.

Detection of HBV episomal form

A specific DNase/TaqMan-based assay was adapted from protocol by Werle-Lapostolle et al to detect HBV episomal form and performed as previously described for AAV2 episomal form detection⁵¹. All samples from the Capture series with enough DNA material were screened for the presence of HBV episomal form: 198 tumors and 155 non-tumor liver tissues from 169 patients. The HBV probes used are listed in **Supplementary Table 4**. When the results were not conclusive, a second step of digestion was performed. The difference between HBV Ct values without and with one or two steps of PS-DNase digestion was analyzed to determine the presence of viral episomal form, or the digestion of the linear viral DNA.

Viral mRNA and specific genes mRNA screening

For viral mRNA, eight specific probe sets were designed to detect HBV mRNA and are listed in **Supplementary Table 4**. Probe sets also designed to detect HCV and HDV mRNA. For TERT expression analysis, we used human catalogue TaqMan probes. We performed qRT-PCR using BioMark Real-Time PCR system. Expression data were normalized with the $2^{-\Delta Ct}$ method relative to ribosomal 18S (Hs03928990_g1). Five normal tissues were used as reference.

RNAseq and transcriptomic analysis

A selection of 265 tumors (130 HBV-positive and 135 HBV-negative) and 25 HBV-positive non-tumor tissues were sequenced using Illumina TruSeq or Illumina TruSeq Stranded mRNA kit on HiSeq2000 sequencer by IntegraGen, Evry, France²¹. Among the 25 non-tumor samples, we used the Bioconductor limma package⁵² to test for differential expression, between samples harboring a HBV clonal integration and other HBV-positive non-tumor tissues, of all genes expressed. We applied a q-value threshold of ≤ 0.05 to define differentially expressed genes. We used an in-house adaptation of the GSEA method⁵³ to identify gene sets from the MSigDB (v. 6.0) database overrepresented among up- and down-regulated genes.

In addition, qRT-PCR of 190 genes using BioMark Real-Time PCR system was performed on the 220 HBV-positive tumors sequenced in viral capture, in order to classify them in the G1-G6 classification as previously described for the LICA-FR cohort^{54,55}.

HBV variants analysis

From viral capture, we extracted the number of each base per position of the HBV sequence. The positions on all HBV genotypes were modified to use the EcoR1 restriction site as the +1. We considered the positions 1762 and 1764 of the Basal Core Promoter region, 1896 of the PreCore region and the positions 646, 667, 670, 671, 739, 741, 836 and 877 of the Reverse transcriptase (RT) domain of HBV polymerase. For RT mutants, we considered the mutations

inducing the following amino acids changes: V173L, L180M, A181T, A181V, M204V, M204I, N236T and M250V^{24,56}. Only samples with a coverage greater than 20 reads at the selected position were used for analysis. A sample was considered as mutated if the number of reads harboring the mutation was greater than 20 or represented more than 10% of the total coverage.

Cell culture, transfection and dual luciferase assay

HuH7 cells were purchased from the American Type Culture Collection (ATCC) and cultured in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum and 100U/mL penicillin/streptomycin. Cells were co-transfected using Lipofectamine 3000 (Life Technologies) with a pGL3 plasmid containing the wild-type TERT promoter or promoter with the two hotspot mutations, or different HBV sequences (normal or scrambled) controlling a firefly luciferase reporter gene and a plasmid encoding Renilla luciferase (Promega). Luminescence from firefly luciferase was normalized on the corresponding Renilla luciferase activity as previously described⁵⁷. A fold change was then calculated relative to the values obtained for the construct containing the wild-type TERT promoter.

Long-read sequencing

Three HBV-positive tumors were selected to perform long-read sequencing using the technology SMRT (single molecule real time technology) with a PacBio-SMRTcell system (ICGex NGS platform, Curie Institute, Paris, France). Samples were selected based on the results of Fragment Analyzer to have an average length of DNA fragments of 10kb. Libraries were prepared based on the PacBio template Prep kit protocol. For analysis, error-prone raw subreads were firstly merged to obtain unique polymerase reads with the Consensus Circular Sequencing (CCS2) algorithm (Pacific Biosciences). The consensus reads obtained were aligned using minimap2 aligner (V.2.16)⁵⁸ on a custom reference genome including the HG19 reference and the HBV sequence of the main genotype previously identified for each sample.

Bionano whole-genome mapping

Two pairs of tumors and non-tumor liver tissues were selected to perform whole-genome mapping using Bionano technology (Bionano Genomics, La Jolla, California, USA). 30mg of tissue were used to extract long molecules of DNA, labeled with Bionano reagents by incorporation of fluorophores at a specific sequence motif along the genome. The labeled genomic DNA was linearized in the SaphyrChip using NanoChannel arrays and single molecules were imaged and digitized. As molecules are uniquely identifiable by distinct distribution of sequence motif labels, they were then assembled by pairwise alignment into de novo genome maps. Genome-wide Structural Variant (SV) calling was performed in tumor samples by aligning these maps and molecules to a reference genome, and SV calls are annotated using the corresponding non-tumor tissue from the same patient.

Statistical analysis

For statistical analysis, we used R version 3.6.0 (R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, (<http://www.R-project.org>). Wilcoxon, Fisher, Chi-square or Pearson's correlation statistical tests were applied with respect to the type of variable. Survival analysis was performed in patients treated for a primary HCC tumor by R0 liver resection as previously described²⁶. We assessed overall survival defined by the interval between surgery and death. Survival curves

were represented using the Kaplan-Meier method compared with the Log Rank test. Multivariate analysis was performed using Cox model. A p value <0.05 was considered as statistically significant.

Data availability

The sequencing data of the LICA-FR cohort reported in this paper have been deposited to the EGA (European Genome-phenome Archive) database (RNA-seq accession [EGAS00001001284], [EGAS00001002879], [EGAS00001003310], [EGAS00001003837] and [XXX]; WES accessions [EGAS00001000217], [EGAS00001001002], [EGAS00001003063], [EGAS00001003130], [EGAS00001003837] and [XXX]; WGS accessions [EGAS00001000706], [EGAS00001002408], [EGAS00001002888], [EGAS00001003063], [EGAS00001003837] and [XXX]), through the ICGC (International Cancer Genome Consortium) data access committee.

Author contributions:

J.Z-R. conceived and directed the research. C.P. and J.Z-R. designed the study and wrote the manuscript. C.P. and T.L-B. performed the experiments. C.P., S.I., T.Z.H., S.C., E.L. and J.Z-R. analyzed and interpreted the data. J.C., V.P., J-F.B., J-C.N. and G.A. provided essential biological resources and collected clinical data. All authors approved the final manuscript and contributed to critical revisions to its intellectual context.

Acknowledgments:

We thank Alain Nicolas, Sylvain Baulande and Sonia Lameiras at Institut Curie for their help in analyzing PacBio results and setting up viral capture. We thank Quentin Bayard and Karl Hong for their help in analyzing Bionano sequencing results. We thank Gabrielle Couchy, Iadh Mami, Bénédicte Noblet, Massih Ningarhari, Jill Pilet and Jie Yang for their help in molecular biology experiments. This work was supported by ANRS (French national agency for research on AIDS and viral hepatitis). The group is supported by the Ligue Nationale contre le Cancer (Equipe Labellisée), Labex OncoImmunology (investissement d’avenir), grant IREB, Coup d’Elan de la Fondation Bettencourt-Shueller, the SIRIC CARPEM, FRM prix Rosen, Ligue Contre le Cancer Comité de Paris (prix René et André Duquesne) and Fondation Mérieux. C.P. was supported by a fellowship from ANRS (French national agency for research on AIDS and viral hepatitis). T.L-B was supported by an “Attractivite IDEX” fellowship from IUH.

Competing interests: Authors have no conflict of interest to declare.

Abbreviations:

HBV: hepatitis B virus; HCC: hepatocellular carcinoma; RNAseq: RNA-sequencing, BCP: basal core promoter; RT: reverse transcriptase; PC: precore; dsLDNA: double-strand linear DNA; CNA: copy-number alteration; WES: whole exome sequencing; WGS: whole genome sequencing; HDV: hepatitis D virus; AFB1: aflatoxin B1; AA: aristolochic acid; NGS: new-generation sequencing

Writing assistance: no writing assistance

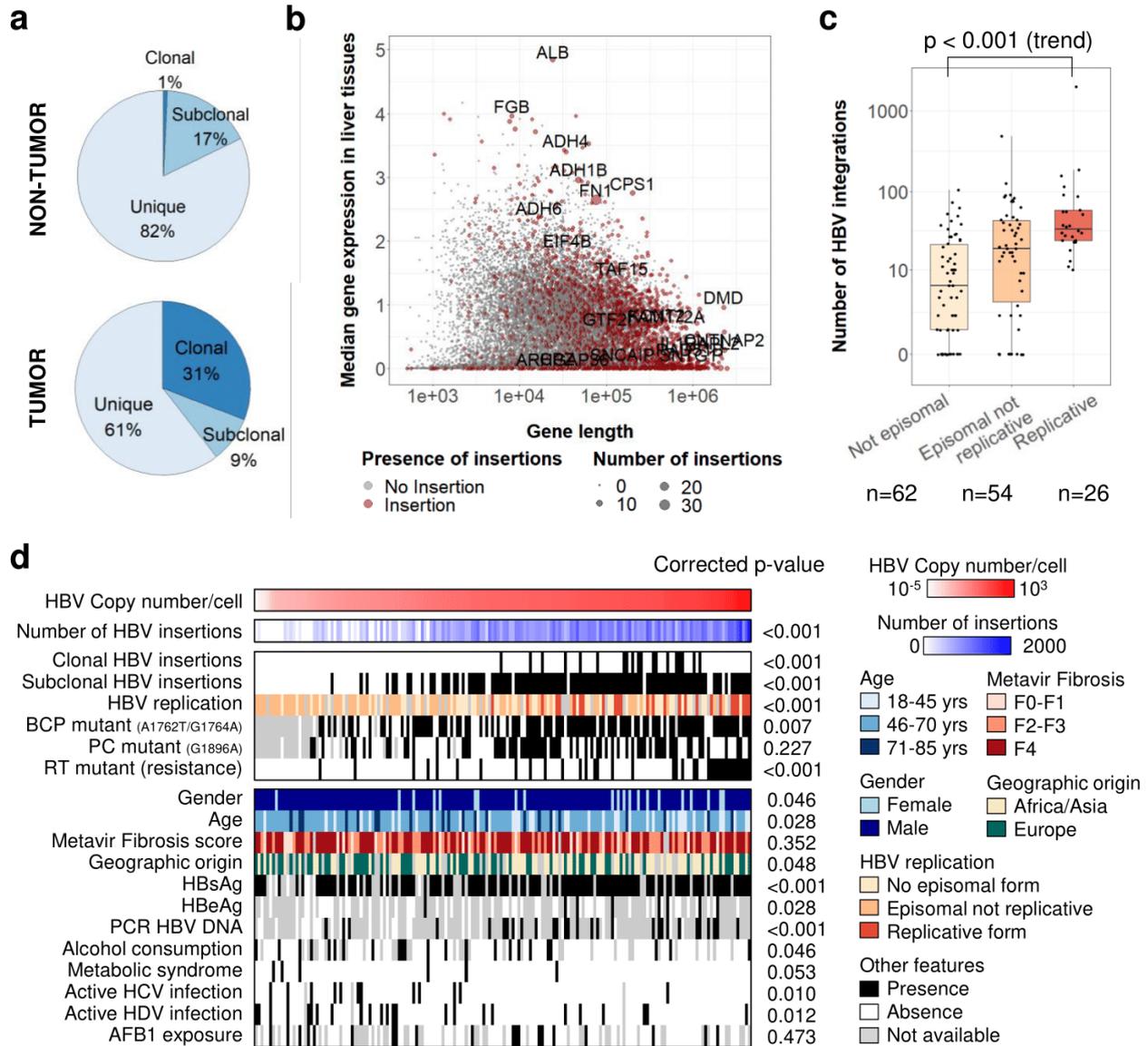


Figure 1 | HBV integrations in non-tumor tissues are associated with viral replication and more frequent in large highly expressed genes.

(a) Repartition of clonality between all HBV integration events detected in viral capture (n=9183). (b) Coding-genes with or without HBV integrations according to the gene length and the median gene expression in non-tumor liver tissues. Genes with recurrent clonal or subclonal HBV integrations were annotated. (c) Number of HBV integrations identified in non-tumor samples (n=142) according to the presence of episomal HBV DNA and replicative HBV DNA (Jonckheere’s trend test). (d) Correlations between HBV copy number/cell in 170 non-tumor liver tissues assessed by viral capture and clinical or molecular features. Positivity for HBsAg, HBeAg, HBV DNA (by PCR) in the patients’ serum and duration of antiviral treatments were obtained from clinical data. Wilcoxon signed-rank, Kruskal-Wallis or Pearson’s correlation statistical tests were applied with respect to the type of variable. P values were adjusted for multiple testing using the Benjamini-Hochberg method (false discovery rate). *AFB1*, Aflatoxin-B1; *BCP*, Basal Core Promoter; *PC*, PreCore; *RT*, Reverse-Transcriptase; *ns*, not significant.

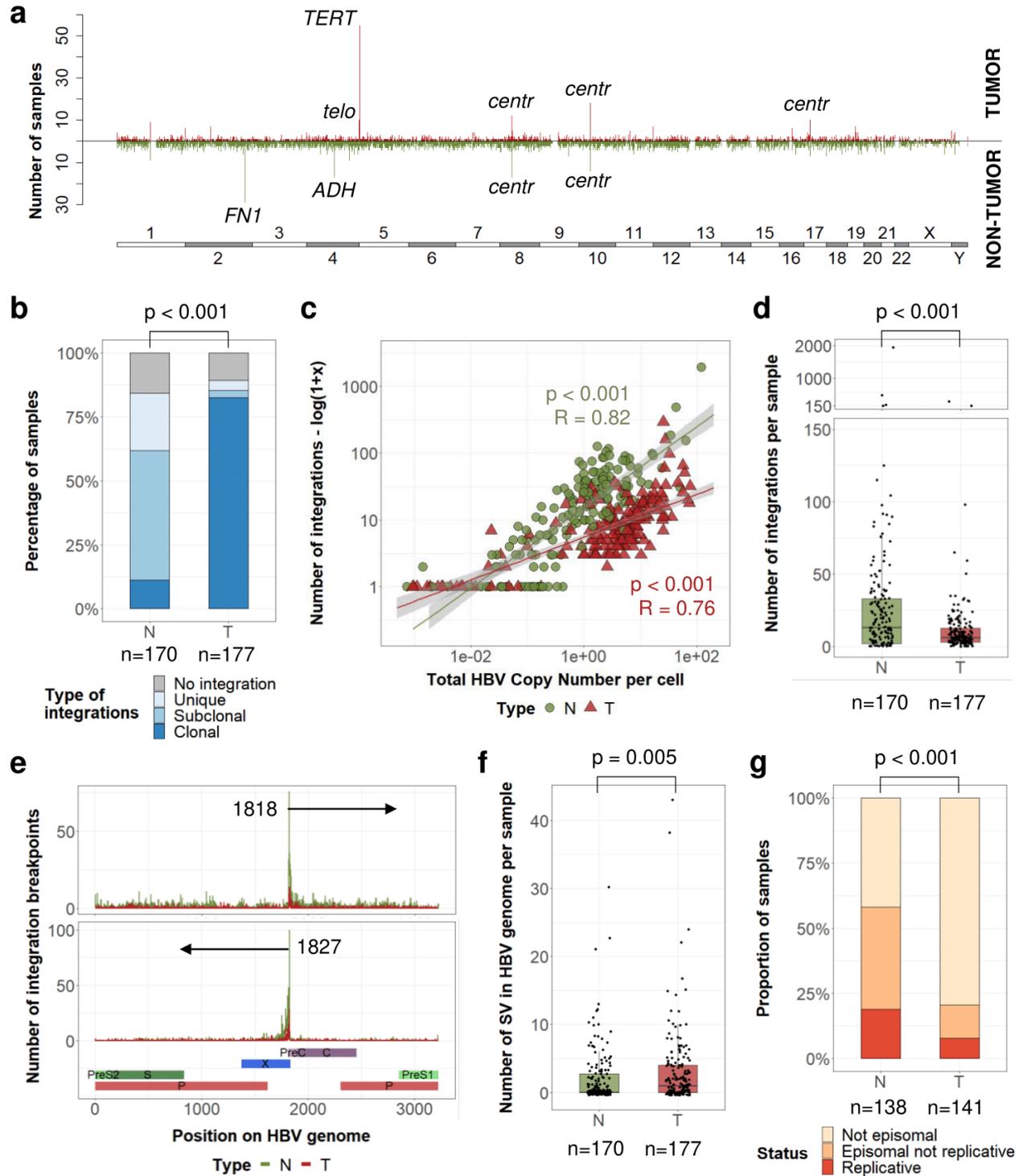


Figure 2 | Tumors and non-tumor tissues have different HBV integrations profiles and structures of HBV sequences

(a) Pan-genomic view of genomic locations of HBV integration breakpoints in tumors (up) or non-tumor tissues (down) in 347 HBV-positive samples from the Capture series (non-tumor liver, n=170, and tumor samples, n=177). A line corresponds to a 1M-bin region. (b) Proportion of non-tumor tissues and tumors harboring only unique integrations, at least one subclonal integration, or at least one clonal integration (Chi-square test). (c) Correlation between the HBV copy number per cell and the number of HBV integration breakpoints (Pearson correlation). (d) Number of HBV integration breakpoints per sample (Wilcoxon signed-rank test). (e) Localization of HBV integration breakpoints along the HBV genome in tumors and non-tumor samples according to the orientation of the integrated sequence. (f) Number of structural variants in HBV genome per sample (Wilcoxon signed-rank test). (g) Proportion of non-tumor and tumor samples containing replicative HBV DNA, episomal not replicative HBV DNA or no HBV episomal form (Chi-square test). *centr*, centromeric; *telo*, telomeric; *SV*, Structural Variant.

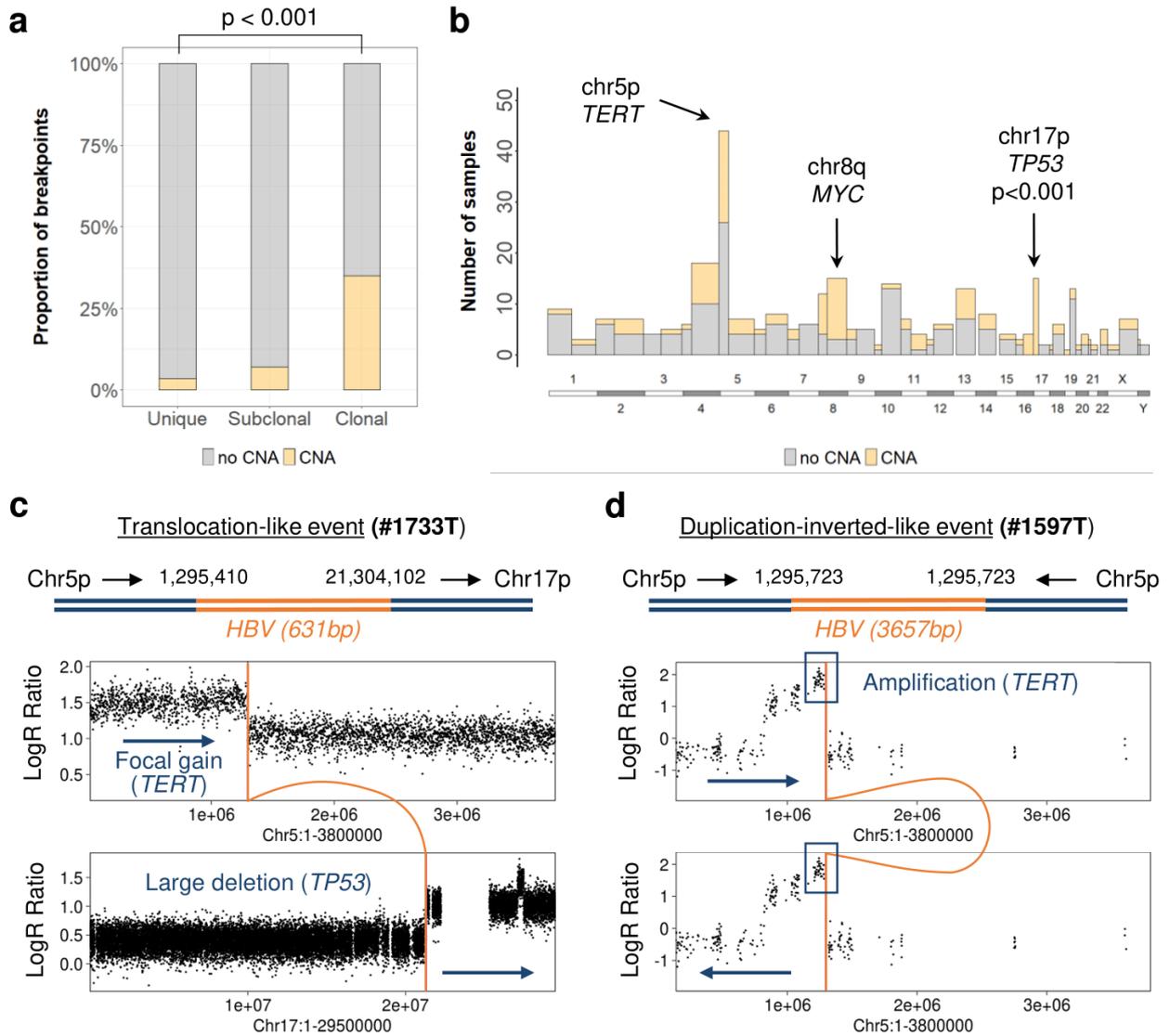


Figure 3 | HBV integrations induce chromosomal rearrangements and distant driver oncogenic alterations

(a) Proportion of HBV integration breakpoints (n=1503) associated with a Copy Number Alteration (CNA) in 121 HBV-positive tumors from the Capture/NGS series (Chi-square test). (b) Pan-genomic view of chromosome arms of HBV integration breakpoints according to their association with a CNA. The three genomic regions containing the higher number of HBV integrations associated with CNA were annotated. Fisher tests were performed to compare the number of integrations with or without CNA and p-values were adjusted for multiple testing. (c) Translocation-like event in tumor #1733T: HBV integration is associated with a focal gain on chr5p and a large deletion on chr17p. (d) Duplication-inverted-like event in tumor #1597T, reconstructed with long-read sequencing: HBV integration is associated with a focal amplification on TERT.

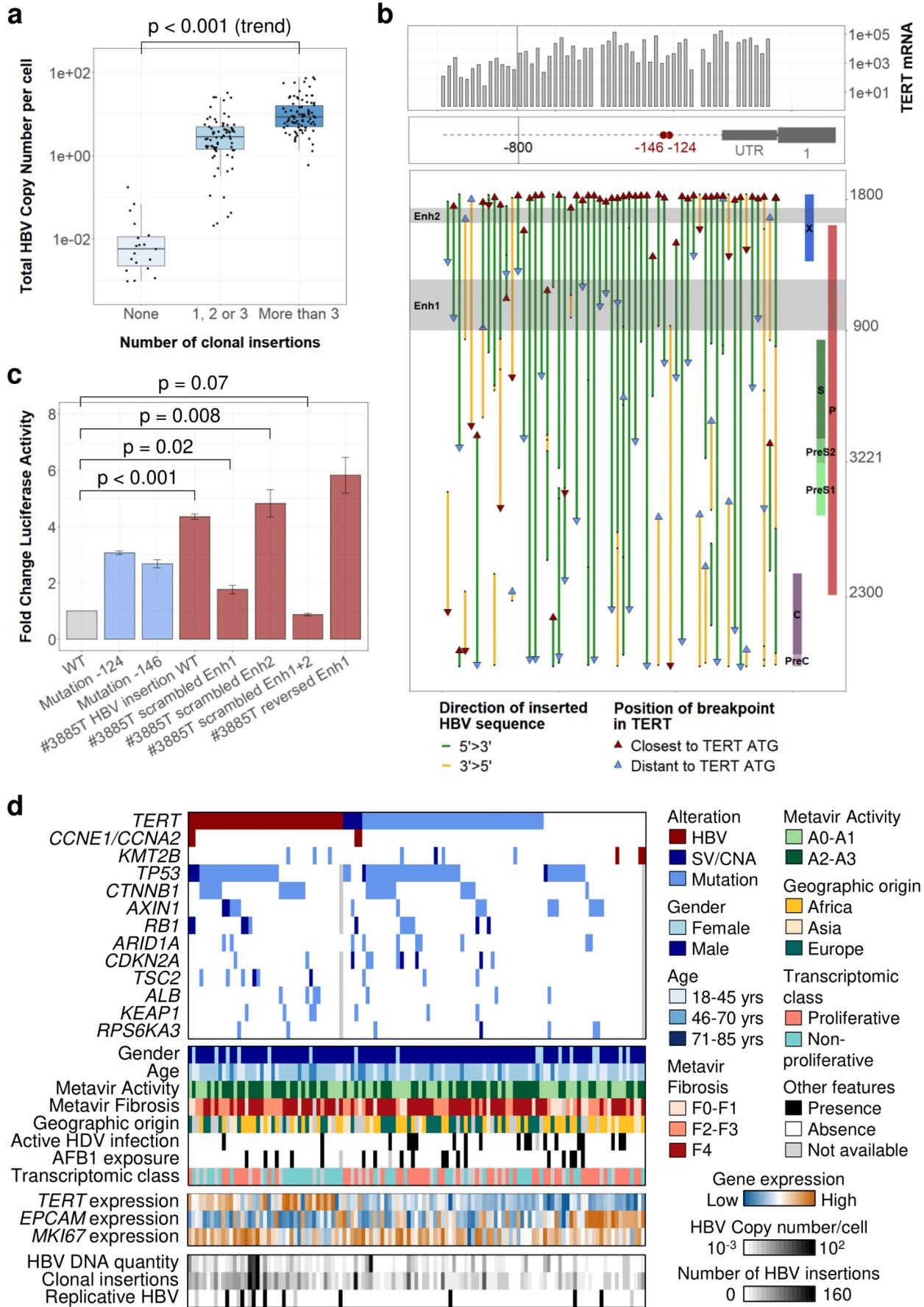


Figure 4 | HBV integrations in the TERT promoter induce a strong activation of TERT promoting HCC development

(a) HBV copy number/cell in 177 tumors from the Capture series according to their number of clonal HBV integration breakpoints (Jonckheere's trend test). (b) Integrated HBV sequences located in the TERT promoter in 57 tumors from the Capture series. The position of the breakpoints, of the viral Enhancer regions and the orientation of the integrated sequences are annotated along the HBV genome. mRNA expression (-ddCT) from RT-qPCR data are represented above. (c) The impact of HBV integration in the TERT promoter was evaluated using promoter luciferase assays in Huh7 liver cell lines. Constructs of TERT promoter containing different HBV integrated sequences with or without scrambled Enhancer regions were compared to the wild-type promoter (WT) and to the TERT promoter with mutations at the -124 or -146 hotspots. Error bars correspond to s.d. of three independent transfections for each plasmid (Student's t test). (d) Molecular profile of 121 HBV-positive tumors with alterations in 14 HCC-associated genes (HBV integration, SV/CNA or mutation), according to clinical and other molecular features. mRNA expression of TERT, EPCAM and MKI67 were obtained from RNA-Seq data. *Enh*, Enhancer; *WT*, Wild Type; *SV*, Structural Variant; *CNA*, Copy Number Alteration; *AFB1*, Aflatoxin B1.

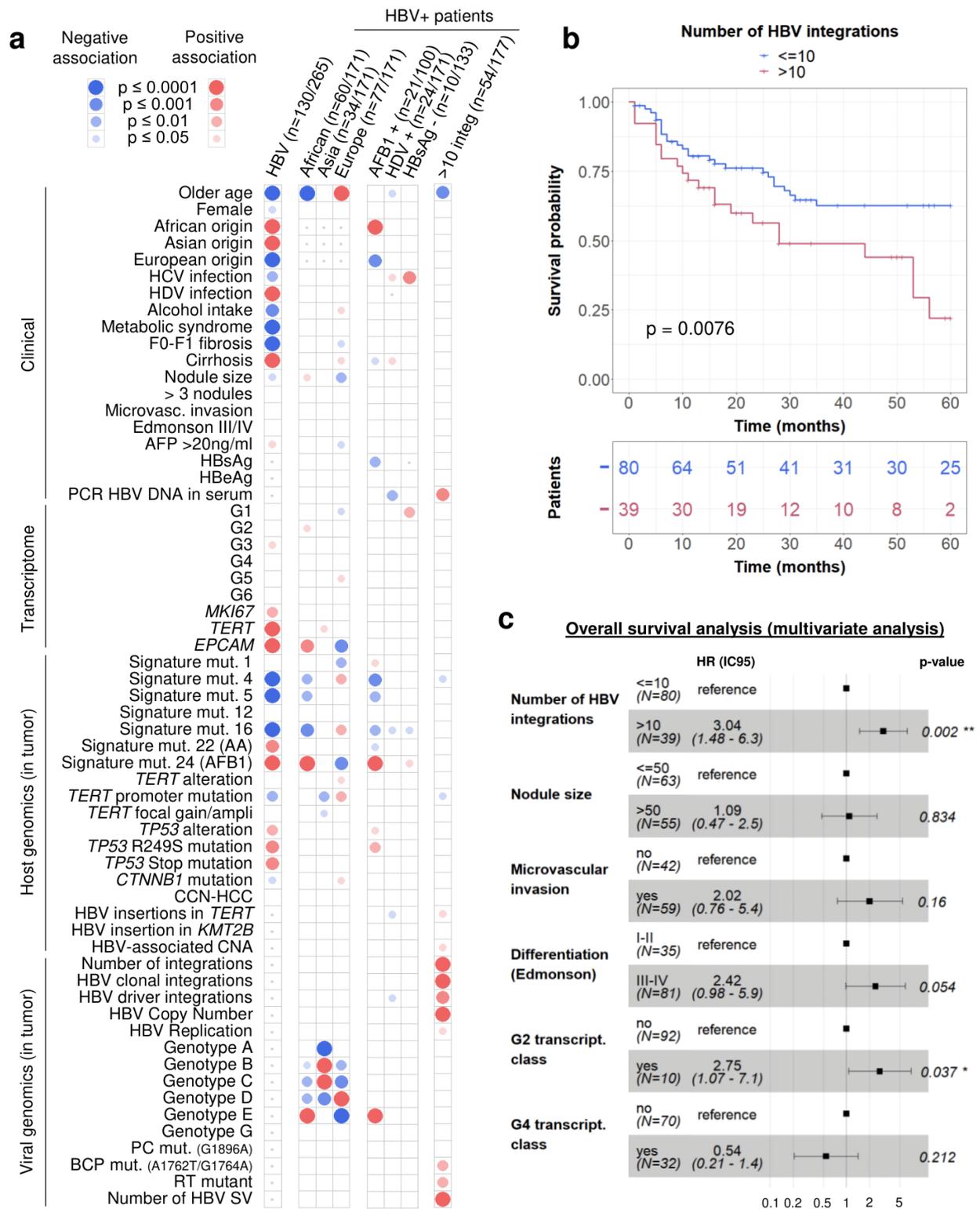
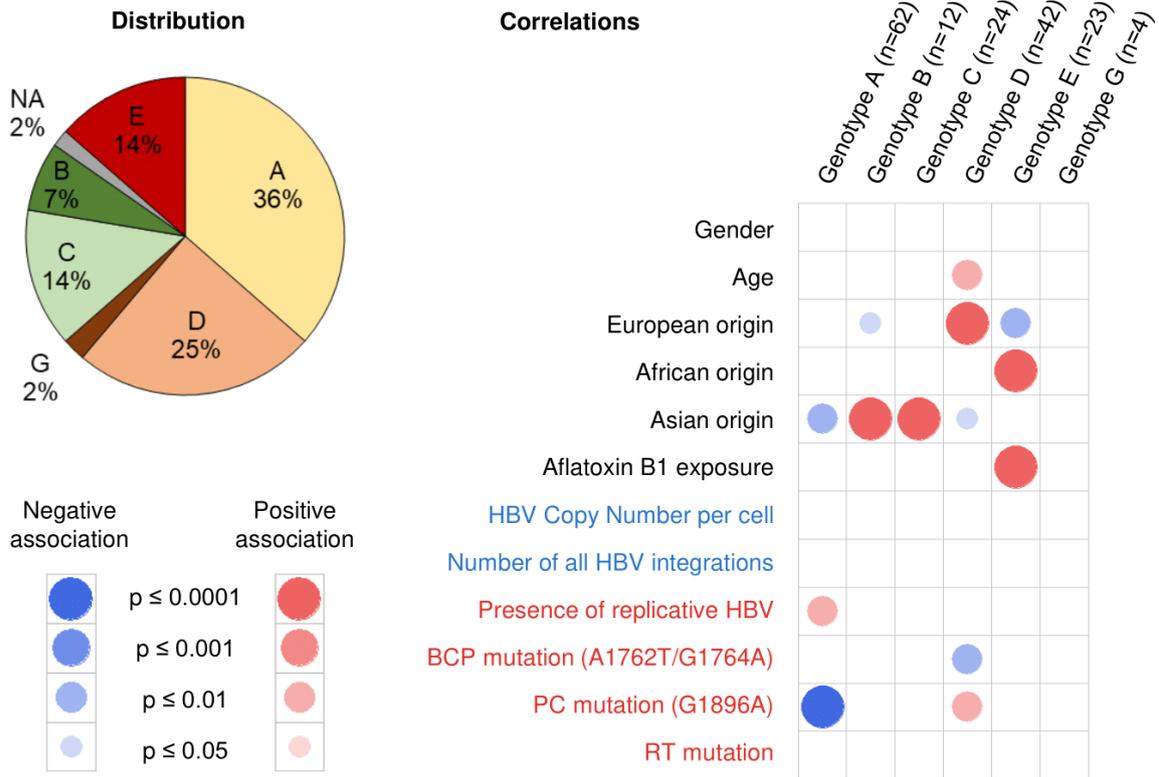
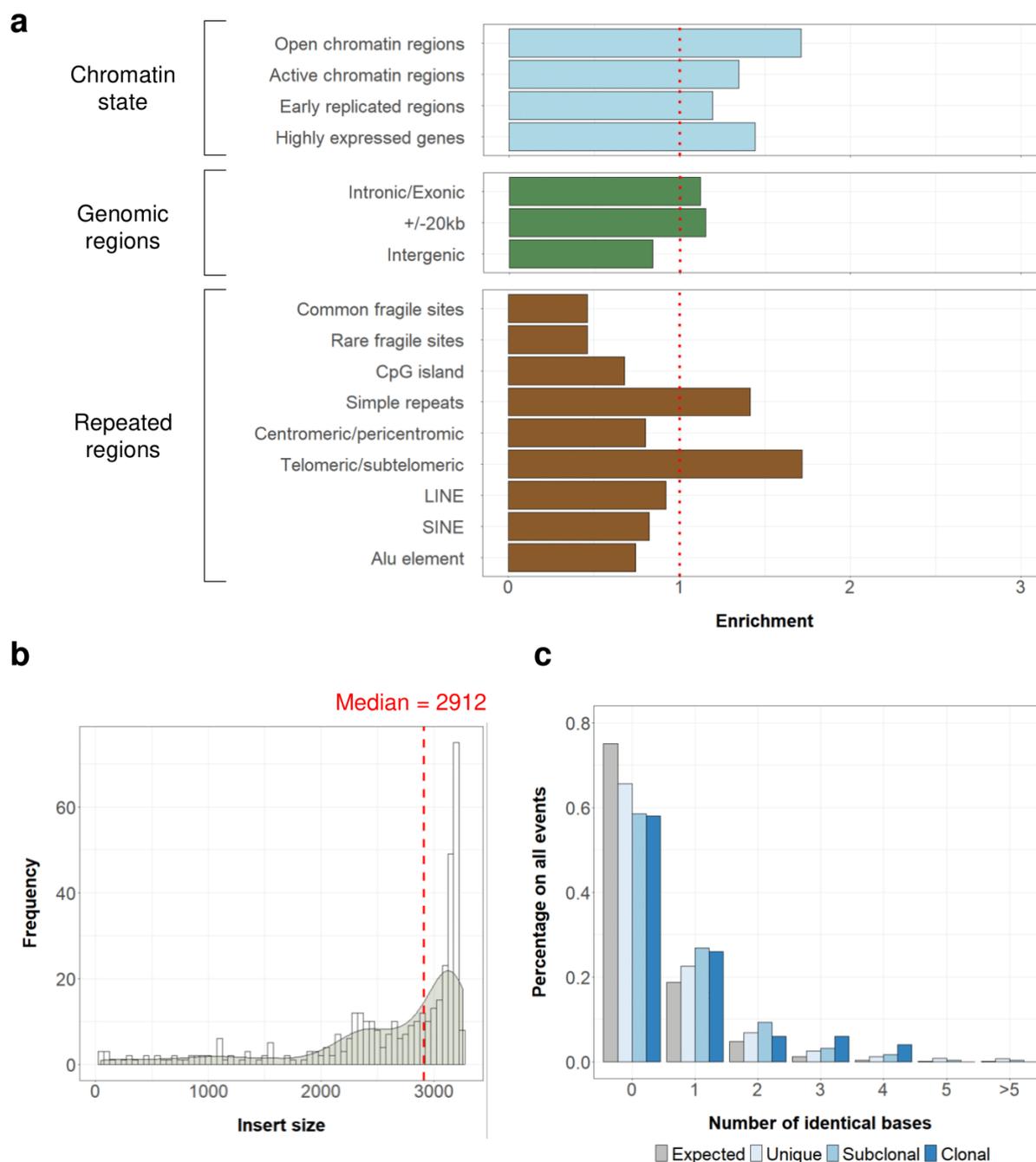


Figure 5 | Integrative analysis reveals a high number of HBV integrations is associated with poor survival of patients

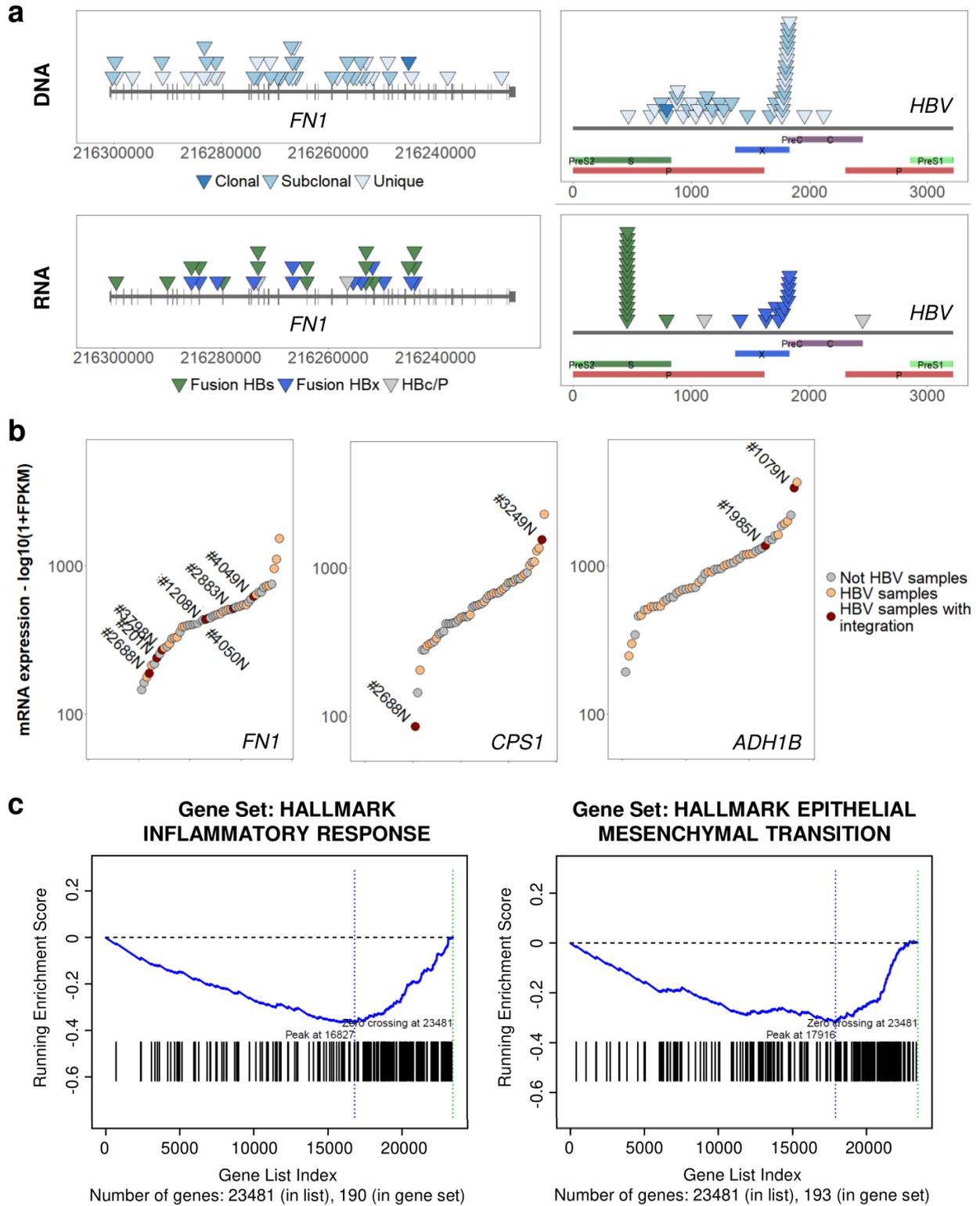
(a) Mosaic plot: association of HBV infection with patient, tumor and viral characteristics in a series of 265 HCC (NGS series) and association of geographic origin, Aflatoxin B1 exposure, HDV infection, HBsAg negativity or number of HBV integrations with patient, tumor and viral characteristics in a series of 177 HBV-positive HCC (Capture series). Blue and red circles indicate negative and positive associations respectively. Color intensities represent different levels of statistical significance. Statistical analysis was performed using chi-square test, Wilcoxon signed-rank test or Pearson correlation with respect to the type of variable. (b) Kaplan-Meier curves for 5-year overall survival from 119 patients after curative R0 resection (n=119 patients). (c) Multivariate Cox regression model for overall survival analysis. HR, Hazard Ratio; IC95, 95% confidence interval.



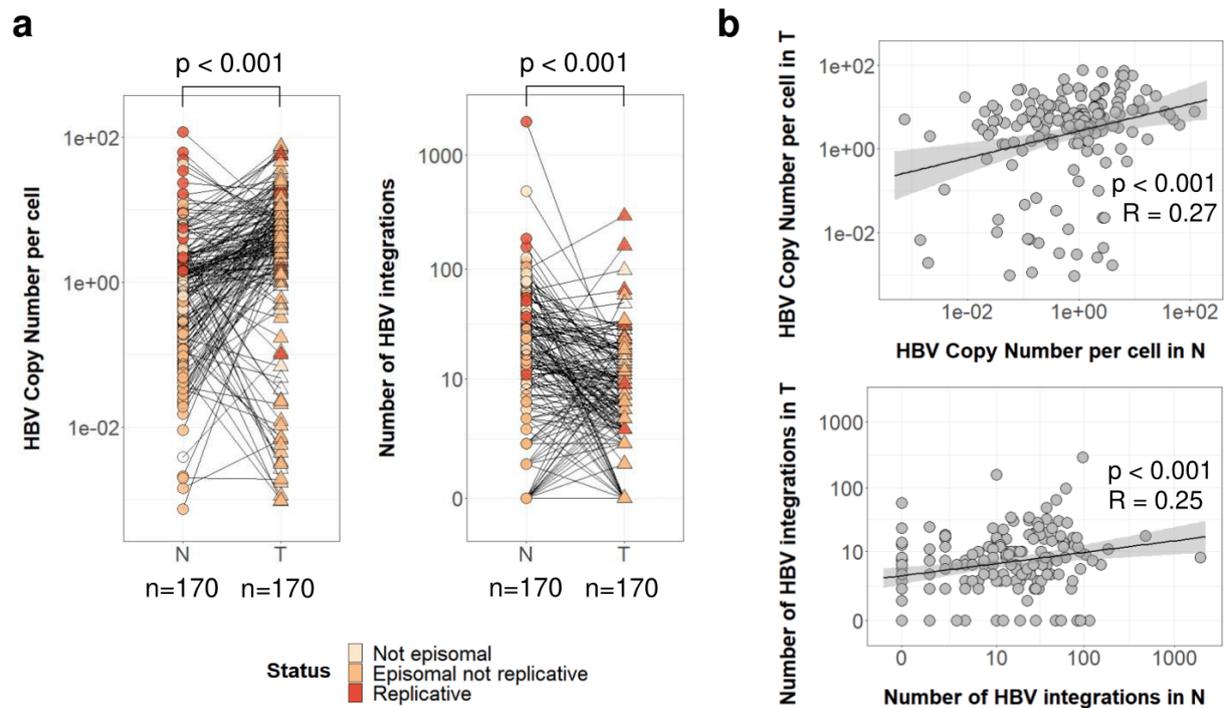
Supplementary Figure 1 | HBV genotypes in non-tumor tissues. Correlation plot between the HBV genotypes identified from viral capture data and clinical or molecular features of non-tumor samples from the capture series (n=170). Blue circles indicate a negative association between features; red circles indicate positive associations. Color intensities represent different levels of statistical significance. Statistical analysis was performed using the chi-square test or Wilcoxon signed-rank test with respect to the type of variable. P values were adjusted for multiple testing using the Benjamini-Hochberg method (false discovery rate). A pie chart showing the distribution of HBV genotypes within the series is represented on the right side.



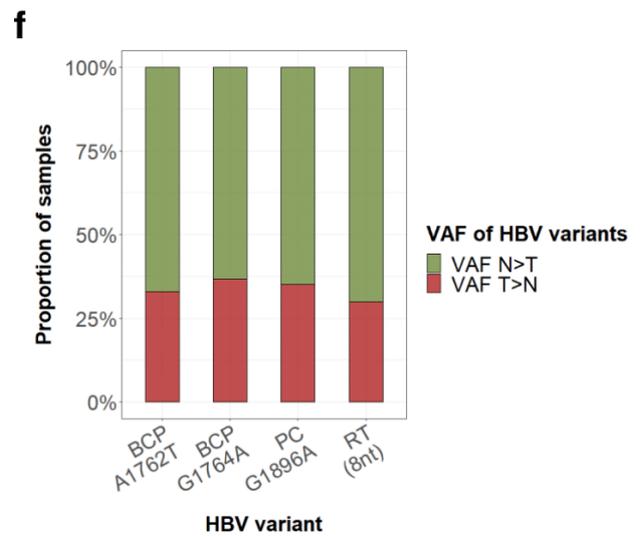
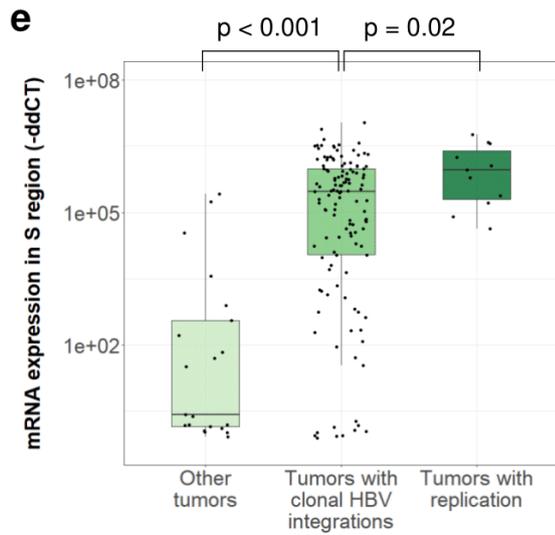
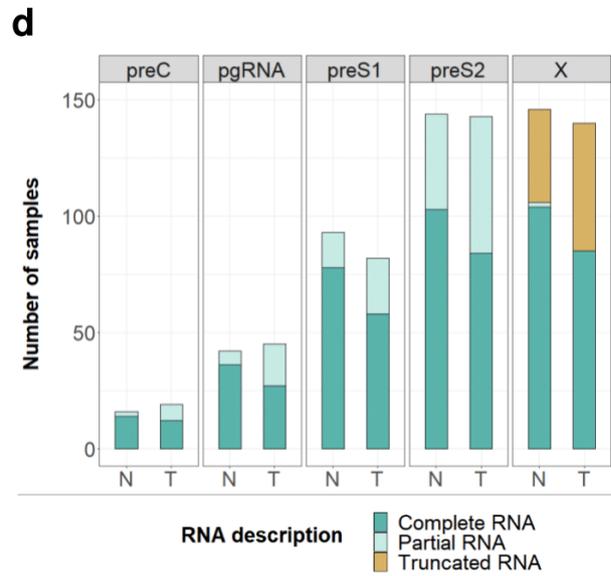
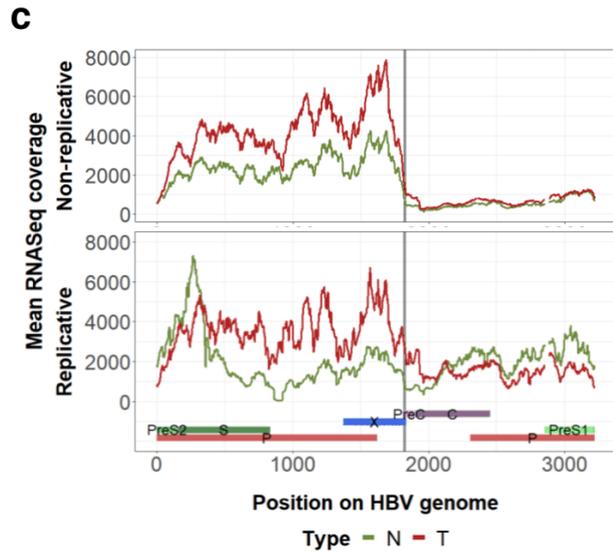
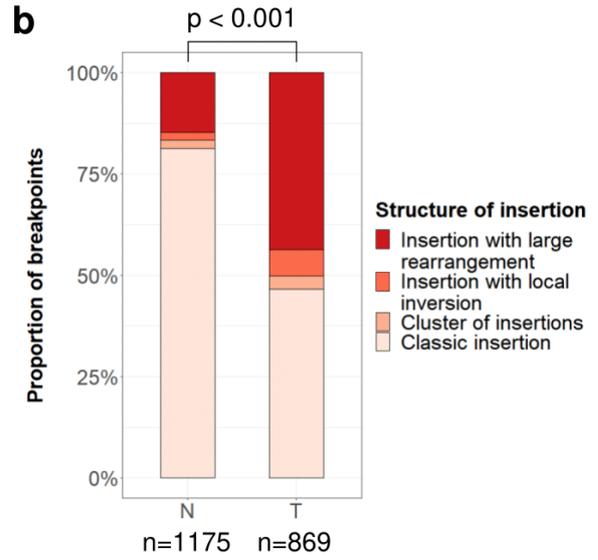
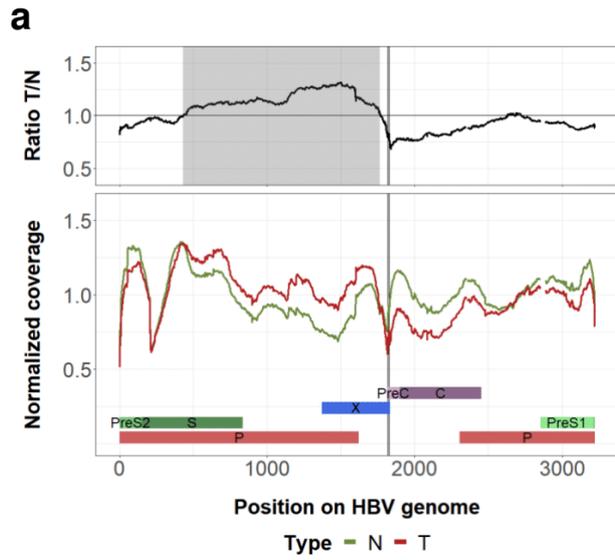
Supplementary Figure 2 | Characterization of localization and sequences of HBV integrations in non-tumor samples. (a) Enrichment of HBV integration breakpoints (n=6454) based on their localization in different regions of the human genome defined by chromatin state, genomic features and repeated motifs. (b) Size distribution of integrated sequences in simple subclonal or clonal integrations (n=394) identified in non-tumor tissues from the Capture series. (c) Homology analysis at HBV integration breakpoints (n=6590) between human genomic sequences and HBV integrated sequences, according to the clonality of the events.



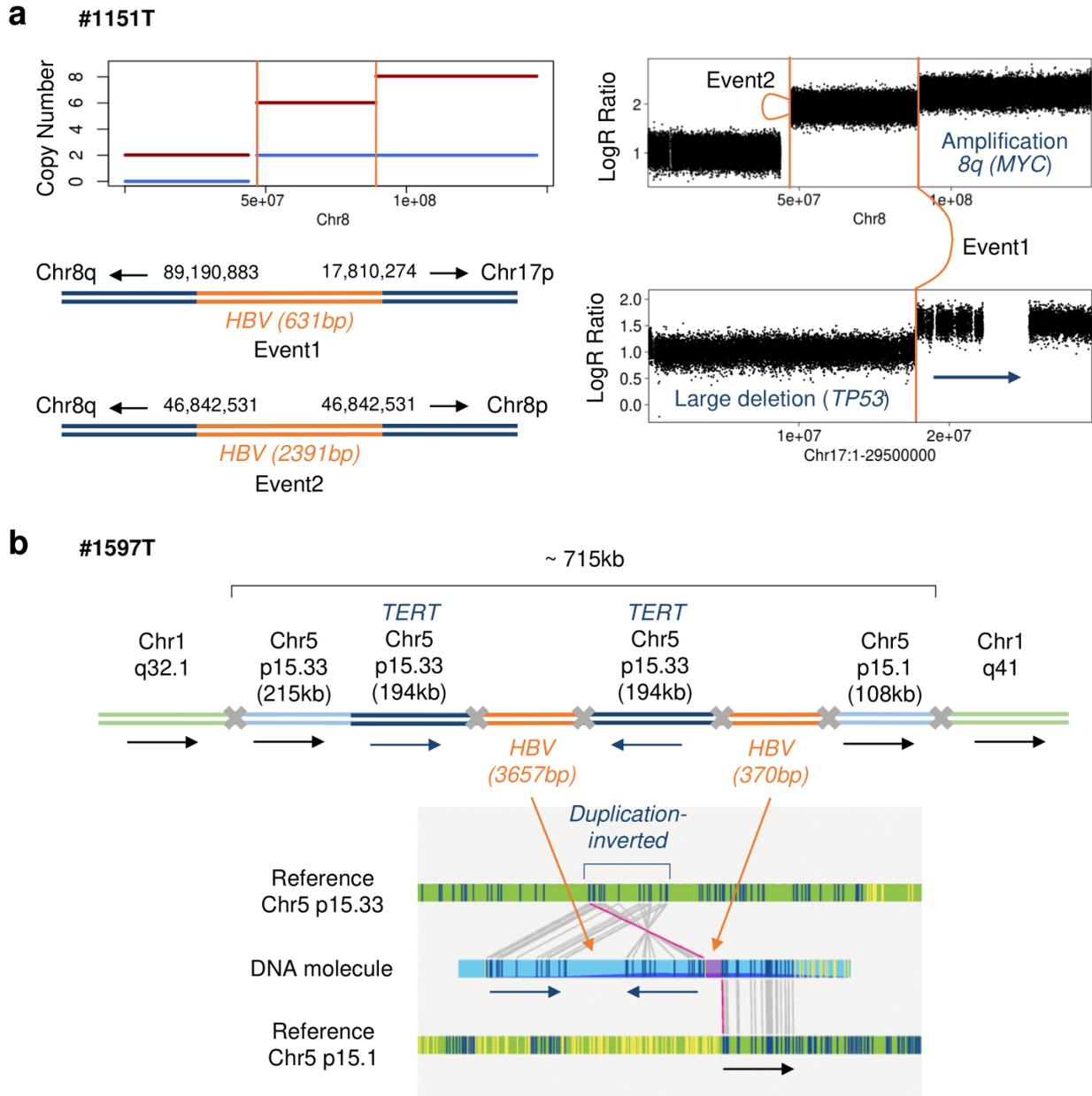
Supplementary Figure 3 | Subclonal/clonal expansion in non-tumor tissues. (a) Breakpoints of HBV integrations in human genome at the *FN1* locus (left) and in HBV genome (right). Breakpoints location were identified from DNA samples in viral capture (up) and RNA samples in RNA-Seq (down). (b) mRNA expression for *FN1*, *CPS1* and *ADH1B* genes from RNA-Seq data, for 25 HBV-positive and 29 HBV-negative non-tumor liver samples. (c) Gene-set enrichment analysis from RNA-Seq data, to compare HBV-positive non-tumor samples with a clonal HBV integration (n=5) or without (n=20). Two gene sets are shown to be downregulated in the first group: one gene set containing 190 genes defining inflammatory response (left) and one gene set containing 193 genes regulated by NF- κ B in response to TNF (right).



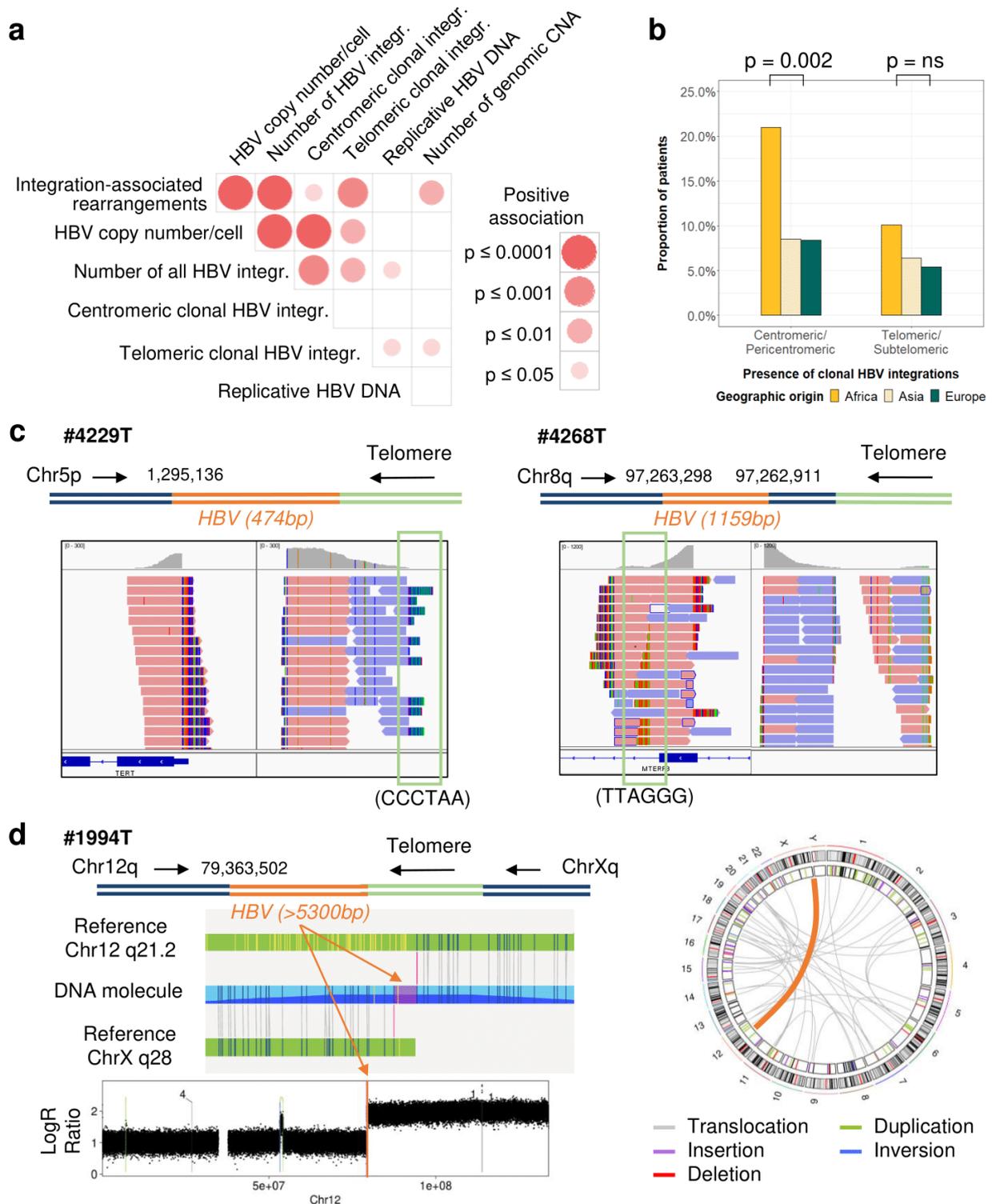
Supplementary Figure 4 | Comparison of non-tumor and tumor samples. (a) HBV copy number per cell (left) and number of HBV integration breakpoints (right) of paired tumor and non-tumor tissues of 170 HBV-positive patients from the Capture series (paired Wilcoxon signed-rank test). (B) Correlation of the HBV copy number per cell (up) and of the number of HBV integration breakpoints (down) assessed by viral capture between tumors and adjacent non-tumor tissues of 170 HBV-positive patients from the Capture series (Pearson correlation).



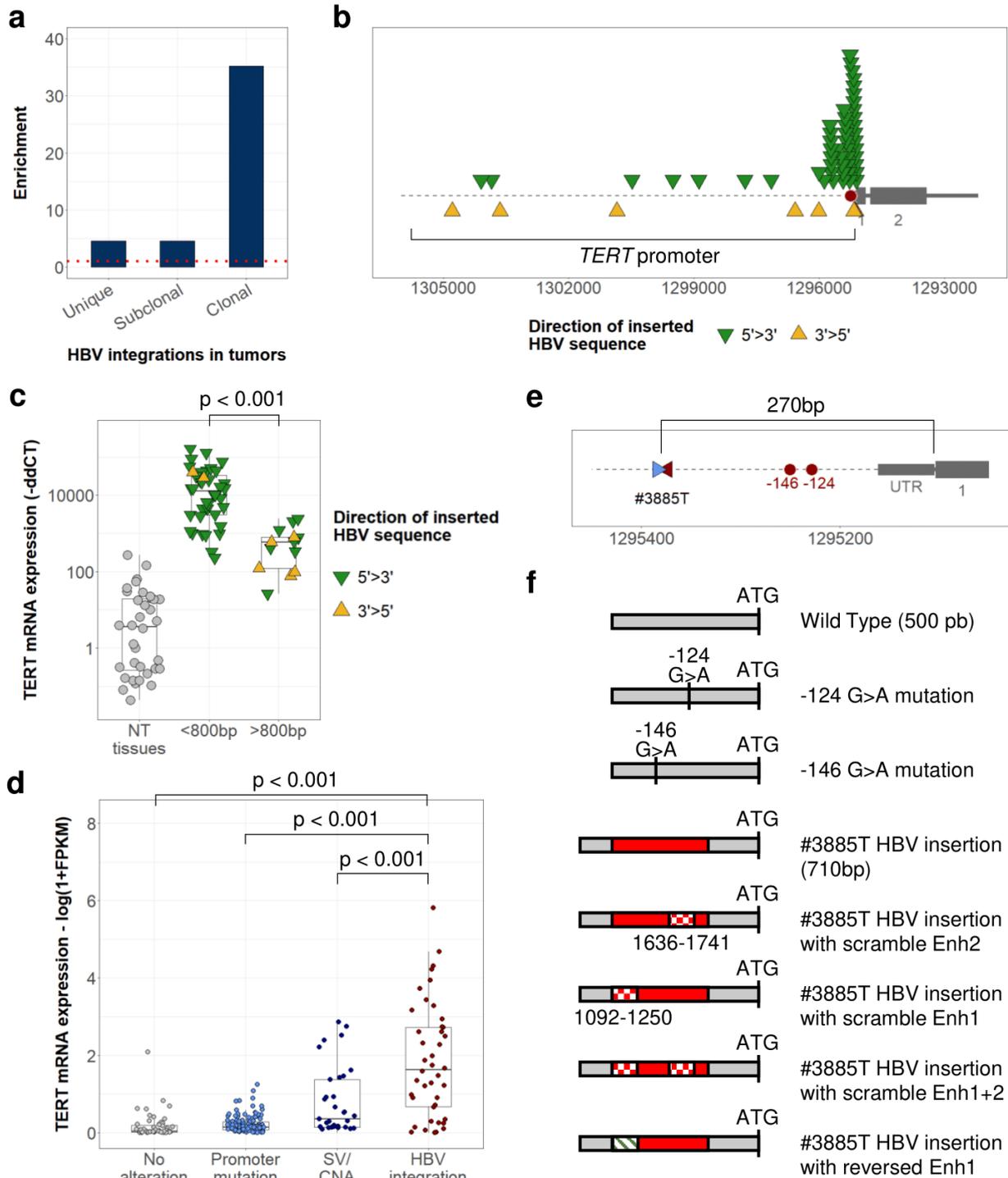
Supplementary Figure 5 | Differences in HBV integrated sequences between non-tumor tissues and tumors. (a) Mean HBV coverage from viral capture of non-tumor samples (n=156) and tumors (n=156). For each sample, the coverage at each position of the HBV genome was normalized on the mean HBV coverage. The ratio of coverage between tumors and non-tumor samples is represented on top of the figure. (b) Distribution of breakpoints in non-tumor and tumor tissues according to the structure of integration (see **Materials and Methods** for definitions) (Chi-square test). (c) Mean HBV coverage from RNA-Seq of non-tumor samples (n=22) and tumors (n=73) according to the presence of replicative HBV DNA. (d) Number of samples positive for HBV mRNA according to the localization of the HBV probes. RNA are annotated as complete if a sample is positive for all probes until the polyA region on the HBV genome. (e) HBV mRNA expression (-ddCT) in the S region of HBV genome in tumors (Wilcoxon signed-rank test). (f) Proportion of non-tumor and tumor samples with HBV variants at different positions of the HBV genome. Variant Allele Frequency was determined in paired HCC and adjacent non-tumor tissue for 57 to 98 patients according to the position coverage on the HBV genome. *VAF*, Variant Allele Frequency; *BCP*, Basal Core Promoter; *PC*, PreCore; *RT*, Reverse-Transcriptase.



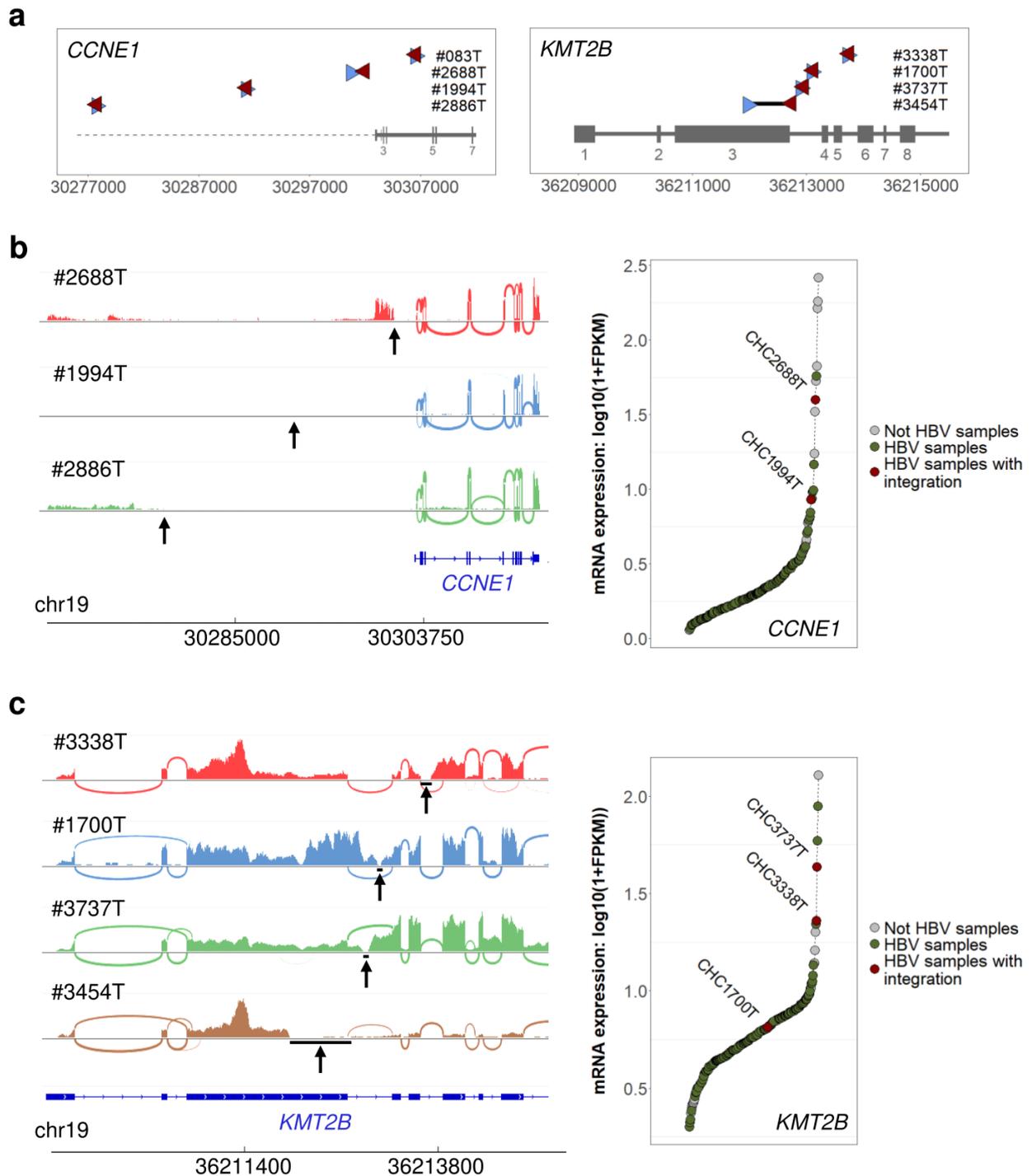
Supplementary Figure 6 | Complex rearrangements involving integrations reconstructed with long-read and Bionano sequencing. (a) Reconstruction with long-read sequencing of two clonal integration events in tumor #1151T located in chr8q and inducing copy number alterations. Profiles of copy number (total in red, minor allele in blue) and LogR ratio are represented. Event 1 is a translocation-like event between chr8q and chr17p. Event 2 is a duplication-inverted-like event in the centromere of chr8. (b) Reconstruction with Bionano whole-genome sequencing of a complex rearrangement in tumor #1597T. The rearrangement is composed of two translocations between chr5p and chr1q, and two HBV clonal integrations inducing an amplification of *TERT*.



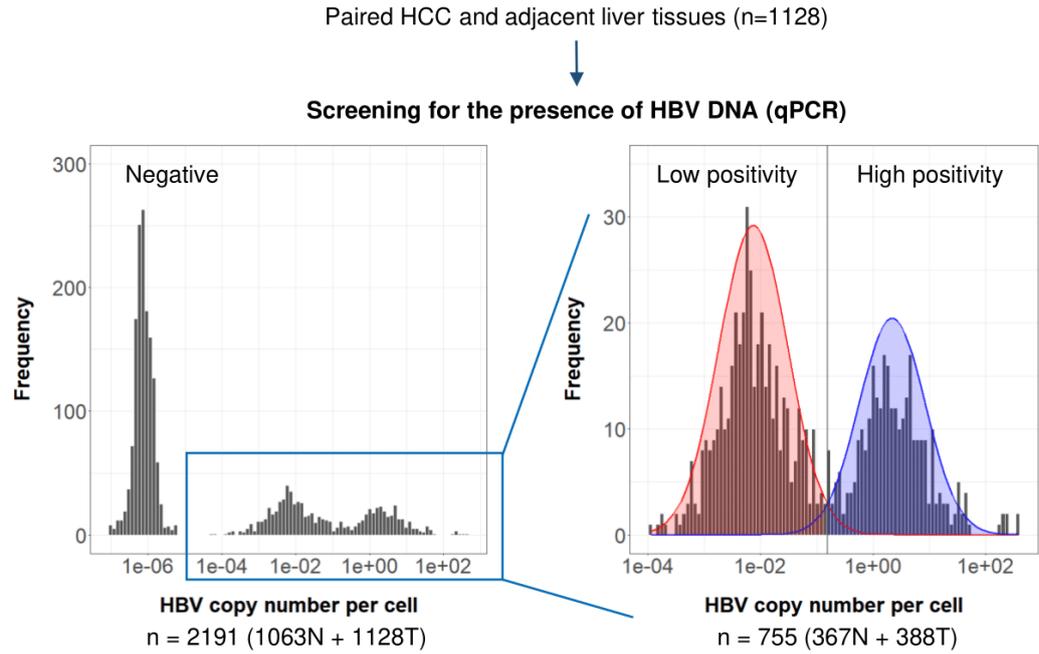
Supplementary Figure 7 | HBV insertions in centromeric/telomeric regions. (a) Correlation plots between the presence of integration-associated rearrangements, viral features and the number of genomic CNA. All associations are positive and color intensities represent different levels of statistical significance. Statistical analysis was performed using the chi-square test, Wilcoxon signed-rank test or Pearson correlation with respect to the type of variable. P values were adjusted for multiple testing using the Benjamini-Hochberg method (false discovery rate). (b) Proportion of HCC according to the patients' geographic origin harboring HBV integrations in centromeric or telomeric regions (Fisher test). (c) Reconstruction of HBV clonal integration events involving telomeric regions in tumors #4229T and #4268T. Views from Integrative Genomics Viewer are shown to visualize telomeric sequences (green square). (d) Reconstruction with Bionano whole-genome sequencing of a translocation in tumor #1994T between chr12q and the telomeric region of chrXq. The logR Ratio in chr12q is shown to see the CNA at HBV integration breakpoint. CNA, Copy Number Alteration; ns, not significant.



Supplementary Figure 8 | HBV insertional mutagenesis: *TERT* activation. (a) Enrichment of HBV integration breakpoints in tumors (n=2154) around 72 HCC-associated genes (+/- 25kb). The HCC-associated genes were selected as previously described⁵⁹. (b) Localization of HBV clonal integration breakpoints in the human genome in the promoter of the *TERT* gene. Red dots indicate the positions of classic mutations of the promoter (-124 and -146 from ATG) (c) mRNA expression (-ddCT from RT-qPCR) of *TERT* in HCC harboring an HBV integration in the *TERT* promoter (n=57) and in the adjacent non-tumor liver samples (n=36) (Wilcoxon signed-rank test). (d) mRNA expression ($\log_{10}(1+FPKM)$) from RNA-Seq data of *TERT* in HCC harboring an different alterations of the *TERT* promoter in the NGS Series (n=265) (Wilcoxon signed-rank test). (e) Localization of HBV clonal integration breakpoints in the promoter of the *TERT* gene for tumor #3885T. Red dots indicate the positions of mutations of the promoter (-124 and -146 from ATG). (f) Representation of plasmids constructs of *TERT* promoter (wild type, harboring -124 or -146 promoter mutations, harboring the HBV insertion identified in tumor #3885T wild type or containing scrambled/reversed Enhancer sequences). *SV*, Structural Variant; *CNA*, Copy Number Alteration; *Enh*, Enhancer.

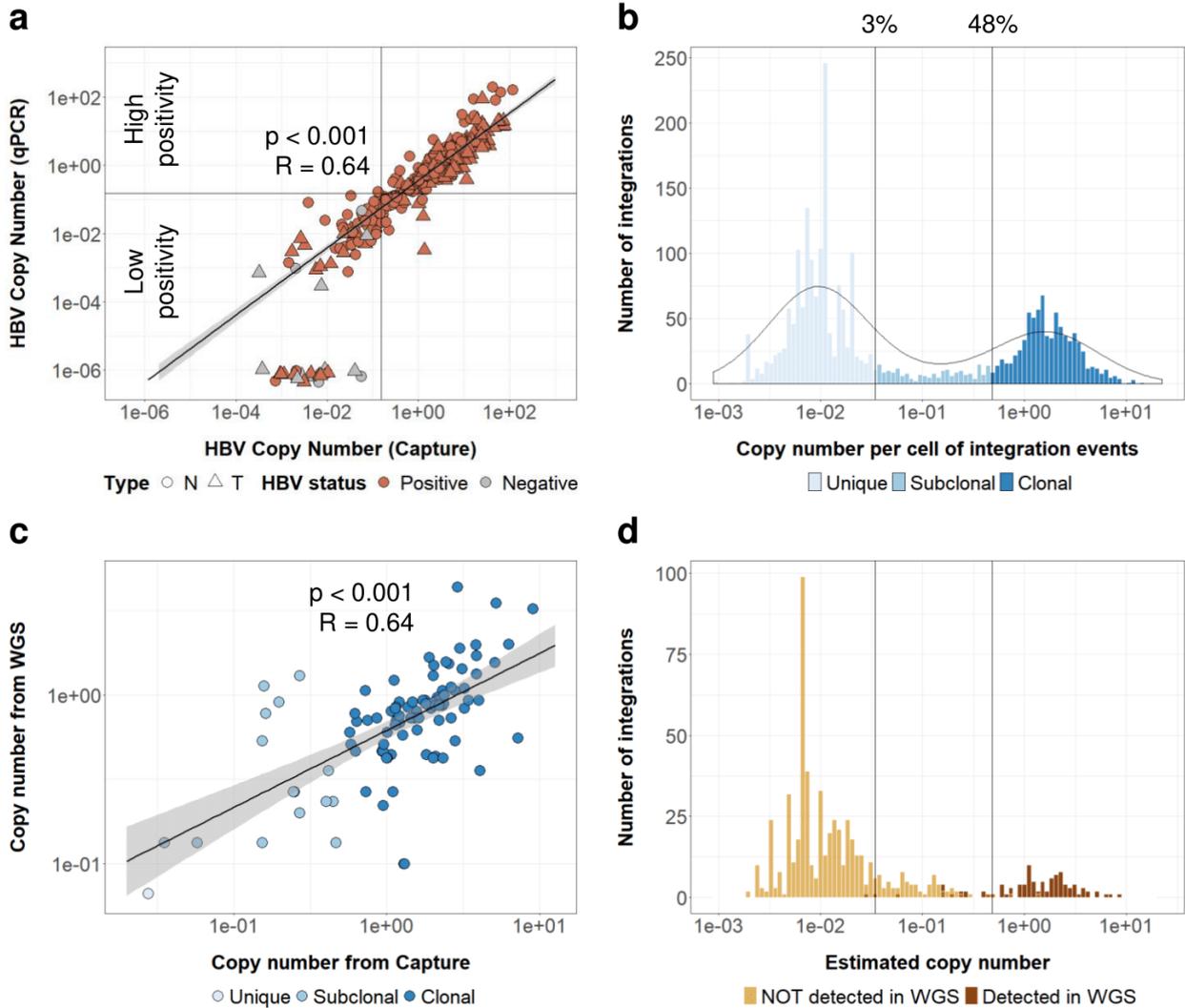


Supplementary Figure 9 | HBV insertional mutagenesis: *CCNE1* and *KMT2B*. (a) Localization of HBV clonal integration breakpoints in the human genome at the loci of *CCNE1* gene (left) and *KMT2B* gene (right). HBV integration in tumor #1994T have been previously described²¹. (b) Consequences of HBV integrations on *CCNE1* mRNA in 3 tumors. IGV-Sashimi plots show RNA-seq alignments (left) and sorted mRNA expression ($\log_{10}(1+FPKM)$) from RNA-Seq data of the NGS series ($n=265$) is represented (right). (c) Consequences of HBV integrations on *KMT2B* mRNA in 4 tumors. IGV-Sashimi plots show RNA-seq alignments (left) and sorted mRNA expression ($\log_{10}(1+FPKM)$) from RNA-Seq data of the NGS series ($n=265$) is represented (right). For IGV-Sashimi plots, alignments in exons are represented as read density, and alignments to splice junctions are shown as an arc connecting a pair of exons, where arc width is proportional to the number of reads aligning to the junction. The canonical transcripts for the genes are shown below. Black arrows indicate the position of HBV integration breakpoints. *FPKM*, fragments per kilobase of exons per million reads.



	TOTAL	High positive	Low positive (≥ 3 probes)	Negative or low positive (≤ 2 probes)			
AgHBs +	n = 1128	n = 187 129/138 (93%)	n = 163 11/90 (12%)	n=778 4/221 (2%)			
Mutation series (Targeted sequencing: WES/Miseq/Sanger)	780	181 124/133 (93%)	120 9/72 (13%)	479 2/169 (1%)			
Capture series (Viral Capture)	190	169 116/124 (94%)	9 3/7 (43%)	12 2/7 (29%)			
NGS series (WGS/WES + RNASeq)	265	117 80/86 (93%)	3 3/3 (100%)	5 3/5 (60%)	38 3/27 (11%)	7 1/4 (25%)	95 0/60 (0%)

Supplementary Figure 10 | Flow-chart of the study. Frozen paired HCC and adjacent liver tissues from 1128 patients were screened to detect the presence of HBV DNA by quantitative PCR with Taqman probes targeting 5 regions of the viral genome. Patients were classified into 3 groups based on the highest value between the two paired samples (tumor and adjacent liver): high positive (for at least 1 probe), low positive (for at least 3 probes), negative or low positive (0, 1 or 2 probes with low positivity). The positivity for AgHBs in patients' serum is indicated in red when available (See patient's characteristics in Supplementary Table 3). *HCC*, hepatocellular carcinoma; *N*, non-tumor; *T*, tumor; *NGS*, next-generation sequencing; *WGS*, whole-genome sequencing; *WES*, whole-exome sequencing; *RNASeq*; RNA sequencing.



Supplementary Figure 11 | Definition of clonality from viral capture. (a) Correlation between the values of HBV copy number per cell obtained from qPCR and viral capture in 419 samples from the Capture series of 190 patients (non-tumor liver, n=185, and tumor samples, n=234). The HBV status is annotated based on clinical data. The two techniques show a good correlation (Pearson correlation). (b) Definition of clonality with all 2505 integration events detected in tumors based on the k-means method. (c) Correlation between the values of copy number per cell for integration events obtained from WGS and from viral capture in 42 samples (non-tumor liver, n=20, and tumor samples, n=22). The two techniques show a good correlation (Pearson correlation). (d) Distribution of integration events detected in viral capture in 42 samples sequenced in WGS, based on their copy number per cell.

LIST OF SUPPLEMENTARY TABLES

Supplementary Table 1 – Characterization of 9183 integration events identified by viral capture

Supplementary Table 2 - Survival analysis in patients treated with R0 curative resection

Supplementary Table 3 – Clinical description of the series

Supplementary Table 4 – List of probe sets and primers

Supplementary Table 5 – List of HBV genotypes and human regions targeted in viral Capture

		HCC Series n=1128	Mutation Series n=780	Capture Series n=190	NGS Series n=265
Gender	M F	929/1123 (83%) 194/1123 (17%)	648/777 (83%) 129/777 (17%)	162/190 (85%) 28/190 (15%)	215/265 (81%) 50/265 (19%)
Age		64 (18-90)	64 (18-90)	56 (18-85)	62 (18-90)
Largest nodule diameter	>50	472/1069 (44%)	328/740 (44%)	75/165 (45%)	139/247 (56%)
Number nodules	≥2	281/1036 (27%)	209/730 (29%)	46/162 (28%)	55/240 (23%)
Preoperative serum AFP	>20ng/ul	335/814 (41%)	252/566 (45%)	72/117 (62%)	104/200 (52%)
Geographic Origin	Europe Africa Asia	890/1023 (87%) 90/1023 (9%) 43/1023 (4%)	570/697 (82%) 86/697 (12%) 41/697 (6%)	90/185(49%) 61/185 (33%) 34/185 (18%)	167/263 (63%) 68/263 (26%) 28/263 (11%)
Risk factors	Alcohol Hepatitis B Hepatitis C Metabolic syndrome Without or other etiology	451/1128 (40%) 226/1128 (20%) 296/1128 (26%) 256/1128 (23%) 181/1128 (16%)	306/780 (39%) 212/780 (27%) 201/780 (26%) 147/780 (19%) 107/780 (14%)	27/190 (14%) 177/190 (93%) 13/190 (7%) 14/190 (7%) 5/190 (3%)	74/265 (28%) 130/265 (49%) 35/265 (13%) 40/265 (15%) 39/265 (15%)
Metavir score	F0-F1 F2-F3 F4	298/1089 (27%) 272/1089 (25%) 519/1089 (48%)	212/759 (28%) 169/759 (22%) 378/759 (50%)	35/185 (19%) 51/185 (28%) 99/185 (54%)	97/259 (37%) 60/259 (23%) 102/259 (39%)
Tumor differentiation	Edmonson I-II Edmonson III-IV	449/1054 (43%) 605/1054 (57%)	308/741 (42%) 433/741 (58%)	54/177 (31%) 123/177 (69%)	80/255 (31%) 175/255 (69%)

Among HBV+ patients:

AgHBs	Positive Negative	143/159 (90%) 16/159 (10%)	136/151 (90%) 15/151 (10%)	123/133 (92%) 10/133 (8%)	91/97 (94%) 6/97 (6%)
AcHBc	Positive Negative	109/121 (90%) 12/121 (10%)	105/114 (92%) 9/114 (8%)	90/99 (91%) 9/99 (9%)	66/73 (90%) 7/73 (10%)
AgHBe	Positive Negative	8/77 (10%) 69/77 (90%)	8/75 (11%) 67/75 (89%)	8/69 (12%) 61/69 (88%)	7/56 (13%) 49/56 (88%)
Treatment	Yes No	33/71 (46%) 38/71 (54%)	33/69 (48%) 36/69 (52%)	29/62 (47%) 33/62 (53%)	25/50 (50%) 25/50 (50%)
HBV natural history	HBeAg+ CI HBeAg+ CH HBeAg- CI HBeAg- CH HBsAg-	2/88 (2%) 6/88 (7%) 10/88 (11%) 54/88 (61%) 16/88 (18%)	2/85 (2%) 6/85 (7%) 10/85 (12%) 52/85 (61%) 15/85 (18%)	2/73 (3%) 6/73 (8%) 8/73 (11%) 48/73 (66%) 9/73 (12%)	2/58 (3%) 5/58 (9%) 9/58 (16%) 36/58 (62%) 6/58 (10%)

Supplementary Table 3 – Clinical description of the series

REFERENCES

1. World Health Organization. *Global hepatitis report, 2017*. (2017).
2. Thomas, D. L. Global Elimination of Chronic Hepatitis. *N. Engl. J. Med.* **380**, 2041–2050 (2019).
3. World Health Organization. *Global health sector strategy on viral hepatitis 2016-2021*. (2016).
4. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **68**, 394–424 (2018).
5. Kew, M. Hepatocellular carcinoma: epidemiology and risk factors. *J. Hepatocell. Carcinoma* **115** (2014) doi:10.2147/JHC.S44381.
6. Papatheodoridis, G. V., Chan, H. L.-Y., Hansen, B. E., Janssen, H. L. A. & Lampertico, P. Risk of hepatocellular carcinoma in chronic hepatitis B: Assessment and modification with current antiviral therapy. *J. Hepatol.* **62**, 956–967 (2015).
7. Wu, C.-Y. *et al.* Association of Nucleos(t)ide Analogue Therapy With Reduced Risk of Hepatocellular Carcinoma in Patients With Chronic Hepatitis B—A Nationwide Cohort Study. *Gastroenterology* **147**, 143-151.e5 (2014).
8. Wong, G. L.-H. *et al.* Entecavir treatment reduces hepatic events and deaths in chronic hepatitis B patients With liver cirrhosis: Hepatology. *Hepatology* **58**, 1537–1547 (2013).
9. Hosaka, T. *et al.* Long-term entecavir treatment reduces hepatocellular carcinoma incidence in patients with hepatitis B virus infection. *Hepatology* **58**, 98–107 (2013).
10. Yang, J. D. *et al.* A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 589–604 (2019).
11. Valaydon, Z. S. & Locarnini, S. A. The virological aspects of hepatitis B. *Best Pract. Res. Clin. Gastroenterol.* **31**, 257–264 (2017).
12. Lamontagne, R. J., Bagga, S. & Bouchard, M. J. Hepatitis B virus molecular biology and pathogenesis. *Hepatoma Res.* **2**, 163 (2016).
13. Levrero, M. & Zucman-Rossi, J. Mechanisms of HBV-induced hepatocellular carcinoma. *J. Hepatol.* **64**, S84–S101 (2016).
14. Tu, T., Budzinska, M. A., Vondran, F. W. R., Shackel, N. A. & Urban, S. Hepatitis B Virus DNA Integration Occurs Early in the Viral Life Cycle in an *In Vitro* Infection Model via Sodium Taurocholate Cotransporting Polypeptide-Dependent Uptake of Enveloped Virus Particles. *J. Virol.* **92**, e02007-17 (2018).
15. Sung, W.-K. *et al.* Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat. Genet.* **44**, 765–769 (2012).
16. Zhao, L.-H. *et al.* Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nat. Commun.* **7**, 12992 (2016).
17. Li, C. *et al.* Androgen Receptor Enhances Hepatic Telomerase Reverse Transcriptase Gene Transcription After Hepatitis B Virus Integration or Point Mutation in Promoter Region. *Hepatology* **69**, 498–512 (2019).
18. Ally, A. *et al.* Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* **169**, 1327-1341.e23 (2017).
19. Fujimoto, A. *et al.* Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* **44**, 760–764 (2012).
20. Zapatka, M. *et al.* The landscape of viral associations in human cancers. *Nat. Genet.* **52**, 320–330 (2020).

21. Bayard, Q. *et al.* Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. *Nat. Commun.* **9**, 5235 (2018).
22. Dong, H. *et al.* Identification of HBV-MLL4 Integration and Its Molecular Basis in Chinese Hepatocellular Carcinoma. *PLOS ONE* **10**, e0123175 (2015).
23. Furuta, M. *et al.* Characterization of HBV integration patterns and timing in liver cancer and HBV-infected livers. *Oncotarget* **9**, (2018).
24. Tong, S. & Revill, P. Overview of hepatitis B viral replication and genetic variability. *J. Hepatol.* **64**, S4–S16 (2016).
25. Tu, T., Budzinska, M., Shackel, N. & Urban, S. HBV DNA Integration: Molecular Mechanisms and Clinical Implications. *Viruses* **9**, 75 (2017).
26. Nault, J. *et al.* Clinical Impact of Genomic Diversity From Early to Advanced Hepatocellular Carcinoma. *Hepatology* **71**, 164–182 (2020).
27. Yang, L. *et al.* Molecular Characterization of HBV DNA Integration in Patients with Hepatitis and Hepatocellular Carcinoma. *J. Cancer* **9**, 3225–3235 (2018).
28. Li, X. *et al.* The function of targeted host genes determines the oncogenicity of HBV integration in hepatocellular carcinoma. *J. Hepatol.* **60**, 975–984 (2014).
29. Yoo, S. *et al.* A pilot systematic genomic comparison of recurrence risks of hepatitis B virus-associated hepatocellular carcinoma with low- and high-degree liver fibrosis. *BMC Med.* **15**, 214 (2017).
30. Yan, H. *et al.* Characterization of the genotype and integration patterns of hepatitis B virus in early- and late-onset hepatocellular carcinoma: HEPATOLOGY, Vol. XX, No. X, 2015. *Hepatology* **61**, 1821–1831 (2015).
31. Yang, W. & Summers, J. Integration of Hepadnavirus DNA in Infected Liver: Evidence for a Linear Precursor. *J. Virol.* **73**, 9710–9717 (1999).
32. Summers, J. *et al.* Hepatocyte turnover during resolution of a transient hepadnaviral infection. *Proc. Natl. Acad. Sci.* **100**, 11652–11659 (2003).
33. Budzinska, M. A., Shackel, N. A., Urban, S. & Tu, T. Sequence analysis of integrated hepatitis B virus DNA during HBeAg-seroconversion. *Emerg. Microbes Infect.* **7**, 1–12 (2018).
34. Mason, W. S. *et al.* HBV DNA Integration and Clonal Hepatocyte Expansion in Chronic Hepatitis B Patients Considered Immune Tolerant. *Gastroenterology* **151**, 986-998.e4 (2016).
35. Tu, T., Budzinska, M. A., Shackel, N. A. & Jilbert, A. R. Conceptual models for the initiation of hepatitis B virus-associated hepatocellular carcinoma. *Liver Int.* **35**, 1786–1800 (2015).
36. Totoki, Y. *et al.* Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat. Genet.* **46**, 1267–1273 (2014).
37. Bill, C. A. & Summers, J. Genomic DNA double-strand breaks are targets for hepadnaviral DNA integration. *Proc. Natl. Acad. Sci.* **101**, 11135–11140 (2004).
38. Sze, K. M. *et al.* HBV- *TERT* Promoter Integration Harnesses Host ELF4 Resulting in *TERT* Gene Transcription in Hepatocellular Carcinoma. *Hepatology* hep.31231 (2020) doi:10.1002/hep.31231.
39. Amaddeo, G. *et al.* Integration of tumour and viral genomic characterisations in HBV-related hepatocellular carcinomas. *Gut* **64**, 820–829 (2015).
40. Mentha, N., Clément, S., Negro, F. & Alfaïate, D. A review on hepatitis D: From virology to new therapies. *J. Adv. Res.* **17**, 3–15 (2019).
41. Benaglia, T., Chauveau, D., Hunter, D. R. & Young, D. **mixtools**: An R Package for Analyzing Finite Mixture Models. *J. Stat. Softw.* **32**, (2009).
42. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

43. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
44. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
45. Letouzé, E. *et al.* Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat. Commun.* **8**, 1315 (2017).
46. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
47. Debacher, K. & Kooy, R. F. Fragile sites and human disease. *Hum. Mol. Genet.* **16**, R150–R158 (2007).
48. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
49. Schulze, K. *et al.* Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat. Genet.* **47**, 505–511 (2015).
50. Nault, J. C. *et al.* High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. *Nat. Commun.* **4**, 2218 (2013).
51. La Bella, T. *et al.* Adeno-associated virus in the liver: natural history and consequences in tumour development. *Gut* **69**, 737–747 (2020).
52. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
53. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545–15550 (2005).
54. Hirsch, T. Z. *et al.* BAP1 mutations define a homogeneous subgroup of hepatocellular carcinoma with fibrolamellar-like features and activated PKA. *J. Hepatol.* S0168827819307184 (2019) doi:10.1016/j.jhep.2019.12.006.
55. Boyault, S. *et al.* Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* **45**, 42–52 (2007).
56. Lazarevic, I. Clinical implications of hepatitis B virus mutations: Recent advances. *World J. Gastroenterol.* **20**, 7653 (2014).
57. Nault, J.-C. *et al.* Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat. Genet.* **47**, 1187–1193 (2015).
58. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

II. Identification d'un sous-groupe spécifique de CHC agressifs présentant des altérations des gènes *CCNA2* et *CCNE1*

L'un des premiers événements oncogéniques d'intégration du Virus de l'Hépatite B à avoir été décrit dans un CHC était une intégration d'une séquence virale dans un intron du gène *CCNA2* (Wang et al., 1990). Cette étude de 1990 avait alors observé une augmentation de la quantité d'ARNs messagers dans la tumeur étudiée par rapport au tissu non-tumoral adjacent. Plusieurs études de CHC liés au VHB utilisant le séquençage haut-débit pour détecter les intégrations virales ont permis plus récemment de décrire le gène *CCNA2* comme un site d'insertion récurrent dans les tumeurs (Li et al., 2019; Sze et al., 2020; Zhao et al., 2016). Ces mêmes travaux ont également identifié de fréquentes intégrations du VHB dans le promoteur de *CCNE1* induisant une surexpression du gène dans des CHC développés principalement sur des foies non-cirrhotiques. Ces deux gènes *CCNA2* et *CCNE1* codant pour des protéines qui jouent un rôle majeur dans la régulation du cycle cellulaire, l'ensemble des études publiées ont avancé l'hypothèse que l'altération d'un de ces gènes serait à l'origine ou favoriserait le développement tumoral en absence de cirrhose. Cependant, le mécanisme moléculaire reliant les intégrations du VHB à la carcinogénèse n'avait pas été précisément exploré.

Par ailleurs, le laboratoire a décrit en 2015 des insertions du virus Adéno-Associé 2 (AAV2) dans des CHC et plus de la moitié des événements observés étaient situées à proximité d'un de ces deux gènes (7 tumeurs sur 11) (Nault et al., 2015) dans des CHC développés sur foie sain et induisaient leur surexpression. Ainsi, à partir du séquençage des exomes ou génomes d'une série de 160 CHC du laboratoire et de deux séries de données publiques (Ally et al., 2017a; Fujimoto et al., 2016), nous avons criblé spécifiquement ces deux gènes afin d'en caractériser les différentes altérations et de mieux comprendre leur implication dans l'hépatocarcinogénèse.

Nous avons alors découvert dans plusieurs CHC la présence de réarrangements structuraux situés aux mêmes loci que les intégrations virales du VHB ou d'AAV2 : autour de l'intron 2 du gène *CCNA2* ou dans le promoteur du gène *CCNE1*. Ces différents types d'événements induisent tous une surexpression de l'ARN messager du gène altéré. D'un côté, les altérations de *CCNA2* affectent directement la structure

du gène et entraînent la synthèse d'une protéine cycline-A2 tronquée ne possédant pas le domaine de régulation de la protéine normale. D'un autre côté, les altérations de *CCNE1* modifient la région régulatrice du gène et ont comme conséquence de le placer sous le contrôle d'un « enhancer » fort pouvant être viral dans le cas d'une intégration ou humain dans le cas d'un réarrangement structural.

L'analyse des caractéristiques moléculaires et cliniques a mis en évidence que les CHC ayant des altérations des gènes *CCNA2* ou *CCNE1* forme un sous-groupe homogène de tumeurs nommé CCN-CHC. Ces tumeurs sont des CHC de grande taille et de mauvais pronostic qui se développent majoritairement en absence de cirrhose. Par ailleurs, bien qu'elles possèdent moins de mutations que l'ensemble des autres CHC, leur nombre de variants structuraux est beaucoup plus élevé. La classification de ces variants selon le type, la taille et la proximité des réarrangements a alors révélé que les CCN-CHC présentent une signature très spécifique (RS1) caractérisée par des duplications en tandem et des translocations inter-chromosomiques. Ces réarrangements sont plus fréquemment situés dans des régions décrites dans le foie normal comme étant répliquées précocement, contenant des gènes fortement exprimés et dans un état chromatinien actif. Enfin, le seul locus du génome qui contient des réarrangements récurrents et dont l'altération semble donc sélectionnée dans les CCN-CHC est le promoteur de *TERT*, ce qui induit la surexpression du gène.

Ce projet a donc permis de mettre en évidence l'existence d'un sous-groupe agressif de CHC développés dans des contextes non-cirrhotiques. Dans ces tumeurs, l'activation des gènes *CCNA2* et *CCNE1* par des intégrations virales ou des réarrangements structuraux entraîne une entrée prématurée des cellules en phase S. Cette dérégulation du cycle cellulaire génère alors un stress réplicatif à l'origine de réarrangements caractéristiques qui peuvent altérer le promoteur de *TERT* et favoriser directement le développement tumoral.

Article 2: Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress.

Quentin Bayard, Léa Meunier*, **Camille Péneau***, Victor Renault, Jayendra Shinde, Jean-Charles Nault, Iadh Mami, Gabrielle Couchy, Giuliana Amaddeo, Emmanuel Tubacher, Delphine Bacq, Vincent Meyer, Tiziana La Bella, Audrey Debaillon-Vesque, Paulette Bioulac-Sage, Olivier Seror, Jean-Frédéric Blanc, Julien Calderaro, Jean-François Deleuze, Sandrine Imbeaud, Jessica Zucman-Rossi** & Eric Letouzé**.

Nature Communications, 2018 Dec 7;9(1):5235.

*Ces auteurs ont contribué de manière égale à ce travail

**Ces auteurs ont supervisé ensemble ce travail

Les données supplémentaires (tables) sont disponibles à l'adresse suivante :
<https://www.nature.com/articles/s41467-018-07552-9#additional-information>



ARTICLE

DOI: 10.1038/s41467-018-07552-9

OPEN

Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress

Quentin Bayard^{1,2,3,4}, Léa Meunier^{1,2,3,4}, Camille Peneau^{1,2,3,4}, Victor Renault⁵, Jayendra Shinde^{1,2,3,4}, Jean-Charles Nault^{1,2,3,4,6,7}, Iadh Mami^{1,2,3,4}, Gabrielle Couchy^{1,2,3,4}, Giulia Amaddeo^{8,9}, Emmanuel Tubacher⁵, Delphine Bacq¹⁰, Vincent Meyer¹⁰, Tiziana La Bella^{1,2,3,4}, Audrey Debailon-Vesque¹¹, Paulette Bioulac-Sage^{12,13}, Olivier Seror^{1,14}, Jean-Frédéric Blanc^{11,12}, Julien Calderaro^{8,15}, Jean-François Deleuze^{5,10}, Sandrine Imbeaud^{1,2,3,4}, Jessica Zucman-Rossi^{1,2,3,4,16} & Eric Letouzé^{1,2,3,4}

Cyclins A2 and E1 regulate the cell cycle by promoting S phase entry and progression. Here, we identify a hepatocellular carcinoma (HCC) subgroup exhibiting cyclin activation through various mechanisms including hepatitis B virus (HBV) and adeno-associated virus type 2 (AAV2) insertions, enhancer hijacking and recurrent *CCNA2* fusions. Cyclin A2 or E1 alterations define a homogenous entity of aggressive HCC, mostly developed in non-cirrhotic patients, characterized by a transcriptional activation of E2F and ATR pathways and a high frequency of *RB1* and *PTEN* inactivation. Cyclin-driven HCC display a unique signature of structural rearrangements with hundreds of tandem duplications and templated insertions frequently activating *TERT* promoter. These rearrangements, strongly enriched in early-replicated active chromatin regions, are consistent with a break-induced replication mechanism. Pan-cancer analysis reveals a similar signature in *BRCA1*-mutated breast and ovarian cancers. Together, this analysis reveals a new poor prognosis HCC entity and a rearrangement signature related to replication stress.

¹INSERM, UMR-1162, Génomique Fonctionnelle des Tumeurs Solides, Equipe Labellisée Ligue Contre le Cancer, Institut Universitaire d'Hématologie, Paris 75010, France. ²Université Paris Descartes, Labex Immuno-Oncology, Sorbonne Paris Cité, Faculté de Médecine, Paris 75006, France. ³Université Paris 13, Sorbonne Paris Cité, Unité de Formation et de Recherche Santé, Médecine, Biologie Humaine, Bobigny 93017, France. ⁴Université Paris Diderot, Sorbonne Paris Cité, Paris 75013, France. ⁵Laboratory for Bioinformatics, Fondation Jean Dausset - CEPH, Paris 75010, France. ⁶Liver unit, Hôpital Jean Verdier, Hôpitaux Universitaires Paris-Seine-Saint-Denis, Assistance-Publique Hôpitaux de Paris, APHP, Bondy 93140, France. ⁷Unité de Formation et de Recherche Santé Médecine et Biologie Humaine, Université Paris 13, Communauté d'Universités et Etablissements Sorbonne Paris Cité, Bobigny 93017, France. ⁸Inserm, U955, Team 18, Université Paris-Est Créteil, Faculté de Médecine, Créteil 94010, France. ⁹Assistance Publique-Hôpitaux de Paris, Service d'Hépatologie, CHU Henri Mondor, Créteil 94010, France. ¹⁰Centre National de Recherche en Génomique Humaine, CEA, Evry 91000, France. ¹¹Service Hépatogastroentérologie et Oncologie Digestive, Hôpital Haut-Lévêque, Centre Hospitalier Universitaire de Bordeaux, Bordeaux 33076, France. ¹²Université Bordeaux, Bordeaux Research in Translational Oncology, Bordeaux 33076, France. ¹³Service de Pathologie, Hôpital Pellegrin, Centre Hospitalier Universitaire de Bordeaux, Bordeaux 33000, France. ¹⁴Radiology Department, Jean Verdier Hospital, Hôpitaux Universitaires Paris-Seine-Saint-Denis, APHP, Bondy 93140, France. ¹⁵Assistance Publique-Hôpitaux de Paris, Département de Pathologie, Hôpital Henri Mondor, Créteil 94010, France. ¹⁶Assistance Publique-Hôpitaux de Paris, Hôpital Européen Georges Pompidou, 75015 Paris, France. These authors contributed equally: Léa Meunier, Camille Peneau. These authors jointly supervised this work: Jessica Zucman-Rossi, Eric Letouzé. Correspondence and requests for materials should be addressed to J.Z.-R. (email: jessica.zucman-rossi@inserm.fr) or to E.Lé. (email: eric.letouze@inserm.fr)

ARTICLE

epatocellular carcinoma (HCC) is the third leading cause of cancer death worldwide. Only 30% of cases are diagnosed at an early stage and are amenable to curative treatment by tumor resection or liver transplantation¹. The multikinase inhibitors sorafenib² and regorafenib³ are currently the only drugs approved for advanced HCC cases, but the median life expectancy of patients with HCC on sorafenib is only 1 year. All phase III clinical trials involving targeted molecular therapies have failed so far for various reasons including liver toxicity, lack of antitumoral potency, and the molecular heterogeneity of the disease⁴. Identifying homogeneous HCC subgroups sharing similar driving mechanisms and vulnerabilities is thus crucial to design successful patient-tailored clinical trials.

Most HCC develop in a cirrhotic liver, associated with various etiologies including hepatitis B virus (HBV) and hepatitis C virus (HCV) infections, alcohol abuse, metabolic disease, and exposure to carcinogenic compounds like aflatoxin B1⁵. The natural history of HCC in cirrhosis follows a well-established sequence with the successive development of dysplastic nodules that can transform into early stage and advanced HCC. *TERT* promoter mutations are the initial oncogenic events already detected in dysplastic nodules⁶ whereas alterations in other HCC drivers^{7–11} involved in cell cycle control (*TP53*, *RBI*, *CCND1*, *CDKN2A*), Wnt/ β -catenin signaling (*CTNBN1*, *AXIN1*), oxidative stress response (*NFE2L2*, *KEAP1*) epigenetic regulation (*ARID1A*, *ARID2*) and the AKT/mTOR and MAP kinase pathway (*RPS6KA3*, *TSC1*, *TSC2*, *PTEN*) only occur in progressed HCC¹².

In 20% of the cases, HCC develops in absence of cirrhosis. These patients usually maintain adequate liver functions and, being less subject to liver toxicity, may be eligible for more treatment options. The etiology of HCC in absence of cirrhosis is largely unknown, but one mechanism of transformation involves insertional mutagenesis by the HBV virus. The first oncogenic HBV insertion was identified in cyclin A2 gene (*CCNA2*)¹³. Since then, recurrent HBV insertions were mapped in several oncogenes including *CCNE1*, *KMT2B* and *TERT*^{14,15}. Recently, we identified adeno-associated virus type 2 (AAV2) insertions as a new etiology for HCC developed in absence of cirrhosis, with recurrent insertions in *CCNA2* and *CCNE1* genes¹⁶. However, the molecular consequences of viral insertions in cyclin genes and their precise role in HCC development remain poorly understood.

Here, we report the systematic screening of *CCNA2* and *CCNE1* alterations in 751 HCC. We identify new mechanisms of cyclin A2/E1 activation, and we explore the clinical and molecular characteristics of this tumor subgroup.

Results

Viral insertions and gene fusions activate cyclin A2. To identify the exhaustive landscape of *CCNA2* and *CCNE1* alterations in HCC, we analyzed 751 HCC comprising an in-house series of 160 tumors (LICA-FR) analyzed by RNA sequencing (RNAseq, $n = 160$), whole exome (WES, $n = 156$) and whole genome sequencing (WGS, $n = 45$) (Supplementary Data 1), the TCGA¹⁷ series (334 HCC with RNA-seq and WES, 48 of which also analyzed by WGS) and the ICGC-JP¹¹ series (257 HCC with WGS data, Supplementary Data 2).

We first screened the LICA-FR series of 160 tumors to characterize the exhaustive mechanisms activating *CCNA2* and *CCNE1* in HCC. We identified one HBV and 5 AAV2 insertions (four previously described in the ref.¹⁶) in *CCNA2* gene (Supplementary Data 3), all but one located within *CCNA2* intron 2 (Fig. 1a). Viral insertions were associated with *CCNA2* mRNA over-expression ($P = 8.2 \times 10^{-9}$, fold-change = 5.6, Fig. 1b), but also altered the transcript and protein structure.

AAV2 and HBV insertions induced the expression of various abnormal transcripts (Supplementary Fig. 1), predicted to generate a truncated cyclin A2 protein starting at methionine 148 or 158 with occasionally a few amino acids translated from the viral genome (Fig. 1c).

In addition we identified novel gene fusions in 4 tumors (Supplementary Data 4), all involving the C-terminal part of *CCNA2* (exons 3–8) at chromosome 4q27 downstream 3 different partner genes: *GSTCD* at 4q24, *SNX29* at 16p13.13 and *TET2* ($\times 2$) at 4q24 (Fig. 1a, d). In the *TET2-CCNA2* and *GSTCD-CCNA2* fusion transcripts, the first untranslated exons of *TET2* and *GSTCD* were linked with *CCNA2* exons 3–8. The *SNX29-CCNA2* fusion revealed an alternative transcription start site (TSS) in *SNX29* intron 14 generating a 448-nucleotide sequence spliced with *CCNA2* exon 3. In all fusions, the predicted translation initiation site of the fused RNA was located at methionine 158 in *CCNA2* exon 3, predicted to generate a truncated cyclin A2 protein of 275 amino acids (32 Kda), lacking the destruction box¹⁸ and ubiquitination targeting sequences¹⁹ but retaining the functional cyclin box, without any protein fragment from the partner genes (Fig. 1e).

Western blot analysis of 9 tumors with viral insertion or gene fusion confirmed the over-expression, as predicted, of a truncated 32 KDa protein (Fig. 1f). Thus, gene fusions and viral insertions in *CCNA2* both lead to the production of a stable protein lacking the N-terminal regulatory domains.

In the TCGA series, we identified 7 *CCNA2* fusions with 5 different partner genes (*FAM160A1*, *KIAA1109 $\times 3$, *LIPC*, *UBA6* and *TDO2*, Fig. 1a, d), all of which involved the first untranslated exon(s) of the partner gene linked with exons 3–8 of *CCNA2*. WGS revealed in another tumor a focal deletion starting in the 5' UTR region and ending in *CCNA2* intron 2 (Supplementary Fig. 2). All these events were predicted to generate the same 32 KDa truncated cyclin A2 protein lacking N-terminal regulatory domains. We also identified one tumor with HBV insertion and 3 tumors with AAV2 insertions in *CCNA2*. Finally, 6 tumors strongly overexpressed *CCNA2* (FPKM > 15), 3 of which displayed 23–48 Mb intra-chromosomal deletions linking the intergenic region downstream *CCNA2* with the highly expressed *ALB*, *AFP*, and *ADH6* genes (Supplementary Fig. 2). The ICGC-JP cohort comprised one HBV insertion in *CCNA2* intron 2 and one fusion between the first untranslated exon of *ANXA5* and exons 3–8 of *CCNA2* (Fig. 1a, d).*

In total, we identified 10 HCC with *CCNA2* activation events in the LICA-FR series (6.2%), 2 in the ICGC-JP series (0.8%) and 18 in the TCGA series (5.4%), associated with a significant increase of *CCNA2* mRNA expression, but also generating a truncated cyclin A2 protein lacking the N-terminal destruction box and the ubiquitination site.

Viral insertions and enhancer hijacking activate cyclin E1. In our series of 160 HCC, we identified 5 AAV2 insertions (three previously described in the ref.¹⁶) and one HBV insertion in the 5' region or upstream the transcription start site (TSS) of *CCNE1* (Fig. 2a, Supplementary Data 3). These viral insertions induced a massive overexpression of the full-length *CCNE1* gene (Fig. 2b), confirmed by western-blot analysis (Supplementary Fig. 3). Interestingly, one case with AAV2 insertion (FR2141T) also displayed an amplification of *CCNE1* locus including the viral sequence (Supplementary Fig. 3), suggesting a two-step selection of *CCNE1* activation in the natural history of this tumor. Four other tumors overexpressed *CCNE1* (FPKM > 6), explained by high-level amplification in one case. In the 3 remaining cases, whole genome sequencing revealed interchromosomal translocation breakpoints in the regulatory region of *CCNE1* (Fig. 2a). Tumor FR2048T

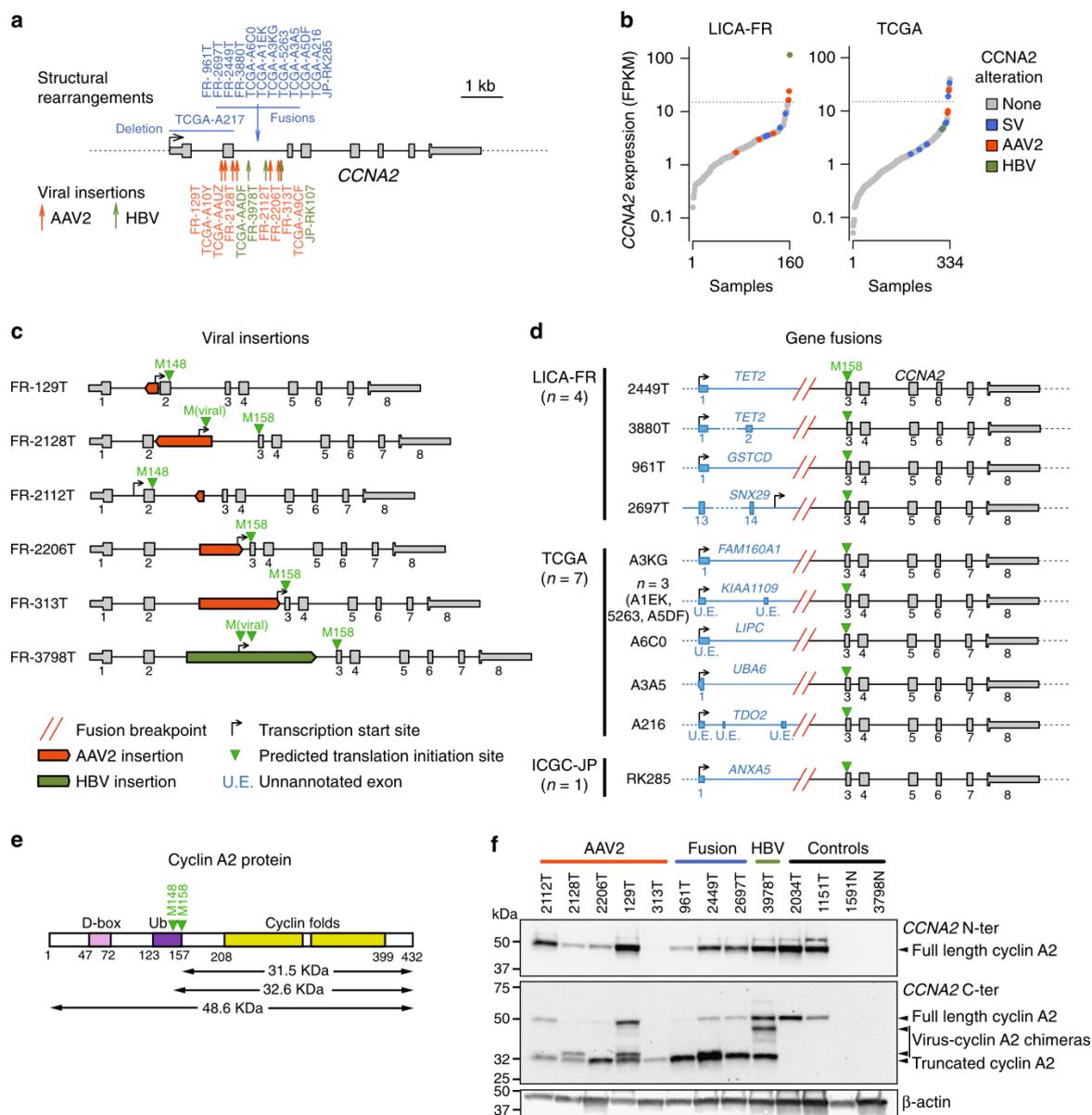


Fig. 1 Diverse mechanisms leading to *CCNA2* activation in HCC. **a** Summary of structural rearrangements (top) and viral insertions (bottom) affecting *CCNA2* gene identified in 751 HCC from the LICA-FR, TCGA and ICGC-JP cohorts. **b** Sorted *CCNA2* expression (log scale) in the LICA-FR and TCGA cohorts. Gene expression was obtained from RNA-seq data and is given in fragments per kilobase of exons per million reads (FPKM). Samples harboring structural variants (SV) or viral insertions are indicated with a color code. **c** Functional consequences of AAV2 and HBV insertions in *CCNA2*. Viral insertions identified in the LICA-FR cohort were precisely mapped using WGS or viral capture data, and RNA-seq reads were aligned on the reconstructed chimeric DNA to identify the transcription start sites and predicted translation initiation sites of abnormal transcripts. **d** *CCNA2* fusions identified in the LICA-FR, TCGA and ICGC-JP cohorts. The transcription start site of the fusion transcript is represented together with the predicted translation initiation site. Fusions with *KIAA1109*, *LIPC* and *TDO2* involve 5' exons not annotated in transcript databases but expressed in normal liver. **e** Schematic representation of cyclin A2 protein with functional domains. D-box Destruction box; Ub, Ubiquitination targeting sequences. **f** Western blot analysis of cyclin A2 using antibodies targeting the N-terminal (top) or C-terminal (middle) domains. Tumors with viral insertions or gene fusions are compared with tumors without *CCNA2* alteration and non-tumoral liver controls

displayed a translocation placing *CCNE1* downstream the first untranslated exon of the highly expressed *ERRF1* gene, leading to a highly expressed *ERRF1-CCNE1* fusion. The two other translocations lead to juxtapose *CCNE1* promoter with enhancer-rich chromatin areas located close to the highly expressed genes *RAPH1*

and *CYB5A* (Fig. 2c). Thus, both viral insertions and structural rearrangements can activate *CCNE1* expression by bringing viral or distal human enhancers in the regulatory region of the gene.

In the TCGA series, 10 tumors overexpressed *CCNE1* (Fig. 2b), including 2 cases with HBV insertion, one with HBV insertion

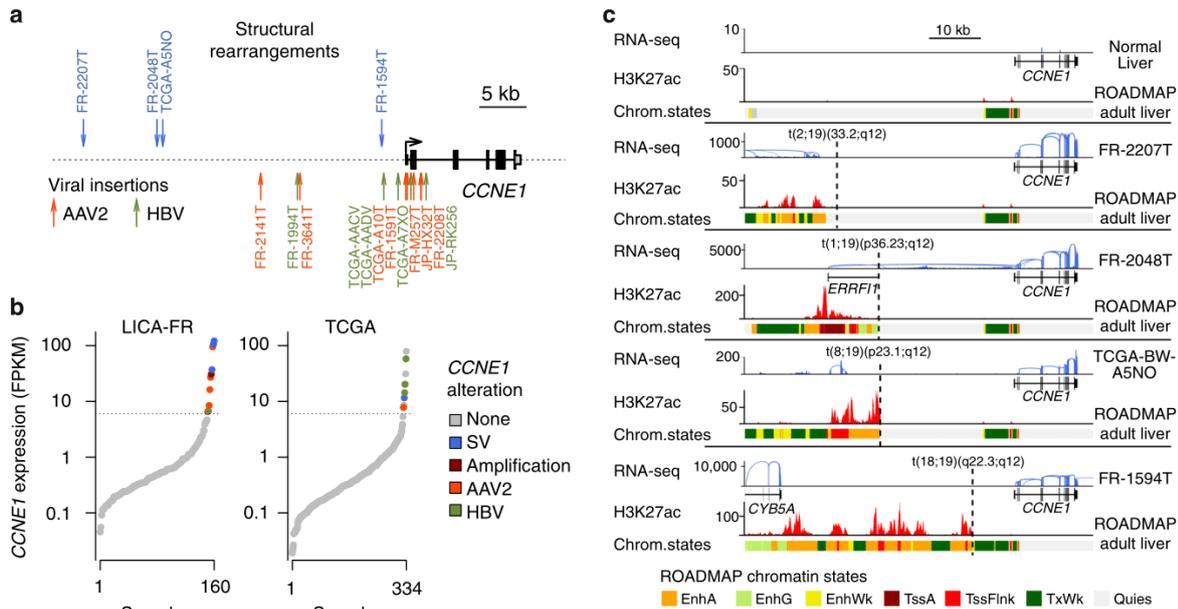


Fig. 2 Viral and non-viral mechanisms of *CCNE1* activation in HCC. **a** Summary of structural rearrangements (top) and viral insertions (bottom) affecting *CCNE1* gene identified in 751 HCC from the LICA-FR, TCGA and ICGC-JP cohorts. **b** Sorted *CCNE1* expression (log scale) in the LICA-FR and TCGA cohorts. Gene expression was obtained from RNA-seq data and is given in fragments per kilobase of exons per million reads (FPKM). Samples harboring structural variants (SV), focal amplifications and viral insertions are indicated with a color code. **c** Functional consequences of structural rearrangements affecting *CCNE1* regulatory region. RNA-seq read counts along *CCNE1* locus are represented in normal liver (top) and in 4 tumors harboring structural rearrangements upstream *CCNE1* transcription start site (TSS). H3K27Ac chromatin immunoprecipitation sequencing (ChIP-seq) signal and chromatin states in adult liver were obtained from the ROADMAP consortium and are depicted below each reconstructed DNA sequence. EnhA: active enhancer; EnhG: genic enhancer; EnhWk: weak enhancer; TssA: active TSS; TssFlnk: flanking TSS; TxWk: weak transcription; Quies: quiescent chromatin

plus high-level amplification, one with AAV2 insertion and one with a translocation between *CCNE1* regulatory region and an enhancer-rich region on chromosome 5 (Fig. 2c). In the 5 remaining cases, the mechanism leading to *CCNE1* overexpression remained unexplained in absence of WGS data. In the ICGC-JP cohort, we identified one AAV2 and one HBV insertion associated with *CCNE1* overexpression. In total, we identified 10 HCC with *CCNE1* activation events in the LICA-FR cohort (6.2%), two in the ICGC series (0.8%) and 10 in the TCGA series (3.0%).

Across the three data sets, 52/751 tumors (6.9%) displayed an activation of cyclin A2 ($n=30$) or E1 ($n=22$) due to viral insertions or structural rearrangements. These are later referred to as CCN-HCC. The proportion of CCN-HCC varied between the cohorts (12.5% in our series, 8.4% in TCGA and 1.6% in ICGC-JP) due to differences in etiological backgrounds (Supplementary Data 2). It was particularly high in our series enriched in cancers developed in a non-fibrotic liver, and low in the ICGC-Japan series dominated by HCV-related cases.

Cyclin A2 or E1 activation defines a homogeneous HCC subgroup. We next explored the molecular and clinical characteristics of CCN-HCC. Gene expression analysis of the LICA-FR and TCGA showed that CCN-HCC defined homogeneous transcriptional clusters (Fig. 3a). They were characterized by an overexpression of cell cycle genes, in particular E2F targets, and an activation of the ATR pathway in response to replication stress (Fig. 3b, Supplementary Data 5). The most significant down-regulated pathways were oxidative phosphorylation, suggesting a metabolic switch to aerobic glycolysis (Warburg effect), and *MYC* targets. We also compared the alteration frequencies of known

liver cancer driver genes¹⁰ between CCN-HCC and others. *CCNA2* and *CCNE1* activation events were remarkably exclusive from *CTNBN1* and *TERT* promoter mutations, but frequently associated with *PTEN* and *RB1* inactivation in both the LICA-FR and TCGA series (Fig. 3b, Supplementary Data 6). *RB1* inactivation may allow cells to overcome oncogene-induced senescence²⁰ in these tumors, whereas *PTEN* inactivation might favor the oncogenic metabolic switch that we observed at the transcriptional level²¹. Compared to the other tumors in the LICA-FR series, CCN-HCC were enriched in large tumors (median largest nodule diameter = 115 vs. 60 mm, $P=0.0033$), of poor prognosis (median overall survival = 21 vs. 69 months, $P=0.0072$, Fig. 3c), developed in younger patients (median age = 57 vs. 67 years old, $P=0.050$) with a non-fibrotic liver (fibrosis stage F0-F1 80 vs. 42%, $P=0.0011$). Thus, CCN-HCC define a homogenous HCC entity with characteristic clinical and molecular features.

CCN-HCC display a unique structural rearrangement signature. To identify mutational signatures associated with CCN-HCC, we analyzed the whole genome sequences of 45 of our 160 HCC (35 were previously published²², 10 new), including 13 CCN-HCC. With a median of 12,463 mutations, CCN-HCC were rather less mutated than others (median = 16,397 mutations, $P=0.065$). Mutational signatures 4, 5, and 16 (COSMIC nomenclature), ubiquitous in liver cancers²², accounted for most mutations in CCN-HCC, with a slight increase of signature 5 (53 vs. 33%, $P=0.036$) and decrease of signature 16 (23 vs. 32%, $P=0.05$) as compared with other HCC (Supplementary Fig. 4).

In contrast, CCN-HCC displayed >3 times more structural variants (median = 415 vs. 126, $P=1.1 \times 10^{-4}$). We identified 6 rearrangement signatures, termed RS1 to RS6, characterized by

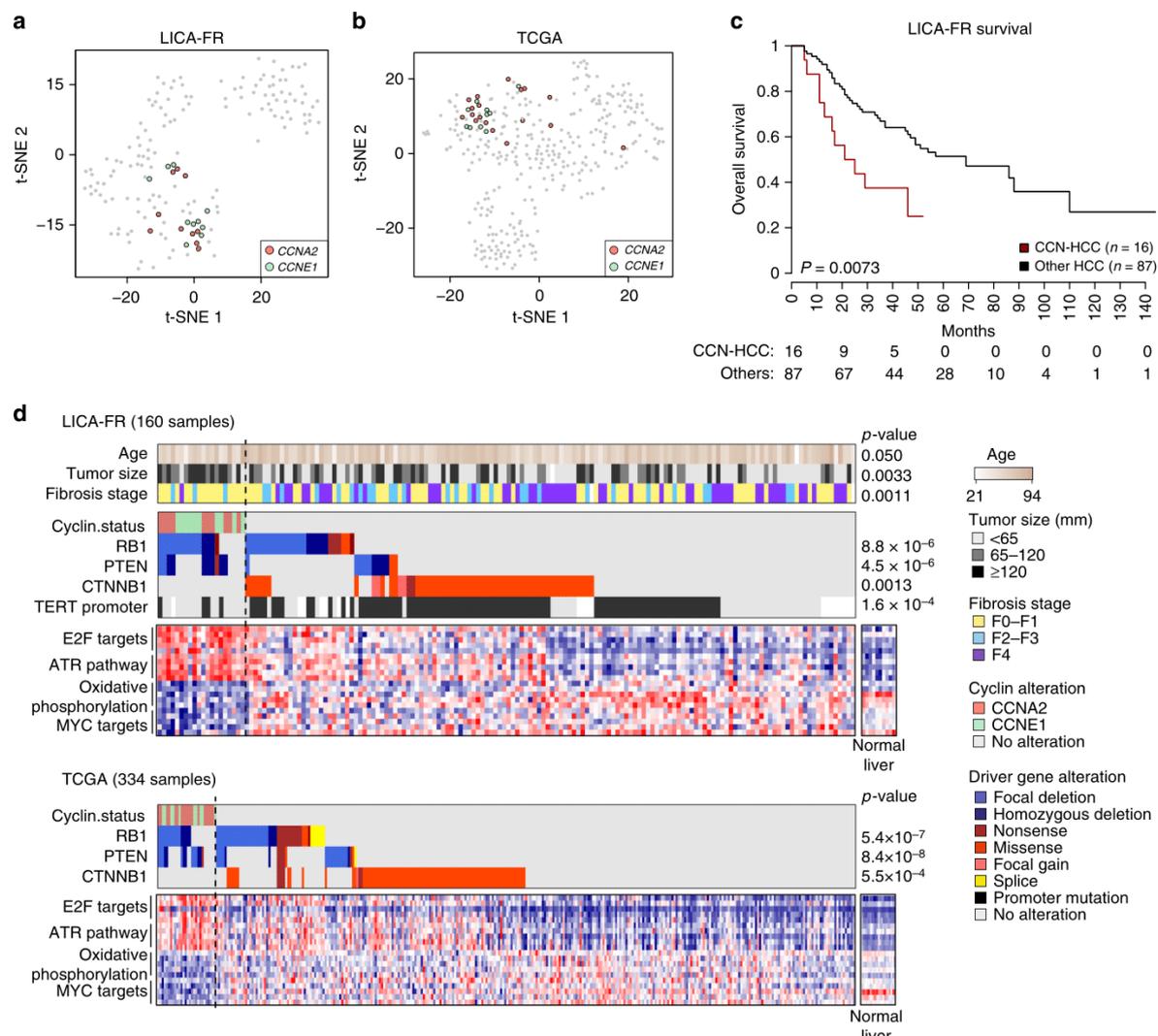


Fig. 3 Clinical and molecular features of cyclin-activated HCC. **a** t-SNE plots depicting the classification of HCC from the LICA-FR and TCGA cohorts based on their transcriptional profiles. Tumors harboring *CCNA2* or *CCNE1* activating alterations are indicated with a color code. **b** Clinical characteristics, driver genes and deregulated pathways associated with CCN-HCC in the LICA-FR (top) and TCGA (bottom) cohorts. **c** Overall survival in CCN-HCC as compared with other HCC in the LICA-FR cohort. Only HCC with curative resection (R0) were included

different combinations of rearrangement categories defined according to the type, size, and clustered nature of rearrangements (Fig. 4a). Strikingly, a high number of rearrangements attributed to signature RS1 (≥ 50 events) was specifically encountered in a cluster of 13 tumors corresponding exactly to CCN-HCC ($P = 1.4 \times 10^{-11}$, Fig. 4b). We validated this association using WGS data from the ICGC-JP series and a subset of 48 samples from the TCGA series (Fig. 4c, Supplementary Data 7). In absence of WGS data for the rest of the TCGA series, we used SNP array data to estimate the number of focal gains (<200 kb) in each tumor as a surrogate marker of the RS1 signature. With a median of 120 events, CCN-HCC displayed a significant increase of focal gains as compared with other HCC in the TCGA series (median = 6, $P < 2.2 \times 10^{-16}$, Supplementary Fig. 5). Thus, CCN-HCC have a relatively low mutation burden but a large number of structural rearrangements with a specific signature.

RS1 features suggest a replication stress-induced mechanism.

Almost all rearrangements in CCN-HCC belonged to signature RS1, characterized by a combination of small tandem duplications (<100 kb) and inter-chromosomal translocations (Fig. 4d). CCN-HCC also displayed a typical copy-number profile showing hundreds of focal gains, usually one copy above surrounding chromosome segments (Supplementary Fig. 6). Surprisingly, overlaying structural rearrangement breakpoints with copy-number profiles revealed that only 68% of these gains were due to tandem duplications, other gains being frequently surrounded by translocation or inversion breakpoints (Fig. 4e, Supplementary Fig. 6). A recurrent feature consisted of several chromosome segments, usually between 10 and 100 kb, stung together and with the same duplication level relative to their source chromosomes. Most of these events involved segments from two (Fig. 4f) or more (Supplementary Fig. 7) different chromosomes, a feature recently described as templated

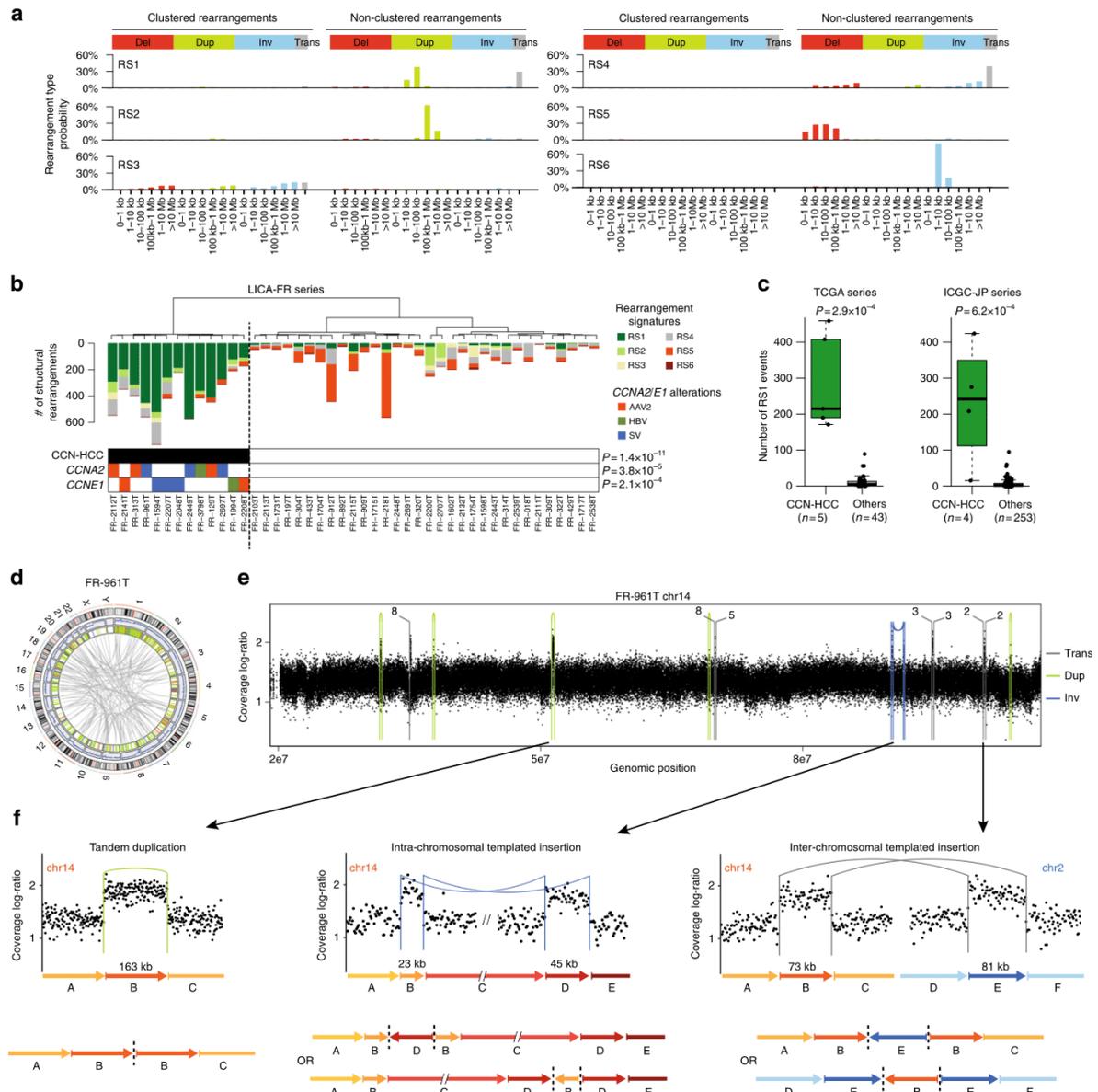


Fig. 4 Cyclin-activated HCC display a specific signature or structural rearrangements. **a** Six rearrangement signatures identified across 350 HCC genomes by non-negative matrix factorization. Structural rearrangements were classified in 38 categories considering their type (del: deletion, dup: tandem duplication, inv: inversion, trans: inter-chromosomal translocation) and size, and distinguishing clustered from non-clustered events. The probability of each rearrangement category in each signature is represented, with rearrangement types indicated above and rearrangement sizes below. **b** Unsupervised classification of 45 HCC from the LICA-FR cohort based on the contribution of rearrangement signatures in each tumor. Significant molecular alterations associated with the cluster of tumors having a high contribution of signature RS1 are represented below. *P*-values were obtained using Fisher’s exact tests. **c** Validation of the association between the RS1 signature and CCN-HCC in the TCGA and ICGC-JP series. The middle bar, median; box, interquartile range; bars extend to 1.5 times the interquartile range. **d** CIRCOS plot representing the structural rearrangement profile of a representative CCN-HCC (FR-961T, harboring a *GSTCD-CCNA2* fusion). **e** Copy-number profile showing the accumulation of focal gains along chromosome 14 in tumor FR-961T. Structural rearrangements are overlaid on the copy-number profile with a color code indicating the type of event. trans: inter-chromosomal translocation; dup: tandem duplication; inv: inversion. **f** Three types of rearrangements leading to focal chromosome gains in CCN-HCC. A representative example of each type of event is shown with a copy-number plot above and a schematic representation of the rearranged chromosome below. Structural rearrangements are represented with the same color code as in **e**. Dashed lines on schematic chromosome reconstructions represent the abnormal junctions detected in WGS data

insertion cycle²³. Inter-chromosomal templated insertions accounted for 11% of focal gains in CCN-HCC. Other events, which we call intra-chromosomal templated insertions, involved distal segments of a same chromosome and appeared as couples of inversions (Fig. 4f) or duplication and deletion (Supplementary Fig. 7), depending on the orientation of the junctions. Intra-chromosomal templated insertions accounted for 7% of focal gains in CCN-HCC. All these events are consistent with a replication-based mechanism in which a DNA polymerase at a stalled replication fork would switch template, replicate one or more other DNA regions and switch back to the original template strand behind the point of departure, generating a duplication on the host chromosome^{23–26}. Such mechanism could be particularly active in CCN-HCC due to replication stress induced by premature S phase entry.

Structural rearrangements activate *TERT* promoter in CCN-HCC. To better understand the functional consequences of the rearrangement phenotype observed in CCN-HCC, we examined the location of 8466 breakpoints attributed to the RS1 signature among the 350 liver cancer genomes from the LICA-FR, TCGA and ICGC cohorts. RS1 breakpoints were not distributed evenly along the genome but formed clusters located almost exclusively within active topologically associated domains (TADs, Fig. 5a) characterized by early replication, high gene expression and active chromatin states in normal liver (Fig. 5b). In particular, RS1 breakpoint hotspots were frequently observed at loci encoding very highly expressed liver enzymes exemplified by the albumin (*ALB*), alcohol dehydrogenase (*ADH*) and hydroxysteroid 17-Beta dehydrogenases (*HSD17B*) loci on chromosome 4 (Fig. 5a, Supplementary Fig. 8). Among the 18 chromatin states defined by the ROADMAP consortium in normal adult liver, active transcription start sites (TSS) and enhancer regions were the most strongly enriched in RS1 breakpoints (fold-change > 3), whereas quiescent and heterochromatin domains were the most depleted (Fig. 5c). TSS and enhancer regions were also enriched, to a lesser extent, in breakpoints related to signature RS2 characterized by large tandem duplications. By contrast, breakpoints related to signature RS6, dominated by inversions < 10 kb, were predominantly observed in heterochromatin and ZNF repeats.

We then used binomial regression²⁷ to model the density of rearrangement breakpoints along the genome considering an extensive set of genomic features (Supplementary Fig. 9) and to identify hotspots harboring more breakpoints than expected by chance from the background model, which may indicate positive selection in CCN-HCC. We identified a single significant locus corresponding to *TERT* promoter region ($q=0.0029$, Fig. 5d). Although *TERT* promoter mutations were rare in CCN-HCC (9 vs. 55% in others, $P=2.4 \times 10^{-5}$), *TERT* promoter rearrangements were highly enriched (82 vs. 7%, $P=1.8 \times 10^{-15}$, Fig. 5e) and involved regions of active chromatin in normal liver, in the vicinity of highly expressed liver enzymes (*ALB*, *FGG*, *SEP15*, *SLC12A7* and *BAAT*) or transcription factors (*HNF4A*, *CEBPA*, and *CEBPB*) (Supplementary Data 8, Supplementary Fig. 10). *TERT* promoter rearrangements induced an over-expression of *TERT*, stronger than promoter mutations but lower than HBV insertions (Supplementary Fig. 11). Of the 18 *TERT* promoter rearrangements identified in CCN-HCC, 16 could be associated with signature RS1 with a probability ≥ 0.5 (Fig. 5f). By contrast, most *TERT* promoter rearrangements in other HCC were related to signature RS4. Thus, structural rearrangements induced by replication stress are enriched at active chromatin regions and can promote CCN-HCC development by activating oncogenes like *TERT*.

CCN-HCC share a similar signature with *BRCA1*-altered cancers. To investigate the prevalence of the RS1 signature in other cancer types, we applied our method to 2606 tumors from the ICGC PanCancer Analysis of Whole Genomes (PCAWG) dataset^{23,28,29}. In this pan-cancer series, we identified 9 rearrangement signatures (Supplementary Fig. 12), including one signature (RS1-pancan) highly similar to the RS1 signature that we identified in liver cancers (cosine similarity = 0.91). The RS1-pancan signature was detected at low frequency in several cancer types (e.g. bladder, lung, esophageal and gastric cancers), and was highly active in breast (18% of samples with ≥ 50 RS1 events) and ovarian (33%) cancers. However, this signature was associated with *CCNA2/E1* rearrangements only in liver cancer (Fig. 6a, Supplementary Data 9). Thus, the relationship between cyclin A2/E1 activation and signature RS1 is specific to liver cancer, and the molecular cause of this signature in other cancer types remains to be elucidated. In ovarian and breast cancer, RS1 signature was not associated with *CCNE1* amplifications but with *BRCA1* inactivation (Fig. 6b, c), consistent with previous reports^{30,31}. Despite sharing a common signature of short tandem duplications and templated insertions, *CCNA2*, *CCNE1* and *BRCA1*-altered tumors displayed slightly different characteristics. First, the number of RS1 rearrangements was higher in *CCNA2*-activated HCC (median = 269) than in *CCNE1*-activated HCC (137) and *BRCA1*-altered breast (132) and ovarian (159) cancers (Fig. 6d). Second, tandem duplications were larger in *CCNE1*-activated HCC (median = 39 kb) than in *CCNA2*-activated HCC (22 kb), and smaller in *BRCA1*-altered breast (9 kb) and ovarian (10 kb) cancers (Fig. 6e). Finally, duplication and translocation breakpoints were strongly enriched in early-replicated regions in CCN-HCC as compared with other HCC, but not in *BRCA1*-altered as compared with other breast and ovarian cancers (Fig. 6f). Cyclin E1 activation was recently shown to induce replication stress by firing novel replication origins located within highly transcribed genes and prone to collapse³². *BRCA1* is implicated in the response to replication stress^{33,34} and its inactivation leads to tandem duplication formation at stalled forks by a replication restart-bypass mechanism³⁵. Cyclin A2/E1 activation in HCC and *BRCA1* inactivation in breast and ovarian cancers may thus converge towards a similar rearrangement signature, with specificities reflecting the different ways by which these genetic alterations induce replication stress or modulate response to it (Fig. 6g).

Discussion

Here, we report the characterization of a homogeneous HCC subgroup driven by the activation of *CCNA2* or *CCNE1* gene. CCN-HCC represent 7% of HCC across the 3 data sets analyzed here, but up to 14% of HCC developed in a non-fibrotic liver. These patients often have atypical clinical presentation, without any history of primary risk factors, and can be remarkably young, exemplified by tumor FR-3880T developed in a 32 year-old woman without any risk factor, due to a *TET2-CCNA2* fusion. CCN-HCC are usually large tumors of poor prognosis but share molecular characteristics, in particular high proliferation and replication stress, that could provide therapeutic opportunities³⁶. First, conventional chemotherapies mainly affect actively dividing cells by generating DNA damage or blocking DNA replication, and the tandem duplicator phenotype was identified as a marker for chemotherapeutic response in breast cancer cell lines and patient-derived xenografts³⁷. Transarterial chemoembolization (TACE) with doxorubicin, cisplatin or epirubicin, usually recommended for patients with intermediate HCC not eligible for surgery, may thus be an interesting option for CCN-HCC. Poly (ADP-ribose) polymerase (PARP) inhibitors, the first clinically

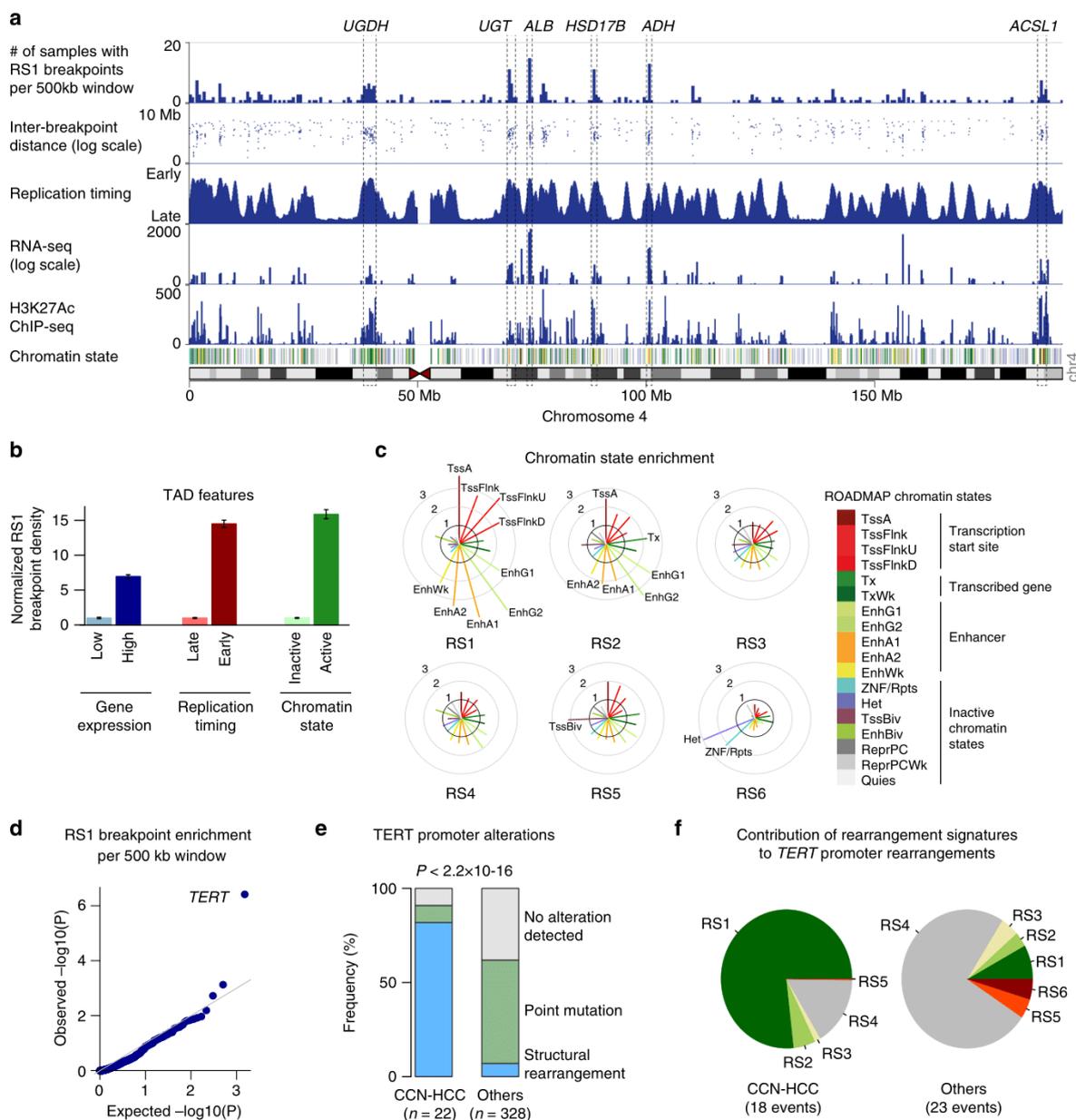
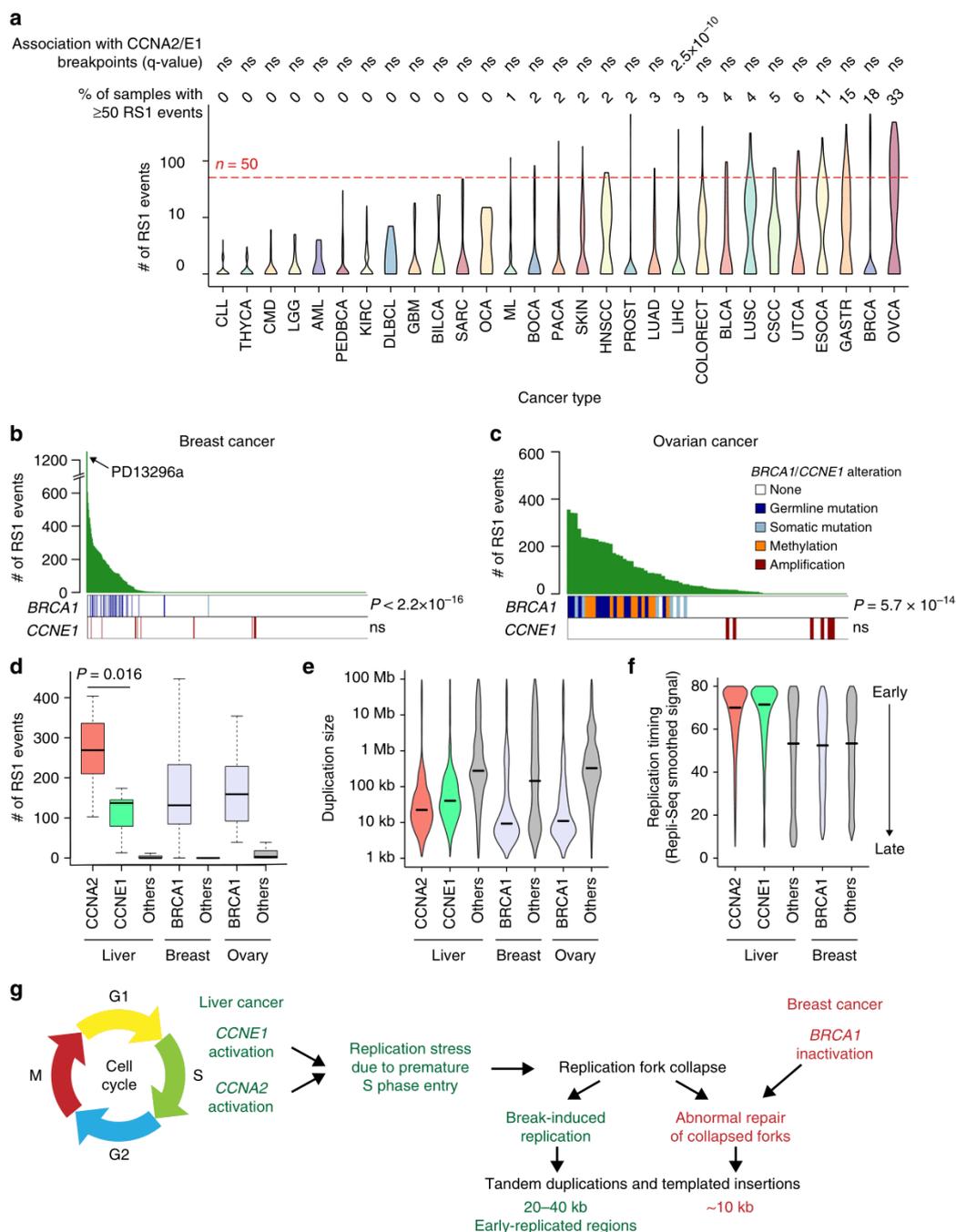


Fig. 5 Hotspot analysis of rearrangement signature 1 (RS1) breakpoints. **a** The density of RS1 breakpoints along chromosome 4 is displayed above replication timing, RNA-seq expression, H3K27Ac ChIP-seq profile and chromatin state. Replication timing was determined using Repli-Seq data from the liver cancer cell line HepG2. RNA-seq profile was generated from a normal liver sample. H3K27Ac and chromatin states in normal adult liver were obtained from the ROADMAP consortium. The legend for chromatin state color codes is displayed in **c**. Hotspots corresponding to highly expressed liver enzymes are annotated (UGDH, UDP-glucose 6-dehydrogenase; UGT, UDP glucuronosyltransferase family cluster; ALB, albumin; HSD17B, hydroxysteroid 17-Beta dehydrogenases 11 and 13, ADH, alcohol dehydrogenase cluster; ACSL1, acyl-CoA synthetase long chain family member 1). **b** RS1 breakpoint density in topologically associated domains (TADs). TADs were defined in human embryonic stem cells (H1) and classified based on gene expression in normal liver, replication timing and chromatin state. For each comparison, breakpoint density was normalized to be 1 in the group with the lowest density. Error bars indicate the 95% confidence interval. **c** Enrichment of rearrangement breakpoints in ROADMAP chromatin states for the 6 rearrangement signatures identified in HCC. For each signature, the fold-change between the observed and expected number of breakpoints falling within each chromatin state is represented, and chromatin states with a >2-fold enrichment are annotated. **d** Quantile-quantile plot of RS1 breakpoint enrichment p-values across 500 kb windows. **e** Proportion of *TERT* promoter alterations in CCN-HCC and other HCC analyzed by WGS. **f** Contribution of the 6 rearrangement signatures to *TERT* promoter rearrangements in CCN-HCC and other HCC



approved drugs designed to exploit synthetic lethality, have demonstrated benefit for patients carrying *BRCA1* mutations³⁸. CCN-HCC do not harbor a DNA repair defect but share with *BRCA1*-altered tumors a signature of genomic instability that could conceivably confer these tumors sensitivity to PARP inhibitors. Finally, there are currently several compounds in phase I and II trials targeting the replication stress response pathway members *ATR*, *CHK1* and *WEE1*³⁹. If brought to the clinic, such compounds would be promising for CCN-HCC treatment, given that the *ATR* pathway is strongly upregulated in CCN-HCC and

overexpression of *CCNE1* has been shown to confer increased sensitivity to *ATR* inhibition⁴⁰.

We describe for the first time recurrent fusions involving *CCNA2* gene and recurrent rearrangements of *CCNE1* promoter region. *CCNA2* fusions are only the second recurrent fusion event identified in hepatocellular carcinoma, after the *PRKACA-DNAJB1* fusion characteristic of the rare fibrolamellar carcinoma subtype⁴¹. These fusions always involve the untranslated 5' region of different partner genes upstream exons 3–8 of *CCNA2*, which constitutes an original mechanism leading to oncogene activation

ARTICLE

Fig. 6 Pan-cancer analysis of the RS1 signature **a** Violin plots representing the number of rearrangements attributed to signature RS1 across patients within each cancer type in the ICGC PCAWG data set. For each cancer type, we assessed the association between tumors with ≥ 50 RS1 events and tumors with a rearrangement breakpoint < 80 kb from *CCNA2* or *CCNE1* gene using Fisher’s exact tests. ns: not significant. The definition of cancer codes and number of samples per cancer type are available in Supplementary Data 9. **b** Number of RS1 events across 524 breast cancer genomes³⁰ and association with *BRCA1* alterations and *CCNE1* amplifications. PD13296a, the only tumor with both *BRCA1* mutation and *CCNE1* amplification, has the highest number of RS1 events in the series. **c** Number of RS1 events across 80 ovarian cancer genomes⁷⁵ and association with *BRCA1* alterations and *CCNE1* amplifications. *P*-values were obtained using one-sided Wilcoxon rank-sum tests. **d** Number of RS1 events in liver, breast and ovarian cancers with or without *CCNA2*, *CCNE1* and *BRCA1* alterations. The middle bar, median; box, interquartile range; bars extend to 1.5 times the interquartile range. **e**, Violin plots representing the distribution of tandem duplication sizes across liver, breast and ovarian cancers with or without *CCNA2*, *CCNE1* and *BRCA1* alterations. **f** Violin plots representing the replication timing of duplication and inter-chromosomal translocation breakpoint loci in liver and breast cancers with or without *CCNA2*, *CCNE1* and *BRCA1* alterations. Replication timing was determined using Repli-Seq data from the HepG2 cell line for liver cancer and from the MCF-7 cell line for breast cancer. **g** Proposed connexion between rearrangement signatures in CCN-HCC and in *BRCA1*-inactivated breast and ovarian cancers

by truncating a regulatory N-terminal domain. Apart from liver cancers, none of the 2606 tumor genomes from the ICGC PCAWG dataset displayed a rearrangement breakpoint in *CCNA2* intron 2. Consistently, a recent RNA-seq analysis of 9,624 TCGA samples from 33 cancer types⁴² did not reveal any *CCNA2* fusion in other cancer types. *CCNA2* fusions thus appear to be specific of liver cancers. Rearrangements affecting *CCNE1* promoter region result in the overexpression of cyclin E1 by bringing active enhancer regions upstream the transcription start site, mirroring the effect of viral enhancers. This mechanism was more frequent than *CCNE1* amplification in the liver cancer series we analyzed. Although HBV and AAV2 insertions were previously identified in *CCNA2* and *CCNE1*^{14,16}, the functional consequences of these insertions were unknown. By integrating WGS and RNA-seq data, we demonstrate here that viral insertions in *CCNA2*, like gene fusions, induce abnormal transcripts leading to truncated proteins lacking N-terminal regulatory domains. By contrast, viral insertions in *CCNE1* region lead to the overexpression of a full-length transcript and protein.

CCN-HCC display a characteristic transcriptional program, with a strong overexpression of E2F targets. Activation of the E2F pathway is expected in *RBI*-altered tumors and was already described in HCC⁴³. However, E2F pathway is also activated in CCN-HCC without *RBI* inactivation event and may be partly explained by the ability of cyclin E/Cdk2 complexes to phosphorylate Rb. Interestingly, E2F-1 overexpression in the liver causes dysplasia and tumors in mice⁴³, and E2F1 was shown to inhibit c-Myc-driven apoptosis by activating PIK3CA/Akt/mTOR and c-Myb/COX-2 pathways⁴⁴.

A striking feature of CCN-HCC is the accumulation of hundreds of tandem duplications and templated insertion cycles. A recent study showed that *CCNE1* activation in U2OS cell lines leads to shortened G1 phase, early S phase entry and firing of normally silenced replication origins in highly expressed genes, prone to collapse and associated with DNA double-strand breaks formation³². Double-strand breaks formed following replication fork breakdown are primarily repaired by break-induced replication (BIR)⁴⁵. In a cyclin E overexpression model of DNA replication stress, BIR was shown to be required for cell cycle progression and to induce duplications < 200 kb⁴⁶. In addition, template switching may occur during BIR and generate complex chromosome rearrangements^{24,25,47}. Thus, the nature of rearrangements identified in CCN-HCC and the enrichment of breakpoints in early-replicated, actively transcribed regions are consistent with a BIR mechanism induced by replication stress. However, future studies addressing the precise molecular mechanism generating templated insertions will be crucial to fully understand the relationship between replication stress and the RS1 rearrangement signature. The mechanism of tandem duplication formation in *BRCA1*-mutant cells was recently identified³⁵. It involves abnormal repair of collapsed replication forks by a

“replication restart bypass” mechanism with extension of the stalled leading strand by a migration bubble mechanism similar to BIR⁴⁸, terminated by end joining or by microhomology-mediated template switching. Thus, structural rearrangements induced by cyclin activation and *BRCA1* deficiency are initiated by replication fork collapse and processed by different repair mechanisms leading to a similar rearrangement signature with subtle differences regarding the size of rearrangements and breakpoint location. Interestingly, *BRCA1* inactivation and *CCNE1* amplification are mutually exclusive in ovarian cancers⁴⁹, and have been shown to be synthetically lethal⁵⁰. The single breast tumor that we identified with both *BRCA1* mutation and *CCNE1* amplification (PD13296a) had the highest number of rearrangements related to the RS1 signature ($n = 1221$) across all the tumors we analyzed.

Contrary to *CCNA2* alterations that seem to be specific of liver cancers, *CCNE1* activation by high-level amplification is frequent across human cancers, in particular in gynecologic cancers⁵¹. Yet, *CCNE1* amplification in breast and ovarian cancers does not lead to the rearrangement phenotype that we observed in CCN-HCC. Several reasons may explain this discrepancy. First, adult hepatocytes are quiescent, rarely divide, and may thus be particularly sensitive to replication stress. Second, *CCNE1* is mostly activated by viral insertions and structural rearrangements of regulatory regions in HCC, rather than chromosome amplifications. These alterations may not have exactly the same functional consequence. Finally, we believe that viral insertions and structural rearrangements activating *CCNA2* or *CCNE1* are early events triggering hepatocarcinogenesis because they occur in patients without cirrhosis and in absence of other oncogenic event like *CTNGB1* mutations. *CCNE1* amplifications may occur later in breast and ovarian tumors, not leaving enough time for rearrangements to accumulate. Fujimoto et al. reported a positive correlation between the number of structural rearrangements and HBV insertion sites, suggesting that double-strand breaks generated by structural rearrangements may provide opportunities for HBV integration¹¹. Here we describe the reciprocal relationship where viral insertions in cyclin genes lead to structural rearrangement formation due to replication stress.

The propensity of signature RS1 breakpoints to occur in enhancer-rich regions makes these rearrangements likely to activate oncogenes in trans. In this limited series of 22 CCN-HCC analyzed by WGS, we identified a single significantly recurrent hotspot at *TERT* promoter. However, the power to identify recurrent somatic rearrangement hotspots increases sharply with sample size²⁷, and future studies of larger CCN-HCC series may uncover additional sites under positive selection in CCN-HCC.

In conclusion, viral insertions and structural rearrangements activating *CCNA2* and *CCNE1* define a homogeneous subgroup of aggressive HCC developed in non-cirrhotic liver, sharing similar transcriptional profiles and frequent inactivation of *RBI*

and *PTEN*. These tumors display a specific rearrangement signature induced by replication stress that sustains tumor growth by activating *TERT* but may constitute a targetable vulnerability.

Methods

Description of the LICA-FR cohort. A series of 160 hepatocellular carcinoma (HCC) samples and their non-tumor counterparts were collected from patients surgically treated in four French hospitals located in Bordeaux and Paris region. The study was approved by institutional review board committees (CCPRB Paris Saint-Louis, 1997, 2004, and 2010, approval number 01-037; Bordeaux, 2010-A00498-31). Written informed consent was obtained in accordance with French legislation. All samples were immediately frozen in liquid nitrogen and stored at -80°C . HCC were enriched in cases developed on a non-cirrhotic liver (107/160, 67%): 75 tumors developed in non-fibrotic (METAVIR F0-F1), 32 in chronic hepatitis (F2-F3) and 53 in cirrhotic liver (F4). Clinicopathological data were available for all cases. A diversity of risk factors were represented in our series, including alcohol ($n = 63$), metabolic syndrome ($n = 37$), HBV ($n = 30$), and HCV infection ($n = 30$). Twenty-nine patients had none of the above risk factors. These 160 samples were analyzed by RNA sequencing, 156 were analyzed by whole exome sequencing (including 96 were previously published¹⁰) and 45 by whole genome sequencing (35 were previously published²²). Detailed clinical characteristics and sequencing details for each sample are provided in Supplementary Data 1.

Whole genome sequencing. Whole genome data from 45 tumors of the LICA-FR series were analyzed in this study, comprising 35 previously published²² and 10 new cases. The whole genomes of 10 new tumor/normal pairs were sequenced for this project at the Center National de Recherche en Génomique Humaine (CNRGH, Evry, France) on an Illumina HiSeq X Five as paired-end 151 bp reads. Sequences were aligned to the hg19 version of the human genome using BWA⁵² version 0.7.12. We used Picard tools version 1.108 (<http://broadinstitute.github.io/picard/>) to remove PCR duplicates and GATK⁵³ version v3.5 for local indel realignment and base quality recalibration, as recommended in GATK best practices⁵⁴. We obtained an average depth of 119-fold for tumors (range 104–126) and 41-fold for matched non-tumor liver samples (range 38–43).

Whole exome sequencing. Whole exome data from 156 tumors of the LICA-FR series were analyzed in this study, comprising 96 previously published¹⁰ and 60 new cases. Sequence capture, enrichment and elution of genomic DNA samples from the 60 new tumor/normal pairs was performed by IntegraGen (Evry, France). Agilent in-solution enrichment was used with the manufacturer's biotinylated oligonucleotide probe library SureSelect Human All-Exon kit v5 + UTRs ($n = 39$) or SureSelect Clinical Research Exome V2 ($n = 21$) according to the manufacturer's instructions. The eluted enriched DNA sample was sequenced on an Illumina HiSeq 2000 ($n = 39$) or HiSeq 4000 ($n = 21$) as paired-end 75 bp reads. Sequencing details for each sample are indicated in Supplementary Data 1.

Somatic mutation calling. We used MuTect2 to call somatic mutations from WES and WGS data by comparing each tumor sample with its matched non-tumor counterpart and a panel of normals (PON) file. We excluded mutations belonging to the ENCODE Data Analysis Consortium blacklisted regions (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeMapability/wgEncodeDacMapabilityConsensusExcludable.bed.gz>) and regions covered by < 6 reads in the tumor or normal sample. We then selected only single nucleotide variants (SNVs) with a MuTect2 flag among "PASS", "clustered_events", "lod_fstar", "alt_allele_in_normal" or "homologous_mapping_event" and small insertions and deletions (indels) with a MuTect2 flag among "PASS", "clustered_events" or "str_contraction". To improve specificity in the calling of mutations with low variant allele frequency (VAF), we quantified the number of high quality variant reads in the tumor (mapping quality ≥ 20 , base quality ≥ 20) and the number of variant reads in the non-tumor sample with no quality threshold using bamreadcount (<https://github.com/genome/bam-readcount>). Only variants matching the following criteria were finally retained: VAF $\geq 2\%$ in the tumor with ≥ 3 variant reads, VAF $\leq 5\%$ in the non-tumor samples with ≤ 2 variant reads, and a VAF ratio ≥ 5 between the tumor and non-tumor sample.

Copy-number and structural rearrangement analysis. We used MANTA⁵⁵ software to identify somatic structural rearrangements in WGS data. To keep only the most reliable events, we selected only rearrangements supported by ≥ 10 reads and with a variant allele fraction $\geq 5\%$. We used cgpBattenberg⁵⁶ algorithm to reconstruct copy-number profiles from WGS data. We used the circular binary segmentation algorithm implemented in the Bioconductor package DNACopy⁵⁷ to reconstruct copy-number profiles from WES data.

RNA sequencing. RNA samples from the 160 tumors of the LICA-FR series were sequenced in several batches with slightly different protocols. RNA samples were enriched for polyadenylated RNA from 5 μg of total RNA, and the enriched samples were used to generate sequencing libraries with the Illumina TruSeq or

Illumina TruSeq Stranded mRNA kit and associated protocol as provided by the manufacturer. Libraries were sequenced by IntegraGen (Evry, France) on an Illumina HiSeq 2000 or 4000 as paired-end 75 or 100 bp reads. Full Fastq files were aligned to the reference human genome hg19 using TopHat2⁵⁸. Sequencing details for each sample and the parameters used for TopHat2 are indicated in Supplementary Data 1. We removed reads mapping to multiple locations, and we used HTSeq⁵⁹ to obtain the number of reads associated to each gene in the Gencode v19 database, restricting to protein-coding genes, pseudogenes, antisense and lincRNAs ($n = 42540$). We used the Bioconductor DESeq2 package⁶⁰ to import raw HTSeq counts for each sample into R statistical software and apply variance stabilizing transformation (VST) to the raw count matrix. FPKM scores (number of fragments per kilobase of exon model and millions of mapped reads) were calculated by normalizing the count matrix for the library size and the coding length of each gene. We used the area under the ROC curve (AUC) to identify and remove 2724 genes with a significant batch effect (AUC > 0.95 between one sequencing project and others).

Gene fusion detection. Fusions detected by TopHat2 (--fusion-search --fusion-min-dist 2000 --fusion-anchor-length 13 --fusion-ignore-chromosomes chrM) were filtered using the TopHatFusion-post algorithm. We kept only fusions validated by BLAST and with at least 10 split-reads or pairs of reads spanning the fusion event, and we removed fusions identified at least twice in a cohort of 36 normal liver samples.

Gene expression analysis. We used t-distributed stochastic neighbor embedding (t-SNE) to classify HCC based on their gene expression profiles. We selected the 1000 most variably expressed genes, and we used 1 minus the weighted Pearson correlation coefficient as the distance measure. Pairwise Pearson correlation was calculated using the wtd.cors function of the weights R package. We used standard deviation subtracted by 0.2 as the weight, giving more variable genes greater influence. The resulting distance matrix was used to perform the t-SNE analysis using the R package Rtsne⁶¹ with default parameters except the following: theta = 0, is_distance = T, pca = F, max_iter = 2000. We used the Bioconductor limma package⁶² to test for differential expression between CCN2-HCC and other HCC of all genes expressed in at least five samples (FPKM > 0). We applied a q -value threshold of ≤ 0.05 to define differentially expressed genes. We used an in-house adaptation of the GSEA method⁶³ to identify gene sets from the MSigDB v6 database overrepresented among upregulated and downregulated genes.

Viral insertion screening. AAV2 insertions had previously been screened by viral capture and whole exome sequencing in 83 tumors from the LICA-FR cohort¹⁶. We extended this screen to AAV2 and HBV insertions in all HCC from the LICA-FR cohort using RNA-seq and WES data. In the ICGC-JP cohort, AAV2 and HBV insertions had already been screened using WGS data and were provided by Fujimoto et al.¹¹ In the TCGA cohort, we screened AAV2 and HBV insertions using RNA-seq data from all tumors and WES data from 37 tumors showing viral reads or overexpression of *CCNA2* or *CCNE1* in RNA-seq data. For each tumor and matched normal sample, the sequence reads were mapped to the AAV2 (AF043303.1) and HBV (X02763, renumbered using the EcoRI restriction site as the +1) reference genomes using BWA⁵². Read pairs with at least one read aligned on the virus were extracted using samtools⁶⁴, and aligned to a custom reference genome including human chromosomes and virus fasta sequences as pseudo-chromosomes. Tumors with ≥ 6 chimeric reads or read pairs aligned on both the human and viral genomes were further analyzed. All viral insertions were validated by visual inspection on IGV⁶⁵. We used chimeric reads to identify insertion breakpoints at base resolution by mapping sequences on both sides of the junctions. Of the 12 LICA-FR tumors with viral insertions detected in *CCNA2* or *CCNE1*, 7 were previously analyzed by viral capture sequencing¹⁶ and 3 were analyzed by whole genome sequencing. For these 10 tumors, we were able to extract reads covering the full length of the inserted viral genome and to reconstruct the complete human-virus-human chimeric sequence.

Consequences of cyclin A2 alterations on protein structure. All tumors from the LICA-FR series harboring AAV2 or HBV insertions in *CCNA2* were analyzed by WGS or viral capture¹⁶ to determine the precise boundaries of viral insertion breakpoints. RNA-seq reads were then aligned on the reconstructed chimeric sequence with TopHat2⁵⁸, and we used Cufflinks v2.2.1⁶⁶ to identify and quantify the different transcripts. We used ElemeNT⁶⁷ to predict transcription initiation sites and Alamut Visual software (Interactive Biosoftware) to identify splicing signals on the chimeric DNA sequence. We used ATGpr⁶⁸ to identify translation initiation sites on abnormal transcripts resulting from viral insertion or gene fusions.

Western blot analysis of cyclin A2 and cyclin E1 proteins. Cell protein extracts were prepared using hot Laemmli buffer (50 mM Tris, pH = 6.8, 2% SDS, 5% glycerol, 2 mM DTT, 2.5 mM EDTA, 2.5 mM EGTA, Protease inhibitor cocktail complete MINI EDTA-free (Roche Applied Science), 1 \times HALT Phosphatase inhibitor (Perbio), 2 mM Na₃VO₄ and 10 mM NaF). Protein concentration was assessed using the BCA Protein Assay Kit (Pierce). Western blot analyses were

conducted using the following primary antibodies: CCNA2 N-ter (#211735, Abcam); CCNA2 C-ter (#32386, Abcam), CCNE1 (#33911, Abcam), and β -actin (#4967, Cell Signaling Technology) used as loading control. Proteins of interest were detected using an anti-rabbit IgG horseradish peroxidase-linked secondary antibody (#7074, Cell Signaling Technology) and the ECL Chemiluminescence Western Blotting Detection Kit (GE Healthcare), according to the provided protocol. Signal detection was performed using the ChemiDoc XRS system and the Image Lab software (Bio-Rad). All antibodies were used at 1:1000 dilution except secondary antibody, which was used at 1:2000.

Mutational and rearrangement signature analysis. We used the *Palimpsest* R package⁶⁹ to extract mutational and rearrangement signatures from WGS data. For point mutations, we quantified the contribution of the 10 mutational signatures referenced on the COSMIC website (<https://cancer.sanger.ac.uk/cosmic/signatures>) and described as operative in liver cancers (signatures 1, 4, 5, 6, 12, 16, 17, 22, 23, 24)²² to each tumor genome. For structural rearrangements, we performed a de novo signature analysis across the 350 HCC genomes from the LICA-FR, TCGA and ICGC-JP datasets. We identified 6 rearrangement signatures that were very similar to the 6 signatures we previously obtained on a smaller dataset²², except that the two initially described deletion signatures were now merged into signature RS5, and that a new signature emerged (RS6, dominated by inversions < 10 kb). We used *Palimpsest* to quantify the contribution of each signature to each tumor genome and to estimate the probability of each structural rearrangement being due to each process.

Identification of rearrangement hotspots. We identified 8466 breakpoints attributed to signature RS1 (probability > 0.5) across the 350 HCC genomes from the LICA-FR, TCGA and ICGC-JP datasets. To account for the uneven distribution of rearrangements in the genome, we then modeled the background distribution of breakpoints considering various genomic features as described by Glodzik et al.²⁷, with some modifications. In short, we divided the genome into 500 kb bins, and we characterized for each bin 17 genomic features likely to influence the density of rearrangements: replication timing in HepG2 cell line (ENCODE⁷⁰), highly expressed (top 25%) and low-expressed (remaining 75%) genes in normal liver, average copy-number in the cohort, repetitive sequences (segmental duplications, ALU elements and other repeats), number of N bases in the reference genome, known fragile sites⁷¹, chromatin staining, DNase hyper-sensitive sites and 6 histone marks (H3K4me1, H3K4me3, H3K9me3, H3K27me3, H3K36me3, H3K27ac) in adult liver (ROADMAP⁷²). All features were normalized to a mean of 0 and standard deviation of 1 across the bins. The total number of RS1 breakpoints were counted for each bin, and we used negative binomial regression to model the distribution of breakpoints according to the 17 normalized features. The model was trained across 4993 bins after removing bins containing validated cancer genes from the Cancer Gene Census⁷³ (<https://cancer.sanger.ac.uk/census>). For signature RS1, the most predictive features of a high breakpoint density were DNase accessibility, H3K27 acetylation and early replication timing. We then used this model to estimate the expected number of breakpoints across 761 bins containing cancer genes, and we compared the number of observed breakpoints to the number of expected breakpoints using a one-sided binomial test. Finally, p-values were corrected for multiple testing using Benjamini-Hochberg procedure.

Chromatin state analysis. We used various genomic features to correlate with structural rearrangement density and to better understand the functional consequences of rearrangements. We used replication sequencing (Repli-seq) wavelet-smoothed signals downloaded generated by the ENCODE⁷⁰ consortium for the liver cancer cell line HepG2 to define early and late-replicating regions. We used ChIP-seq data for various histone modifications (H3K4me1, H3K4me3, H3K9me3, H3K27me3, H3K36me3, H3K27ac) and chromatin states derived from these modifications in normal adult liver by the ROADMAP consortium⁷². Topologically associated domain (TAD) boundaries in human embryonic stem cells (H1) were provided by Tsigirgos et al.⁷⁴

Pan-cancer analysis of structural rearrangement signatures. Somatic structural rearrangements called by a uniform pipeline over 2,606 tumor genomes were downloaded from the ICGC PanCancer Analysis of Whole Genomes (PCAWG) project^{23,28,29}. Using *Palimpsest*⁶⁹, we identified 9 rearrangement signatures in this data set, including one (RS1-pancan) very similar to the RS1 signature identified in CCN-HCC, and we quantified the contribution of each signature to each tumor genome. In each cancer type, we tested if the presence of ≥ 50 rearrangements attributed to signature RS1-pancan was associated with the presence of rearrangement breakpoints < 80 kb from *CCNA2* or *CCNE1* gene using Fisher's exact test. We analyzed two additional series of breast ($n = 524$)³⁰ and ovarian ($n = 80$)⁷⁵ cancer genomes to correlate the amount of RS1-pancan events with *CCNE1* amplifications and *BRCA1* alterations.

Clinical associations. We tested the association of CCN-HCC in the LICA-FR cohort with gender, age, etiology, liver fibrosis, Edmonson grade, and vascular invasion using Wilcoxon rank sum test for continuous variables, Fisher's exact test for binary variables and Chi square test for trend for categorical variables. We used

log-rank test and Kaplan–Meier method to compare overall survival between CCN-HCC and others, considering only HCC with curative resection (R0) and excluding patients who died within 3 months after surgery.

Computing codes. The functions used to perform the signatures analysis and associated figures are available as an open-source R package, *Palimpsest*, available on Github: <https://github.com/FunGeST/Palimpsest>.

URLs. ICGC data portal, <https://dcc.icgc.org/>; COSMIC database, <https://cancer.sanger.ac.uk/cosmic/>; ENCODE project, <https://www.encodeproject.org/>; GENCODE v19, <http://www.gencodegenes.org/releases/19.html>; ROADMAP project, <http://www.roadmapepigenomics.org/>; NCI GDC data portal, <https://portal.gdc.cancer.gov/>.

Data availability

The sequencing data reported in this paper have been deposited to the EGA (European Genome-phenome Archive) database (RNA-seq accession [EGAS00001002879]; WES accessions [EGAS00001000217], [EGAS00001001002] and [EGAS00001003063]; WGS accessions [EGAS00001002408], [EGAS00001000706] and [EGAS00001002888]) and the International Cancer Genome Consortium (ICGC) data portal (<http://dcc.icgc.org/>; release 27, April 2018).

Received: 23 July 2018 Accepted: 8 November 2018
Published online: 07 December 2018

References

- European Association for the Study of the Liver & European Organisation for Research and Treatment of Cancer EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. *J. Hepatol.* **56**, 908–943 (2012).
- Llovet, J. M. et al. Sorafenib in advanced hepatocellular carcinoma. *New Engl. J. Med.* **359**, 378–390 (2008).
- Bruix, J. et al. Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet* **389**, 56–66 (2017).
- Llovet, J. M. & Hernandez-Gea, V. Hepatocellular carcinoma: reasons for phase III failure and novel perspectives on trial design. *Clin. Cancer Res.* **20**, 2072–2079 (2014).
- Llovet, J. M. et al. Hepatocellular carcinoma. *Nat. Rev. Dis. Prim.* **2**, 16018 (2016).
- Nault, J. C. et al. Telomerase reverse transcriptase promoter mutation is an early somatic genetic alteration in the transformation of premalignant nodules in hepatocellular carcinoma on cirrhosis. *Hepatology* **60**, 1983–1992 (2014).
- Guichard, C. et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat. Genet.* **44**, 694–698 (2012).
- Fujimoto, A. et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* **44**, 760–764 (2012).
- Ahn, S.-M. et al. Genomic portrait of resectable hepatocellular carcinomas: Implications of RB1 and FGF19 aberrations for patient stratification. *Hepatology* **60**, 1972–1982 (2014).
- Schulze, K. et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat. Genet.* **47**, 505–511 (2015).
- Fujimoto, A. et al. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet.* **48**, 500–509 (2016).
- Zucman-Rossi, J., Villanueva, A., Nault, J.-C. & Llovet, J. M. Genetic landscape and biomarkers of hepatocellular carcinoma. *Gastroenterology* **149**, 1226–1239.e4 (2015).
- Wang, J., Chenivresse, X., Henglein, B. & Br  chet, C. Hepatitis B virus integration in a cyclin A gene in a hepatocellular carcinoma. *Nature* **343**, 555–557 (1990).
- Sung, W.-K. et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat. Genet.* **44**, 765–769 (2012).
- Ding, D. et al. Recurrent targeted genes of hepatitis B virus in the liver cancer genomes identified by a next-generation sequencing-based approach. *PLoS Genet.* **8**, e1003065 (2012).
- Nault, J.-C. et al. Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat. Genet.* **47**, 1187–1193 (2015).
- Cancer Genome Atlas Research Network. Electronic address: wheeler@bcm.edu & Cancer Genome Atlas Research Network. Comprehensive and

integrative genomic characterization of hepatocellular carcinoma. *Cell* **169**, 1327–1341.e23 (2017).

18. Geley, S. et al. Anaphase-promoting complex/cyclosome-dependent proteolysis of human cyclin A starts at the beginning of mitosis and is not subject to the spindle assembly checkpoint. *J. Cell. Biol.* **153**, 137–148 (2001).
19. Fung, T. K., Yam, C. H. & Poon, R. Y. C. The N-terminal regulatory domain of cyclin A contains redundant ubiquitination targeting sequences and acceptor sites. *Cell Cycle* **4**, 1411–1420 (2005).
20. Sage, J., Miller, A. L., Pérez-Mancera, P. A., Wysocki, J. M. & Jacks, T. Acute mutation of retinoblastoma gene function is sufficient for cell cycle re-entry. *Nature* **424**, 223–228 (2003).
21. Garcia-Cao, I. et al. Systemic elevation of PTEN induces a tumor suppressive metabolic state. *Cell* **149**, 49–62 (2012).
22. Letouzé, E. et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat. Commun.* **8**, 1315 (2017).
23. Li, Y. et al. Patterns of structural variation in human cancer. Preprint at <https://www.biorxiv.org/content/early/2017/08/27/181339>, <https://doi.org/10.1101/181339> (2017).
24. Lee, J. A., Carvalho, C. M. B. & Lupski, J. R. A. DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* **131**, 1235–1247 (2007).
25. Hastings, P. J., Ira, G. & Lupski, J. R. A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.* **5**, e1000327 (2009).
26. Carvalho, C. M. B. et al. Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. *Nat. Genet.* **43**, 1074–1081 (2011).
27. Glodzik, D. et al. A somatic-mutational process recurrently duplicates germline susceptibility loci and tissue-specific super-enhancers in breast cancers. *Nat. Genet.* **49**, 341–348 (2017).
28. Campbell, P. J., Getz, G., Stuart, J. M., Korbel, J. O. & Stein, L. D. Pan-cancer analysis of whole genomes. Preprint at <https://www.biorxiv.org/content/early/2017/07/12/162784>, <https://doi.org/10.1101/162784> (2017).
29. Wala, J. A. et al. Selective and mechanistic sources of recurrent rearrangements across the cancer genome. Preprint at <https://www.biorxiv.org/content/early/2017/09/14/187609>, <https://doi.org/10.1101/187609> (2017).
30. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
31. Waszak, S. M. et al. Germline determinants of the somatic mutation landscape in 2,642 cancer genomes. Preprint at <https://www.biorxiv.org/content/early/2017/11/01/208330>, <https://doi.org/10.1101/208330> (2017).
32. Macheret, M. & Halazonetis, T. D. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature* **555**, 112–116 (2018).
33. Schlacher, K., Wu, H. & Jasin, M. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. *Cancer Cell* **22**, 106–116 (2012).
34. Pathania, S. et al. BRCA1 haploinsufficiency for replication stress suppression in primary cells. *Nat. Commun.* **5**, 5496 (2014).
35. Willis, N. A. et al. Mechanism of tandem duplication formation in BRCA1-mutant cells. *Nature* **551**, 590–595 (2017).
36. Forment, J. V. & O'Connor, M. J. Targeting the replication stress response in cancer. *Pharmacol. Ther.* <https://doi.org/10.1016/j.pharmthera.2018.03.005> (2018).
37. Menghi, F. et al. The tandem duplicator phenotype as a distinct genomic configuration in cancer. *Proc. Natl Acad. Sci. USA* **113**, E2373–E2382 (2016).
38. Lord, C. J. & Ashworth, A. PARP inhibitors: synthetic lethality in the clinic. *Science* **355**, 1152–1158 (2017).
39. O'Connor, M. J. Targeting the DNA damage response in cancer. *Mol. Cell* **60**, 547–560 (2015).
40. Toledo, L. I. et al. A cell-based screen identifies ATR inhibitors with synthetic lethal properties for cancer-associated mutations. *Nat. Struct. Mol. Biol.* **18**, 721–727 (2011).
41. Honeyman, J. N. et al. Detection of a recurrent DNAJB1-PRKACA chimeric transcript in fibrolamellar hepatocellular carcinoma. *Science* **343**, 1010–1014 (2014).
42. Gao, Q. et al. Driver fusions and their implications in the development and treatment of human cancers. *Cell Rep.* **23**, 227–238.e3 (2018).
43. Conner, E. A. et al. Dual functions of E2F-1 in a transgenic mouse model of liver carcinogenesis. *Oncogene* **19**, 5054–5062 (2000).
44. Ladu, S. et al. E2F1 inhibits c-Myc-driven apoptosis via PIK3CA/Akt/mTOR and COX-2 in a mouse model of human liver cancer. *Gastroenterology* **135**, 1322–1332 (2008).
45. Kramara, J., Osia, B. & Malkova, A. Break-induced replication: the where, the why, and the how. *Trends Genet.* **34**, 518–531 (2018).
46. Costantino, L. et al. Break-induced replication repair of damaged forks induces genomic duplications in human cells. *Science* **343**, 88–91 (2014).
47. Smith, C. E., Llorente, B. & Symington, L. S. Template switching during break-induced replication. *Nature* **447**, 102–105 (2007).
48. Saini, N. et al. Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature* **502**, 389–392 (2013).
49. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).
50. Etemadmoghadam, D. et al. Synthetic lethality between CCNE1 amplification and loss of BRCA1. *Proc. Natl Acad. Sci. USA* **110**, 19489–19494 (2013).
51. Sanchez-Vega, F. et al. Oncogenic signaling pathways in The Cancer Genome Atlas. *Cell* **173**, 321–337.e10 (2018).
52. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
53. McKenna, A. et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
54. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
55. Chen, X. et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**, 1220–1222 (2016).
56. Nik-Zainal, S. et al. The life history of 21 breast cancers. *Cell* **149**, 994–1007 (2012).
57. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
58. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
59. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
60. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
61. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
62. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
63. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
64. Li, H. et al. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
65. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
66. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
67. Slutskin, A. et al. ElemeNT: a computational tool for detecting core promoter elements. *Transcription* **6**, 41–50 (2015).
68. Nishikawa, T., Ota, T. & Isogai, T. Prediction whether a human cDNA sequence contains initiation codon by combining statistical information and similarity with protein sequences. *Bioinformatics* **16**, 960–967 (2000).
69. Shinde, J. et al. Palimpsest: an R package for studying mutational and structural variant signatures along clonal evolution in cancer. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bty388> (2018).
70. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
71. Bignell, G. R. et al. Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893–898 (2010).
72. Roadmap Epigenomics Consortium. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
73. Futreal, P. A. et al. A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
74. Gong, Y. et al. Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries. *Nat. Commun.* **9**, 542 (2018).
75. Patch, A.-M. et al. Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489–494 (2015).

Acknowledgements

We thank Hidewaki Nakagawa for fruitful discussions and providing *TERT* promoter mutation data for the ICGC-JP series. We thank Rameen Beroukhi and Joachim Weischenfeldt for helping us access ICGC structural variant tables, and Aristotelis Tsigirios for providing topologically associated domain boundaries. We thank Tatiana Popova and Céline Vallot for critical discussion of the results. We thank the principal investigators of the liver cancer TCGA (Lewis Roberts, David Wheeler) and ICGC-JP (Tatsuhiko Shibata, Hidewaki Nakagawa) projects, and the ICGC consortium as a whole for providing the high quality data sets used in this study. We thank all the clinician surgeons and pathologists who have participated to this work: Jean Saric, Christophe Laurent, Laurence Chiche, Brigitte Le Bail, Claire Castain (CHU Bordeaux), Alexis Laurent, Daniel Cherqui, Daniel Azoulay (CHU Henri Mondor, Créteil,

ARTICLE

APHP), Marianne Ziol, Nathalie Ganne-Carrié and Pierre Nahon (Jean Verdier Hospital, Bondy, APHP). We also thank the Réseau national CRB Foie (BB-0033-0085), the tumor banks of CHU Bordeaux (BB-0033-00036), Jean Verdier Hospital (APHP) and CHU Henri Mondor (APHP) for contributing to the tissue collection. This work was supported by INCa within the ICGC project, MUTHEC project (INCa translationnel PRTK2014), France Génomique, Cancéropole Ile de France (ExhauTrans project), ITMO Cancer AVIESAN (Alliance Nationale pour les Sciences de la Vie et de la Santé, National Alliance for Life Sciences & Health) within the framework of the Cancer Plan ("HTE program-HetColi network" and "Cancer et environnement program"), BPI France (ICE project), ANRS and the French Liver Biobanks network – INCa, BB-0033-00085, Hepatobio bank. The group is supported by the Ligue Nationale Contre le Cancer (Equipe Labellisée), Labex Oncolmmunology (investissement d'avenir), Coup d'Élan de la Fondation Bettencourt-Shueller, the SIRIC CARPEM and Fondation Mérieux. QB and LM are supported by a fellowship from the HOB doctoral school and the ministry of Education and Research, TLB is supported by an "Attractivité IDEX" fellowship from IUH and CP is supported by a doctoral fellowship funded by ANRS.

Author contributions

J.Z.-R. and E.L. conceived and directed the research. Q.B., L.M., C.P., S.I., J.Z.-R., and E.L. designed the study and wrote the manuscript. C.P., I.M., G.C., and T.L.-B. performed the experiments. Q.B., L.M., C.P., I.M., T.L.-B., S.I., J.Z.-R., and E.L. analyzed and interpreted the data. D.B., V.M., and J.-F.D. generated whole-genome sequencing data. Q.B., L.M., C.P., V.R., J.S., E.T., D.B., V.M. S.I., and E.L. performed bioinformatics and statistical analysis. J.-C.N., G.A., A.D.-V., P.B.-S., O.S., J.-F.B., and J.C. provided essential biological resources and collected clinical data. All authors approved the final manuscript and contributed to critical revisions to its intellectual context.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-07552-9>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

Supplementary Information

Cyclin A2 and E1 genomic alterations define a specific subclass of hepatocellular carcinomas

Bayard *et al.*

Supplementary Figures

Supplementary Figure 1. Consequences of viral insertions in *CCNA2* gene

Supplementary Figure 2. Deletions associated with *CCNA2* deregulation

Supplementary Figure 3. Cyclin E1 overexpression induced by viral insertions and structural rearrangements

Supplementary Figure 4. Mutational signature analysis of CCN-HCC

Supplementary Figure 5. SNP array analysis of focal duplications in CCN-HCC from the TCGA series

Supplementary Figure 6. Characteristic copy-number profile of CCN-HCC

Supplementary Figure 7. Examples of intra-chromosomal templated insertions and templated insertion cycles involving several chromosomes

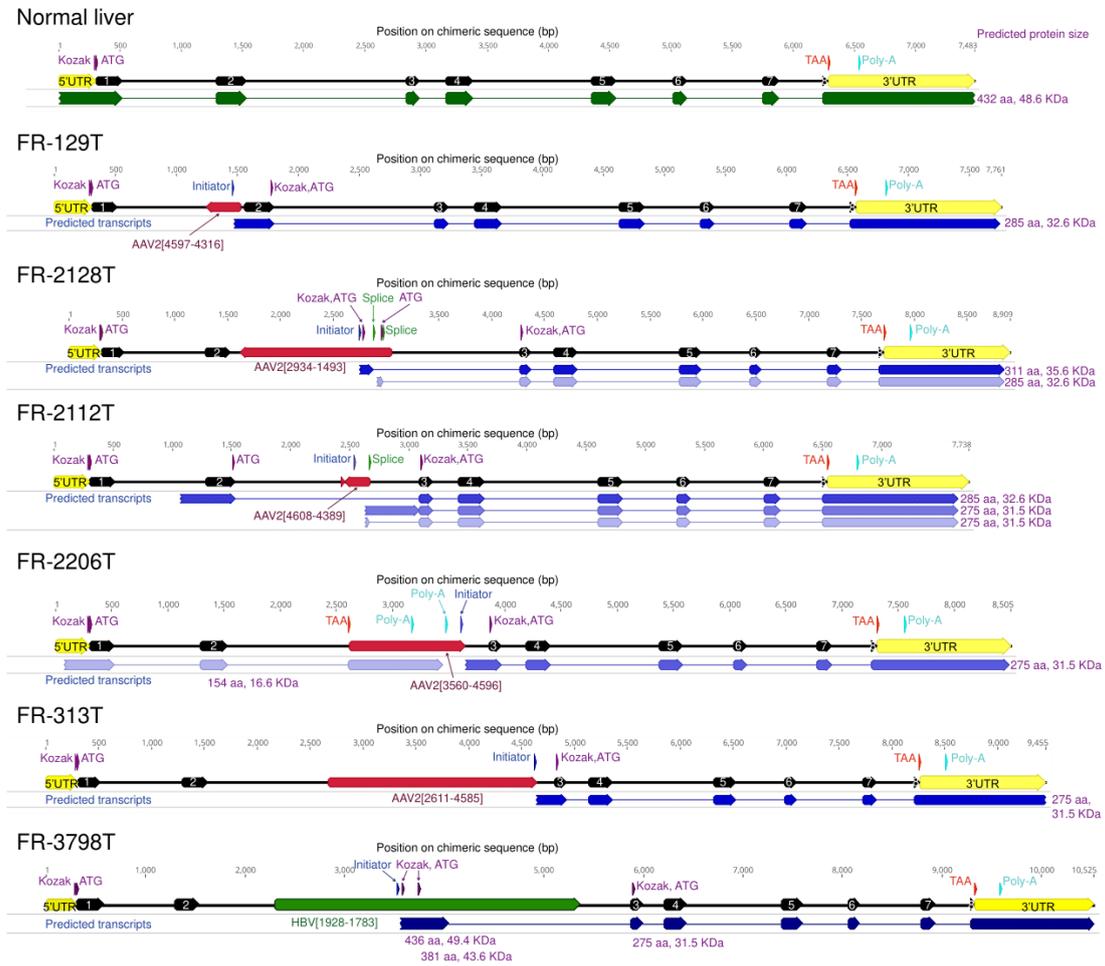
Supplementary Figure 8. RS1 breakpoint hotspots involving highly expressed liver enzymes

Supplementary Figure 9. Binomial regression modeling of rearrangement breakpoint density

Supplementary Figure 10. Rearrangements affecting *TERT* promoter region in CCN-HCC

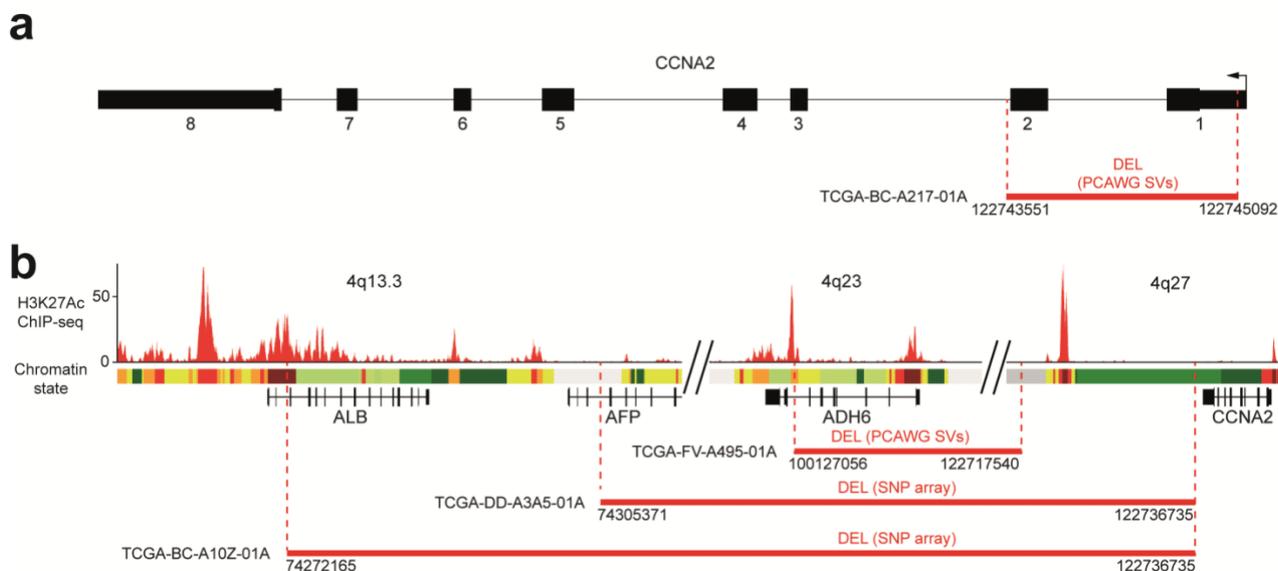
Supplementary Figure 11. Modulation of *TERT* expression by different types of genomic alterations.

Supplementary Figure 12. Rearrangement signatures identified in the pan-cancer series



Supplementary Figure 1. Consequences of viral insertions in *CCNA2* gene

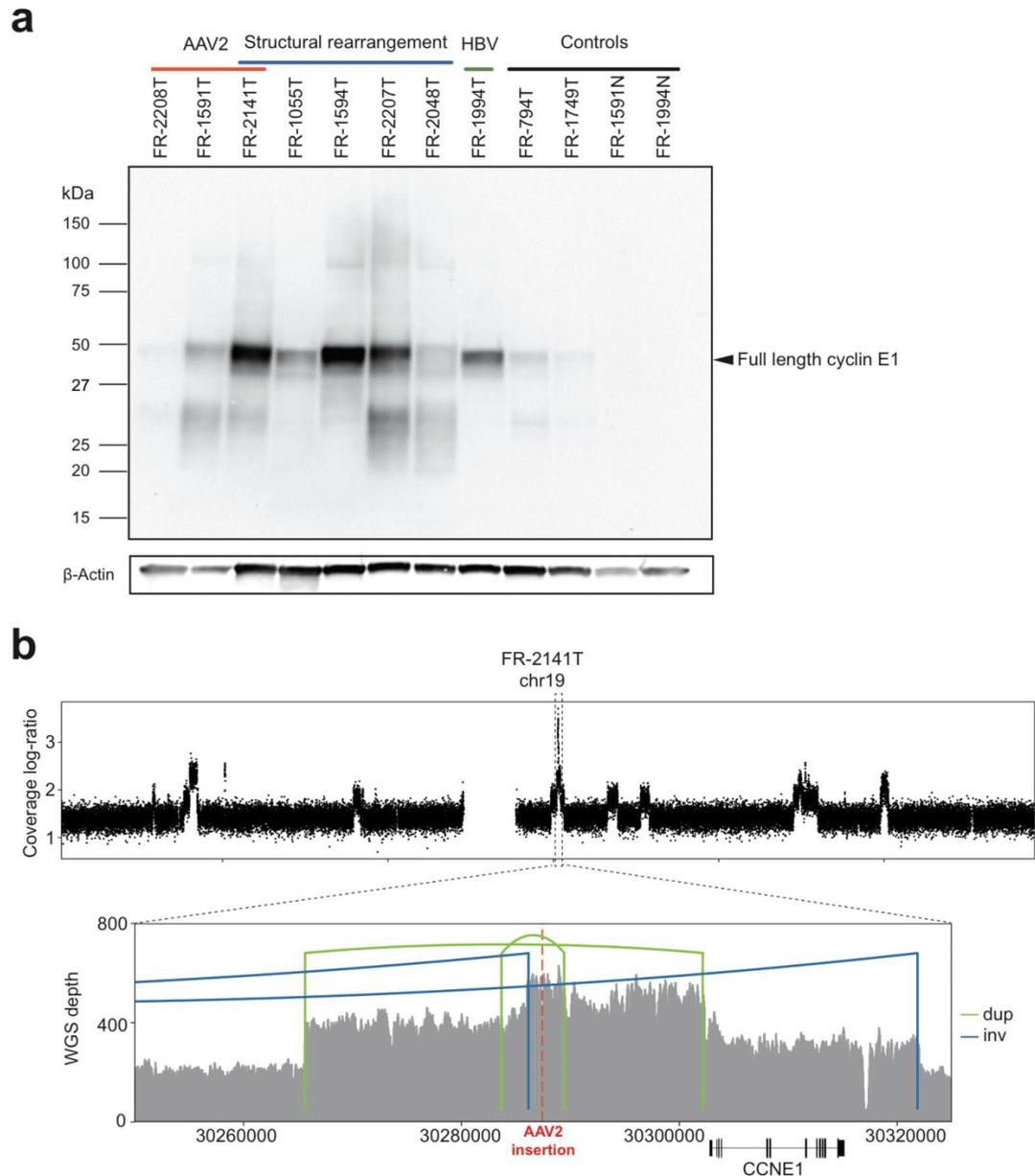
Five AAV2 and 1 HBV insertions were identified in *CCNA2* in the LICA-FR series. Precise insertion boundaries were identified by WGS or viral capture, and RNA-seq reads were aligned on the chimeric sequence. Here, the different transcripts predicted by Cufflinks are represented for each case, ordered by transcript abundance. Predicted functional elements (Transcription initiator, splice and poly-A sites, Kozak sequence, initiator and terminator codons) are annotated on the chimeric DNA sequence, and the predicted protein sizes resulting from the translation of each transcript are annotated on the right. Only the most abundant abnormal transcripts were represented in **Fig. 1c**.



Supplementary Figure 2. Deletions associated with *CCNA2* deregulation

a Focal deletion of *CCNA2* exons 1 & 2 identified in TCGA tumor TCGA-BC-A217.

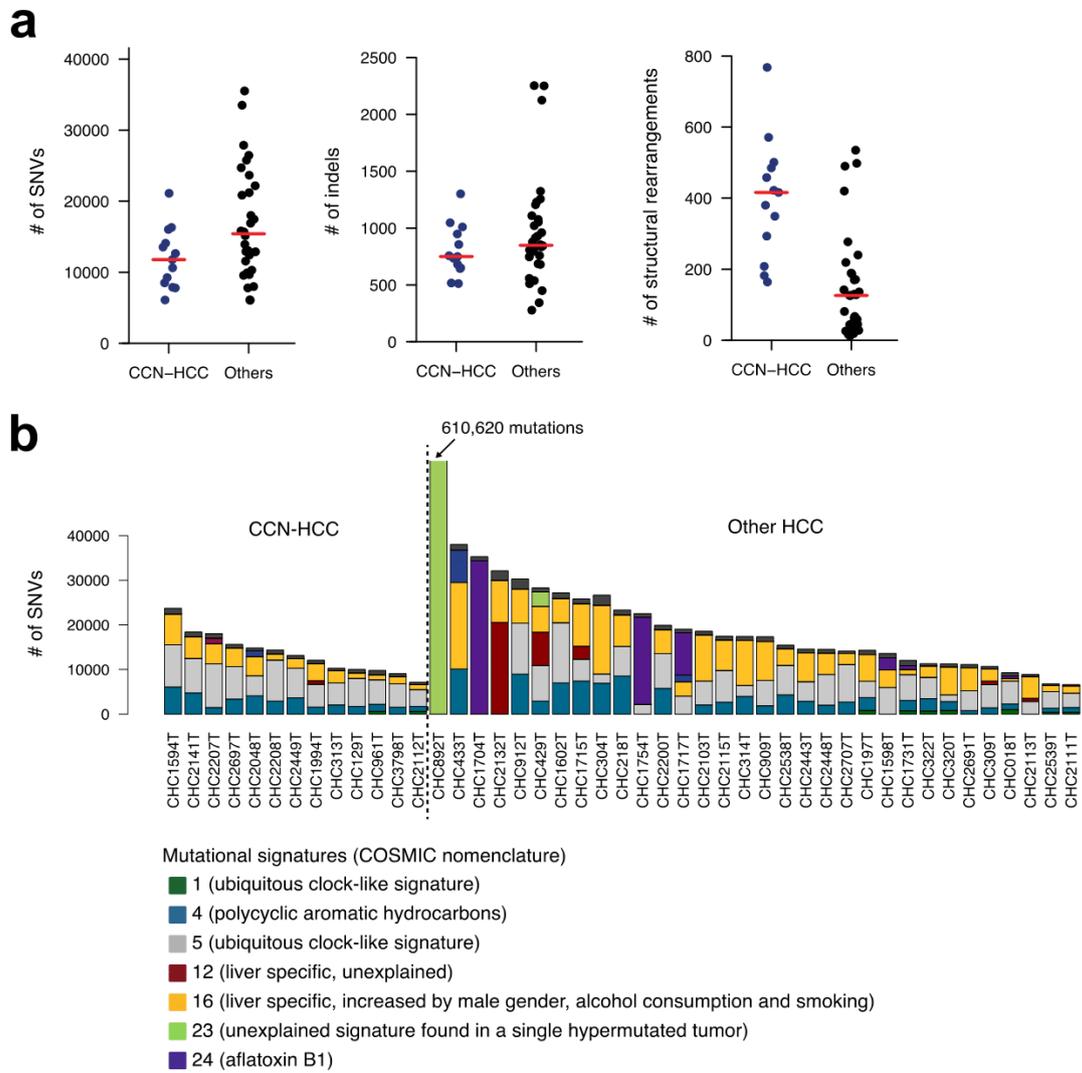
b Deletions identified in 3 TCGA tumors linking *CCNA2* downstream region with the highly expressed genes *ALB*, *AFP* and *ADH6*.



Supplementary Figure 3. Cyclin E1 overexpression induced by viral insertions and structural rearrangements

a Western blot analysis of cyclin E1. Tumors with viral insertions or structural rearrangements are compared with tumors without *CCNE1* alteration and non-tumoral liver controls.

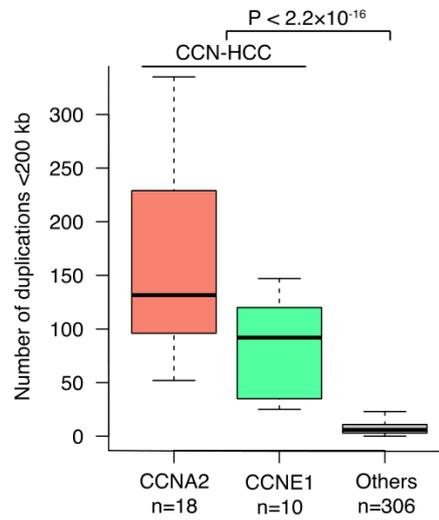
b Tumor FR-2141T displays both an AAV2 insertion in *CCNE1* regulatory region and a high-level amplification of the locus. The top panel displays the coverage log-ratio along chromosome 19 in this tumor. The bottom panel displays the coverage of WGS reads aligned to the chimeric sequence of *CCNE1* locus including AAV2 insertion together with structural rearrangement breakpoints. It shows that the most strongly amplified region includes *CCNE1* regulatory region, and in particular the locus of AAV2 insertion.



Supplementary Figure 4. Mutational signature analysis of CCN-HCC

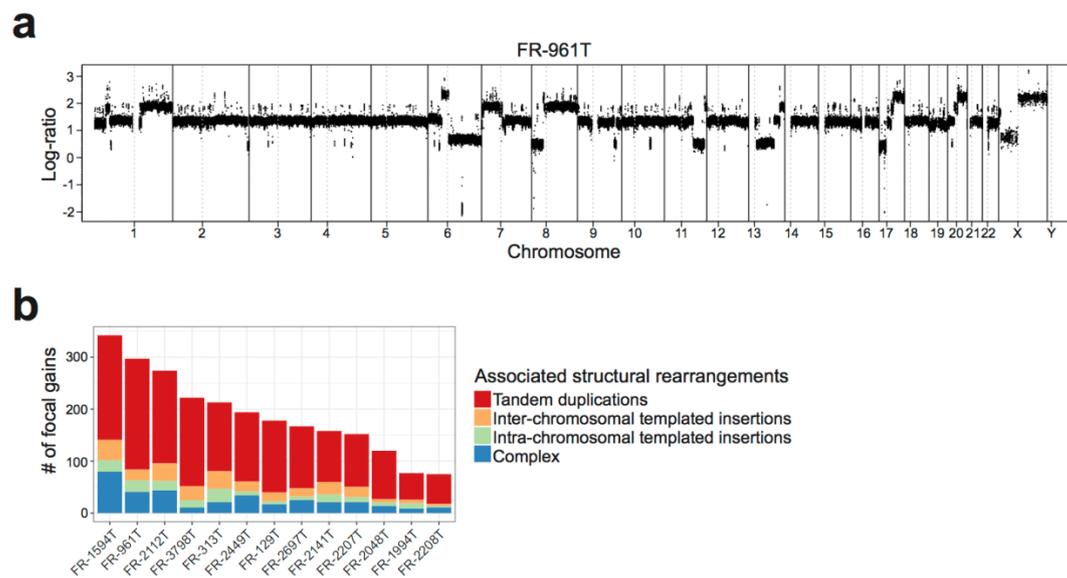
a Comparison of the number of single-nucleotide variants (SNVs), indels and structural rearrangements in CCN-HCC vs others (LICA-FR data, 45 WGS). Red line represent median.

b Contribution of the different mutational signatures known to be operative in liver cancers to the mutational burden of CCN-HCC and other HCC.



Supplementary Figure 5. SNP array analysis of focal duplications in CCN-HCC from the TCGA series.

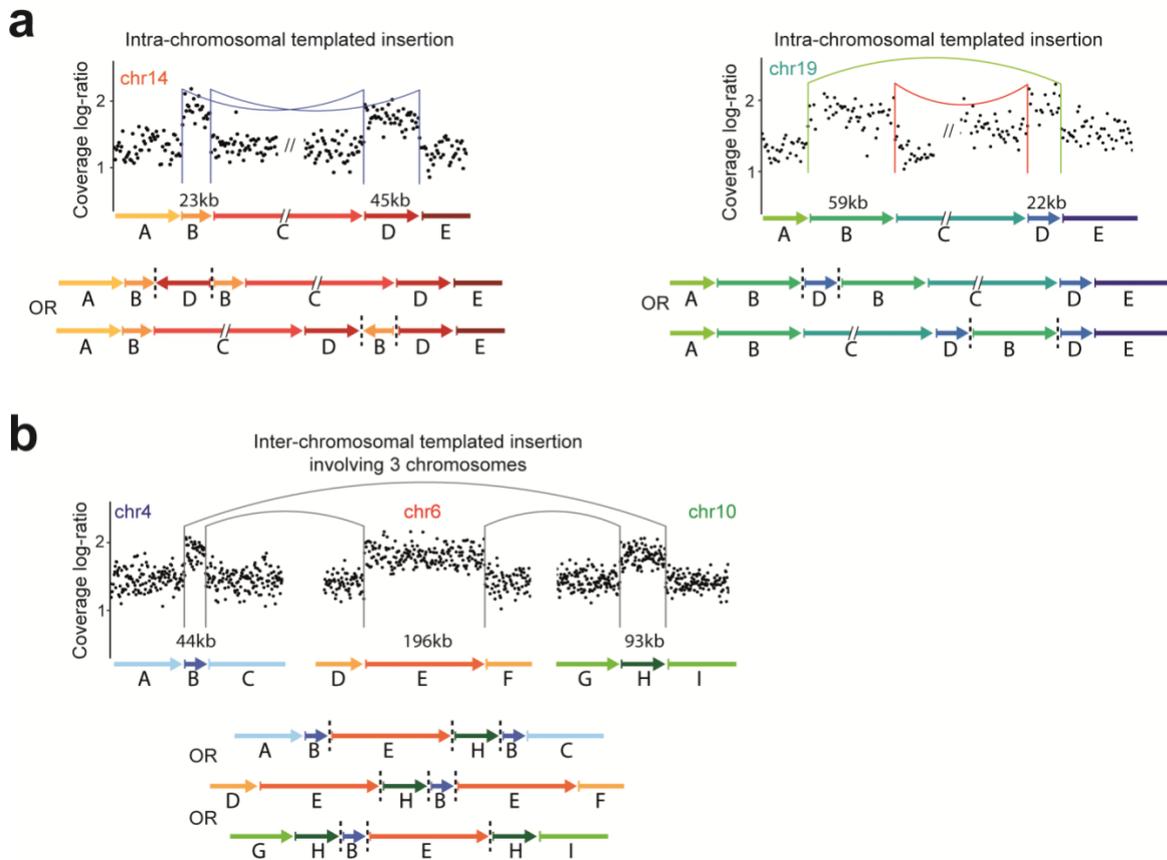
Focal deletions (<200 kb) were quantified across 334 tumors from the TCGA series as a surrogate marker of the RS1 signature, and compared between CCNA2-activated, CCNE1-activated and other HCC. The middle bar, median; box, interquartile range; bars extend to 1.5 times the interquartile range.



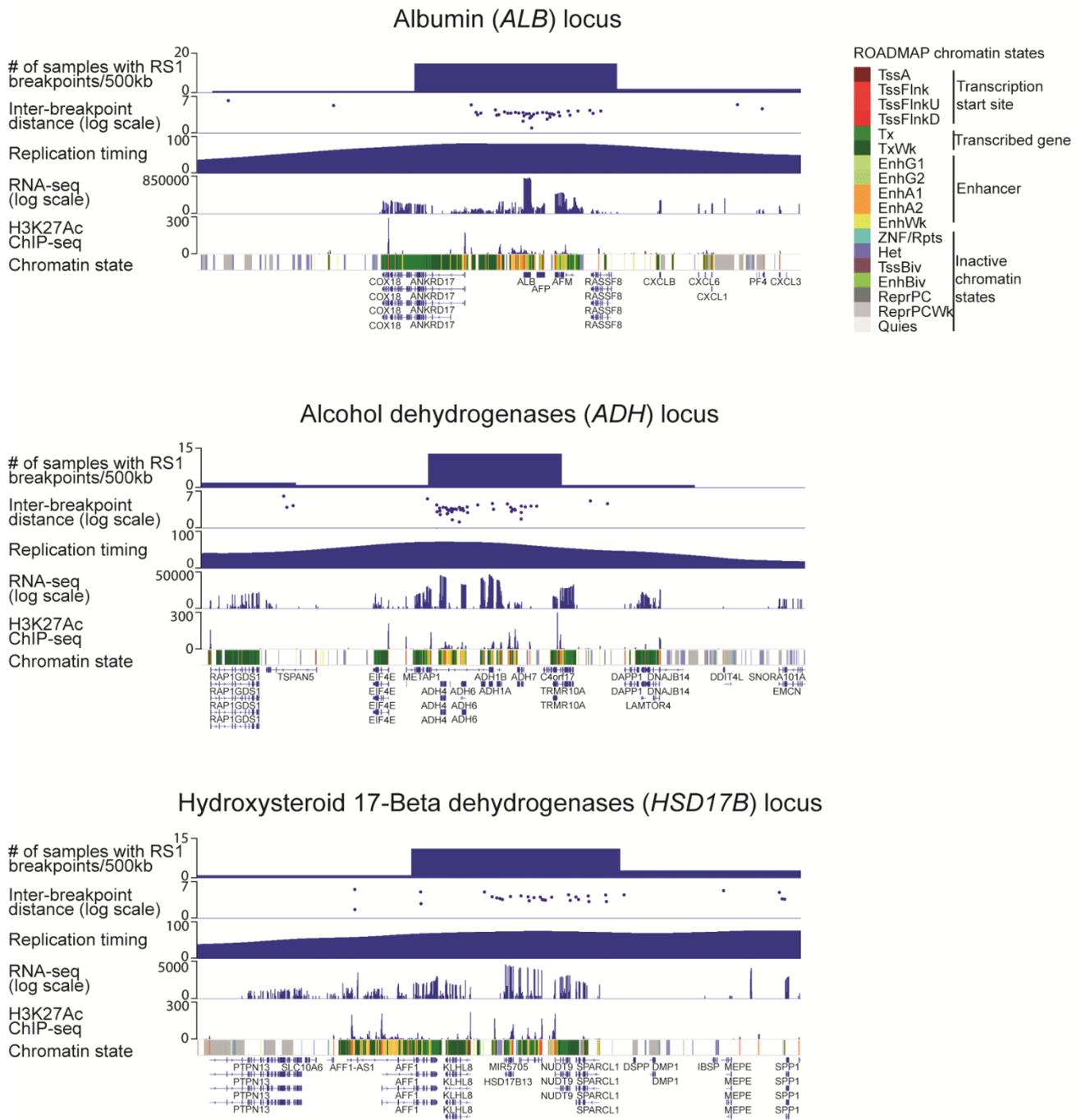
Supplementary Figure 6. Characteristic copy-number profile of CCN-HCC

a Typical copy-number profile of a CCN-HCC, showing hundreds of focal gains scattered throughout the genome.

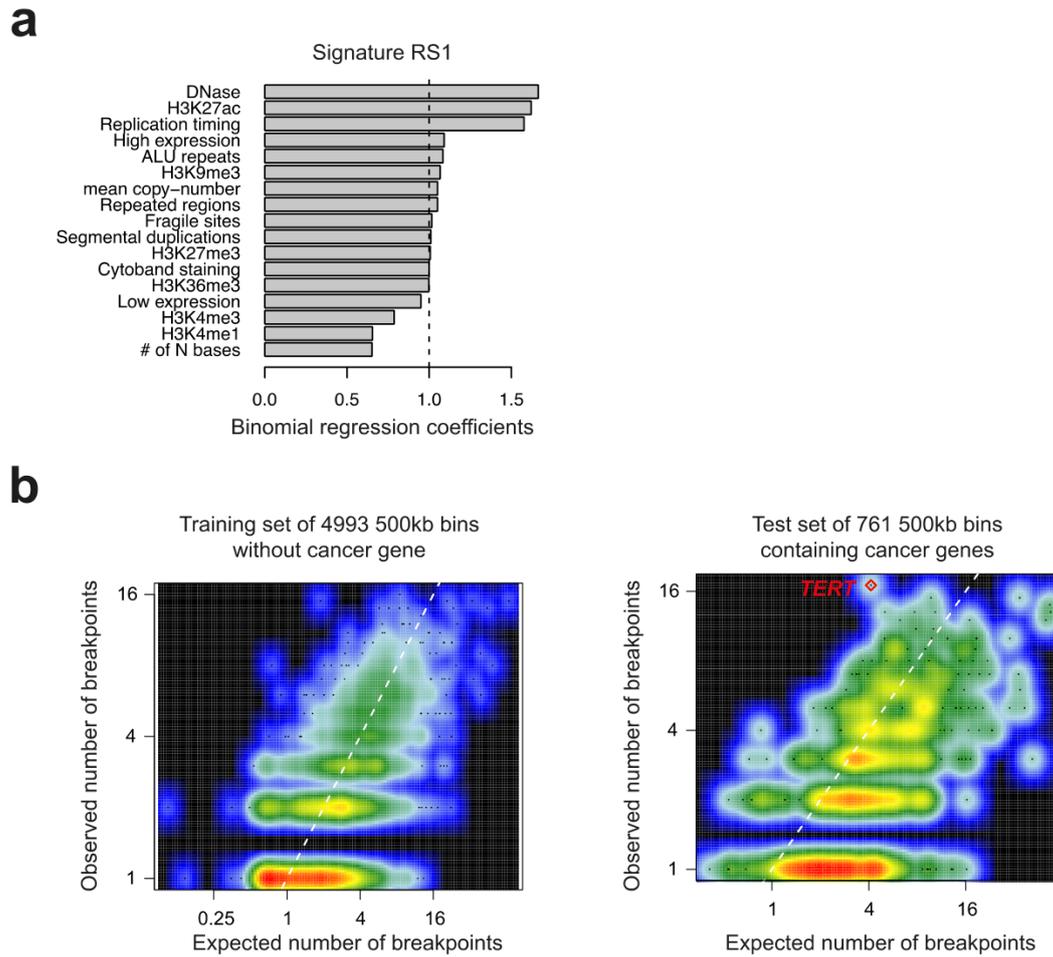
b Proportion of focal gains in each CCN-HCC attributed to each rearrangement mechanism.



Supplementary Figure 7. Examples of intra-chromosomal templated insertions and inter-chromosomal templated insertion cycles involving several chromosomes
a Intra-chromosomal templated insertions appear as couples of inversions (left) or deletion and duplication (right) depending on the orientation of aberrant junctions.
b Example of inter-chromosomal templated insertion involving 3 different chromosomes.

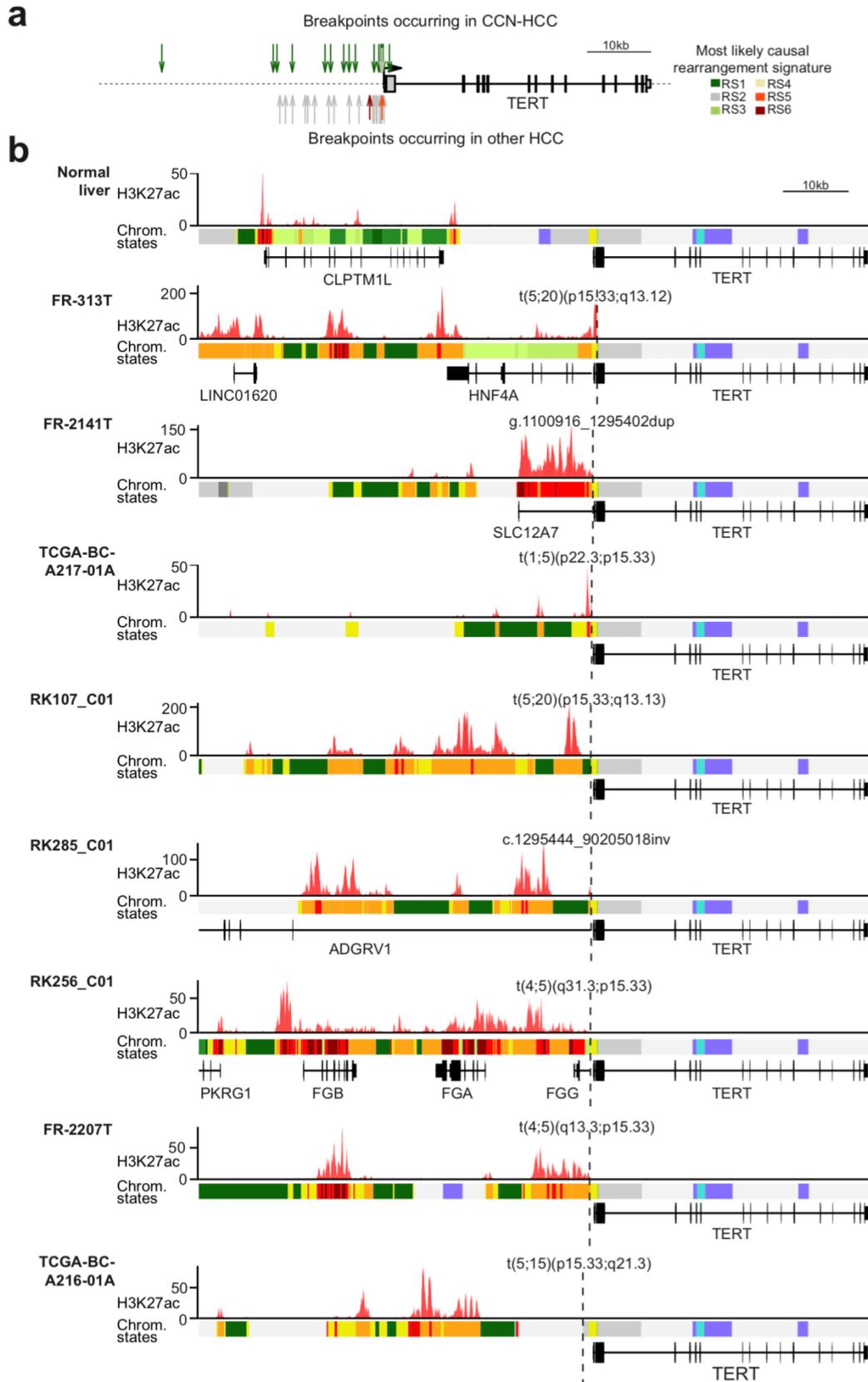


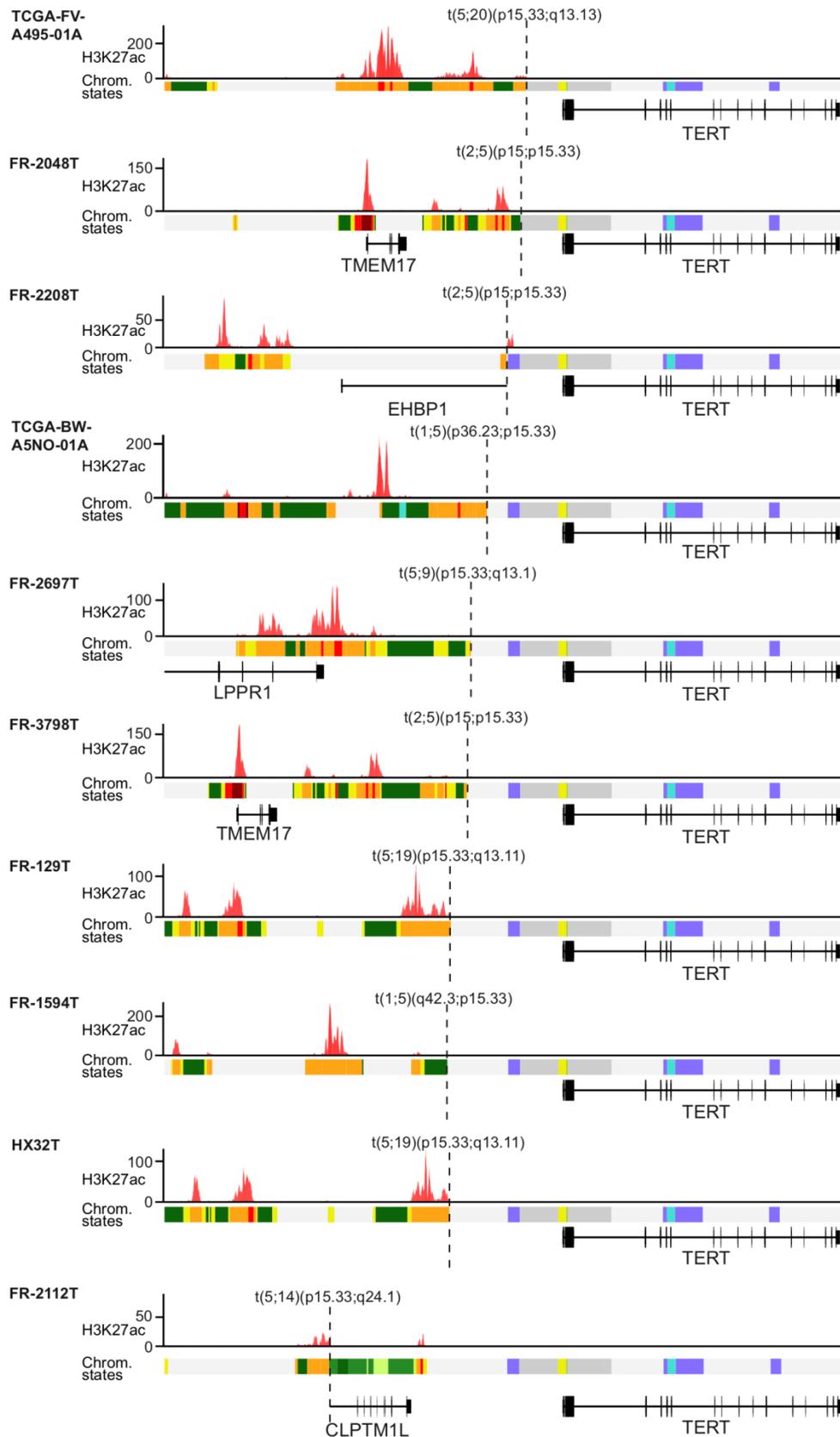
Supplementary Figure 8. RS1 breakpoint hotspots involving highly expressed liver enzymes
 The density of RS1 breakpoints, replication timing, RNA-seq expression, H3K27Ac and ROADMAP chromatin states are displayed for 3 representative hotspots involving the very highly expressed liver enzymes albumin, alcohol dehydrogenases and hydroxysteroid 17-Beta dehydrogenases.



Supplementary Figure 9. Binomial regression modeling of rearrangement breakpoint density
a Regression coefficients of the 17 genomic features used to predict the density of signature RS1 breakpoints.

b Correlation between the number of observed RS1 breakpoints per 500 kb bin and the expected number predicted by the binomial regression model. Left: Within 4933 bins without any cancer gene used as training set. Right: Within 761 bins containing cancer genes (test set). The bin corresponding to *TERT* promoter region is highlighted in red.

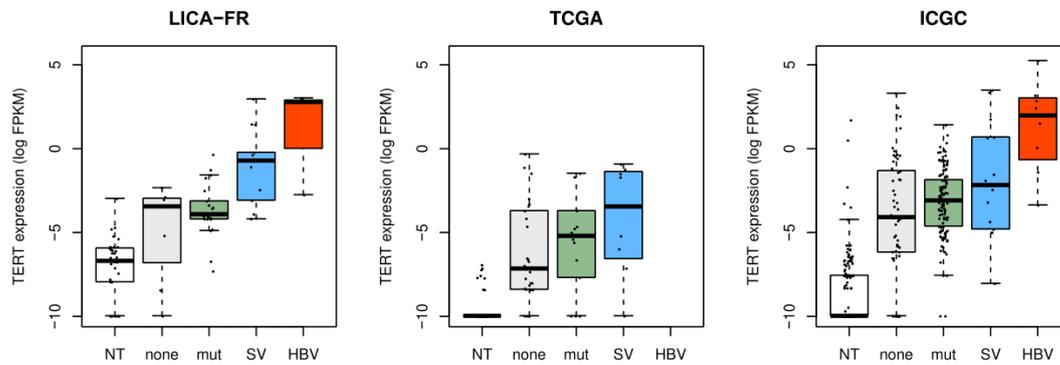




Supplementary Figure 10. Rearrangements affecting *TERT* promoter region in 350 HCC genomes

a Summary of rearrangement breakpoints in *TERT* regulatory region. Each arrow indicates a rearrangement breakpoint. The color indicates the rearrangement signature of the most likely causal process. Breakpoints occurring in CCN-HCC are represented above the scheme. Those occurring in other HCC are represented below.

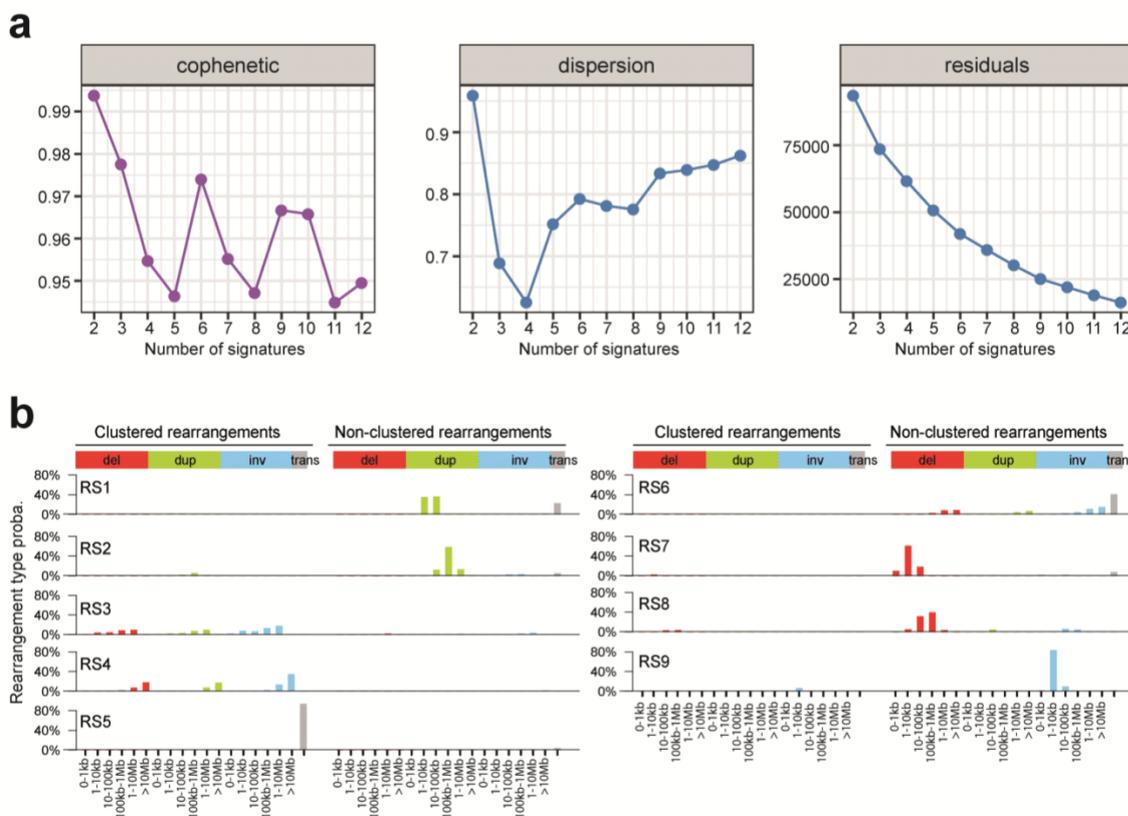
b Functional consequences of structural rearrangements affecting *TERT* regulatory region in CCN-HCC. Chromatin states and H3K27 acetylation in normal adult liver (top) are compared with predicted chromatin states and H3K27 acetylation resulting from the 18 structural rearrangements of *TERT* regulatory regions identified in CCN-HCC. H3K27Ac chromatin immunoprecipitation sequencing (ChIP-seq) signal were obtained from the ROADMAP consortium. The color code of chromatin states is the same as in **Supplementary Fig. 8**



Supplementary Figure 11. Modulation of *TERT* expression by different types of genomic alterations.

The expression of *TERT* is shown in the 3 data sets as a function of the type of alteration identified in *TERT* promoter. Only samples with both whole genome and RNA sequencing were considered. No HBV insertion in *TERT* was identified among the 48 TCGA samples with whole genome sequencing.

NT: non-tumor liver sample; none: HCC without *TERT* alteration identified; mut: HCC with *TERT* promoter mutation; SV: HCC with structural rearrangement affecting *TERT* promoter; HBV: HCC with HBV insertion in *TERT* promoter. The middle bar, median; box, interquartile range; bars extend to 1.5 times the interquartile range.



Supplementary Figure 12. Rearrangement signatures identified in the pan-cancer ICGC series

a Non-negative matrix factorization (NMF) metrics used to determine the optimal number of signatures. With 9 signatures, we obtain good cophenetic coefficient and dispersion score.

b Frequency of the 38 structural rearrangement categories in the 9 signatures.

Description of Additional Supplementary Files
(www.nature.com/articles/s41467-018-07552-9#additional-information)

File Name: Supplementary Data 1

Description: Clinical annotations for the 160 samples of the LICA-FR series

File Name: Supplementary Data 2

Description: Main clinical characteristics of the LICA-FR, TCGA and ICGC-JP cohorts

File Name: Supplementary Data 3

Description: Viral insertions identified at CCNA2 and CCNE1 loci

File Name: Supplementary Data 4

Description: Structural rearrangements identified at CCNA2 and CCNE1 loci

File Name: Supplementary Data 5

Description: Significantly de-regulated pathways in CCN-HCC

File name: Supplementary Data 6

Description: Significantly enriched and depleted driver genes in CCN-HCC

File Name: Supplementary Data 7

Description: Contribution of rearrangement signatures to the genomes of 350 tumors from the LICA-FR, TCGA and ICGC serie

File Name: Supplementary Data 8

Description: Rearrangements affecting TERT promoter region in 350 HCC genomes

File Name: Supplementary Data 9

Description: Number of RS1 events and association with CCNA2/E1 alterations across cancer types

III. Histoire naturelle et potentiel oncogénique de l'adénovirus-associé 2 (AAV2)

L'adénovirus-associé est un virus à ADN simple-brin de petite taille possédant une capsid non-enveloppée (Gonçalves, 2005). C'est un virus défectif car il nécessite la présence d'un autre virus pour se répliquer (un adénovirus ou un herpès simplex virus) et établit dans le cas contraire une infection latente (Schultz and Chamberlain, 2008; Smith, 2008). Ces virus AAV sont transmis par voie respiratoire et présents chez 30 à 80% de la population générale. Douze sérotypes distincts ont été décrits et AAV2 est le plus fréquent chez l'humain (Boutin et al., 2010). Ils sont considérés comme non-pathogéniques pour l'homme et la transfection de vecteurs AAV recombinants ayant une grande efficacité et une faible immunogénicité, ces virus sont largement utilisés en thérapie génique (Kotterman and Schaffer, 2014).

Des intégrations clonales récurrentes d'AAV2 ont cependant été identifiées dans 5% d'une série de près de 200 CHC du laboratoire en 2015. Les tumeurs associées à ces événements s'étant développées dans un contexte non-cirrhotique et en absence de facteur étiologique connu, l'intégration virale semble être chez ces patients directement à l'origine du développement tumoral via un mécanisme de mutagénèse insertionnelle (Nault et al., 2015). Bien qu'il ait été montré que le foie est le site principal d'infection des virus AAV (Gao et al., 2004), peu d'information était alors disponible sur l'histoire naturelle de l'infection et sur le rôle précis joué par l'intégration virale dans la transformation maligne.

Ainsi nous avons voulu étudier de façon approfondie l'infection par AAV2 en criblant par PCR quantitative une série de tissus hépatiques tumoraux et non-tumoraux du laboratoire provenant de 1319 patients afin de détecter la présence d'ADN viral dans le foie. La cohorte constituée était composée à la fois de tumeurs malignes (principalement des CHC) et de tumeurs bénignes (principalement des adénomes hépatocellulaires). Les tissus fortement positifs ont ensuite été séquencés de façon ciblée par capture virale pour identifier précisément les formes et séquences virales.

De l'ADN viral d'AAV a été détecté chez 18% des patients, plus fréquemment dans les tissus non-tumoraux que dans les tumeurs, et principalement chez des patients jeunes de sexe féminin n'ayant pas de cirrhose. La reconstruction des génomes entiers du

virus dans les tissus hépatiques de 57 patients ayant des tumeurs malignes ou bénignes ont permis de mettre en évidence la présence de deux génotypes viraux : le type AAV2 déjà décrit et une séquence recombinée des types AAV2/AAV13.

Afin de caractériser les formes virales, nous avons mis au point une expérience de PCR quantitative utilisant une DNase digérant l'ADN sous forme linéaire pour décrire la présence de formes épisomales dans 26% des tissus non-tumoraux contenant de l'ADN d'AAV, qui était fortement associée à l'expression d'ARN viraux et suggère donc que le virus est transcriptionnellement actif dans le foie de ces patients. AAV ayant besoin d'un autre virus pour se répliquer, un deuxième criblage des virus décrits dans la littérature a permis d'identifier le virus de l'Herpès Humain 6 (HHV-6) comme principal virus co-infectant les tissus hépatiques permettant l'infection active d'AAV.

Seulement 8% des tumeurs analysées contenaient de l'ADN viral d'AAV, et la proportion était identique entre les tumeurs malignes et les tumeurs bénignes. Cependant, les tumeurs malignes avaient une quantité d'ADN plus importante induite par une fréquence plus élevée d'intégrations virales clonales dans ces tissus. Au total, 19 intégrations clonales ont pu être identifiées dans des tumeurs (dont 11 avaient déjà été décrites dans l'étude de 2015) et presque aucune forme épisomale du virus. Les gènes altérés par la mutagenèse insertionnelle d'AAV étaient *CCNA2*, *CCNE2*, *TERT*, *TNFSF10*, *KMT2B* et *GLI1/INHBE*. La présence dans la séquence virale intégrée de régions « Enhancer », de sites de liaison à des facteurs de transcription et d'un site poly-A, entraîne dans les tissus une surexpression du gène cible et la synthèse de transcrits en excès ou sous une forme tronquée, qui ont des conséquences oncogéniques spécifiques.

Ce projet a donc permis de décrire les formes virales d'AAV présentes dans les tissus hépatiques tumoraux et non-tumoraux afin d'améliorer la compréhension de l'histoire naturelle de l'infection de ce virus chez l'homme. Bien que cela soit un événement rare, l'intégration virale d'AAV à des régions précises du génome humain peut induire ou favoriser la carcinogénèse hépatique dans un contexte non-cirrhotique. Ces résultats ont donc confirmé qu'AAV est le troisième virus associé à la survenue de CHC avec le VHB et le VHC, via un mécanisme de mutagenèse insertionnelle.

Article 3: Adeno-associated virus in the liver: natural history and consequences in tumor development.

Tiziana La Bella*, Sandrine Imbeaud*, **Camille Péneau**, Iadh Mami, Shalini Datta, Quentin Bayard, Stefano Caruso, Theo Z Hirsch, Julien Calderaro, Guillaume Morcrette, Catherine Guettier, Valérie Paradis, Giuliana Amaddeo, Alexis Laurent, Laurent Possenti, Laurence Chiche, Paulette Bioulac-Sage, Jean-Frédéric Blanc, Eric Letouzé, Jean-Charles Nault, Jessica Zucman-Rossi.

Gut, 2020 Apr;69(4):737-747

*Ces auteurs ont contribué de manière égale à ce travail

ORIGINAL RESEARCH

Adeno-associated virus in the liver: natural history and consequences in tumour development

Tiziana La Bella,^{1,2} Sandrine Imbeaud ,^{1,2} Camille Peneau,^{1,2} Iadh Mami,^{1,2} Shalini Datta,^{1,2,3} Quentin Bayard,^{1,2} Stefano Caruso,^{1,2} Theo Z Hirsch,^{1,2} Julien Calderaro ,^{1,2,4} Guillaume Morcrette,^{1,2,5,6} Catherine Guettier,^{5,6} Valerie Paradis,^{2,7,8} Giuliana Amaddeo,^{9,10} Alexis Laurent,¹¹ Laurent Possenti,¹² Laurence Chiche,¹³ Paulette Bioulac-Sage,^{14,15} Jean-Frederic Blanc,^{12,14,15} Eric Letouze,^{1,2} Jean-Charles Nault ,^{1,2,16} Jessica Zucman-Rossi ,^{1,2,17}

► Additional material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2019-318281>)

For numbered affiliations see end of article.

Correspondence to

Professor Jessica Zucman-Rossi; jessica.zucman@gmail.com, jessica.zucman@gmail.com

TLB and SJ contributed equally.

Received 14 January 2019
Revised 28 June 2019
Accepted 29 June 2019
Published Online First
2 August 2019



© Author(s) (or their employer(s)) 2020. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: La Bella T, Imbeaud S, Peneau C, *et al.* *Gut* 2020;**69**:737–747.

ABSTRACT

Objective Adeno-associated virus (AAV) is a defective mono-stranded DNA virus, endemic in human population (35%–80%). Recurrent clonal AAV2 insertions are associated with the pathogenesis of rare human hepatocellular carcinoma (HCC) developed on normal liver. This study aimed to characterise the natural history of AAV infection in the liver and its consequence in tumour development.

Design Viral DNA was quantified in tumour and non-tumour liver tissues of 1461 patients. Presence of episomal form and viral mRNA expression were analysed using a DNase/TaqMan-based assay and quantitative RT-PCR. In silico analyses using viral capture data explored viral variants and new clonal insertions.

Results AAV DNA was detected in 21% of the patients, including 8% of the tumour tissues, equally distributed in two major viral subtypes: one similar to AAV2, the other hybrid between AAV2 and AAV13 sequences. Episomal viral forms were found in 4% of the non-tumour tissues, frequently associated with viral RNA expression and human herpesvirus type 6, the candidate natural AAV helper virus. In 30 HCC, clonal AAV insertions were recurrently identified in *CCNA2*, *CCNE1*, *TERT*, *TNFSF10*, *KMT2B* and *GLI1/INHBE*. AAV insertion triggered oncogenic overexpression through multiple mechanisms that differ according to the localisation of the integration site.

Conclusion We provided an integrated analysis of the wild-type AAV infection in the liver with the identification of viral genotypes, molecular forms, helper virus relationship and viral integrations. Clonal AAV insertions were positive selected during HCC development on non-cirrhotic liver challenging the notion of AAV as a non-pathogenic virus.

INTRODUCTION

Adeno-associated virus (AAV) is a small non-enveloped DNA virus with an icosahedral capsid that contains a 4.7 kb linear single-stranded genome.^{1,2} AAV genome codes for non-structural proteins (Rep78, 68, 52 and 40), capsid proteins (VP1, VP2, VP3) and the assembly activating protein (AAP).^{3,4} At the extremities, inverse tandem repeats (ITR) are important for the integration in host genome.^{5,6} AAV is a defective

Significance of this study

What is already known on this subject?

- The seroprevalence of adeno-associated virus (AAV) in general population is 40%–80% and AAV2 is the most frequent serotype in human.
- AAV has a biphasic life cycle characterised by latent and lytic phases.
- The presence of a helper virus is required for the AAV replication.
- It is commonly believed that adenovirus is the natural AAV helper virus.
- Although AAV is considered a non-pathogenic virus, recurrent clonal AAV2 insertions were associated with hepatocellular carcinoma (HCC) development.

What are the new findings?

- Two viral subtypes are present in 21% of the liver tissues: AAV2 and hybrid AAV2/13 sequences.
- Episomal AAV forms are found in 4% of non-tumour liver tissues, mainly in young, female patients without liver fibrosis.
- Human herpesvirus type 6 is the most frequent AAV helper virus in the liver.
- The 2% of patients with HCC displayed clonal AAV integration in cancer driver genes.
- AAV clonal insertion in HCC activates oncogenes using various mechanisms.

virus that requires a helper virus for an active infection, otherwise it can establish a latent infection through integration into host genome or maintenance as circular episomal form.^{7–9} AAV seroprevalence showed that the infection is endemic in human populations (30%–80%) starting during childhood.^{10–12} Twelve distinct serotypes and more than 100 natural variants have been identified, among which AAV2 is the most frequent type in human.^{13–16}

This small virus is attractive for gene therapy because of the lack of identifiable associated disease and the remarkable ability of recombinant AAV (rAAV) vectors to transduce dividing and non-dividing cells with high efficiency, long-term

Hepatology

Significance of this study

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

- ▶ These findings are important to understand wild-type AAV biology and its association with hepatocarcinogenesis.
- ▶ Our data are particularly relevant considering the large usage of AAV vector in liver-targeted gene therapy.
- ▶ Even if rare, AAV insertional mutagenesis is a new risk factor of HCC development, therefore the notion of AAV as non-pathogenic virus should be reviewed.

transgene expression, low immunogenicity and specific tissue tropism.¹⁷ Although AAV was discovered in 1965, many questions regarding the natural history of AAV infection in human remain unanswered.² It is well known that the vector predominantly persists in the nucleus as episomal form with sustained RNA expression raising question on putative episomal AAV form in wild-type infection.⁸ Several helper viruses have been identified but their precise association with wild-type liver AAV infection remains unclear. The frequency of the different AAV genotypes in the human population and AAV persistence in tissues after first infection remains to be determined.¹⁸ Moreover, AAV link with tumour development is controversial, with some studies reporting an oncogenic effect of AAV infection in animal model and others suggesting a tumour suppressive role.^{19–24}

Recently, we reported the involvement of AAV2 in the pathogenesis of human hepatocellular carcinoma (HCC) developed on normal liver in the absence of classical HCC risk factors such as infection with HBV and HCV, high alcohol intake, haemochromatosis or aflatoxin B1 exposure.²⁵ Similar to HBV, recurrent AAV2 clonal insertions were described in *TERT*, *CCNE1* and *CCNA2* cancer driver genes, leading to their overexpression.^{25–28} The AAV insertions can activate oncogenes located nearby in the human genome by a liver promoter recently identified within the minimal common AAV inserted sequence adjacent to the 3'ITR of the virus.²⁹

In this work, we investigated the natural history of wild-type AAV infection in the liver and its consequences in tumour development in a large cohort of 1461 patients with benign or malignant liver tumours.

MATERIALS AND METHODS

Patients and tissue samples

A series of 1461 patients was included in the study approved by our local institutional review board (IRB) committees (CCPRB Paris Saint-Louis, 1997 and 2004; Bordeaux 2010-A00498-31, Ile-de-France VII: projects C0-15-003 and PP 16-001). Liver tissues were frozen immediately after surgery in French hospitals. Tumour and non-tumour counterparts were analysed in 1269 patients, only the tumour or non-tumour tissues were investigated for 138 and 54 patients, respectively. The present series included HCC (n=936), hepatocellular adenomas (HCA, n=225), focal nodular hyperplasia (FNH, n=97), hepatoblastoma or transitional tumours (n=87), cholangiocarcinoma (n=46), fibrolamellar carcinoma (n=36) and other tumours (n=34, online supplementary table 1).

Viral DNA screening

Genomic DNA were analysed for the presence of viral DNA by quantitative RT-PCR (qRT-PCR) on Fluidigm 96.96 dynamic

arrays using the BioMark Real-Time PCR system with TaqMan probe sets designed with Primer3Plus software (online supplementary figure 1A and table 2). Results were analysed using the Fluidigm Real-Time PCR Analysis software (V.4.1.3) and reported to a reference gene, HMBS. The quantification was expressed in viral copy number/cell. Copy number/cell values were tested for unimodal and bimodal distribution using normal-mixEM function of *mixtools* package in R.³⁰

Isolation of human AAV using viral capture sequencing

Viral capture of genomic DNA was performed for tumour and matched normal sample, sequence as previously described using 120-mer primers recognising all AAV genotypes 1–13 already described with around 305 probes/genotype.²⁵ Viral reads were mapped to all AAV1 to AAV13 reference sequences using Burrows-Wheeler Aligner (V.0.7.15).³¹ The number of AAV reads correlates with the number of viral copies/cell (online supplementary figure 1B). Read pairs with at least one read aligned on the virus were extracted using samtools (V.1.3),³² and aligned to a custom reference genome including human chromosomes and virus sequences. We calculated the number of reads mapping the AAV/human chimeric and mate regions in each samples by generating a 20k-bin size bed for hg19 genome, which was used for computations with bedtools multicov utility.³³ For each bin, we calculated the mean of coverage in the samples displayed in a pan genomic plot. We used chimeric reads to identify insertion breakpoints at base resolution by mapping sequences on both sides of the junctions. Clonal events were considered when >25 reads overlapped the same locus, putative subclonal insertions when 4–24 overlapping reads were identified. All viral insertions were validated by visual inspection on Integrative Genomics Viewer. Sequences have been deposited in the Genbank database MK231253 to MK231264 and KT258720 to KT258730.

The analysis of full-length human-AAV sequences is detailed in online supplementary materials and methods. Sequences have been deposited in the Genbank database MK139243 to MK139299 and MK163929 to MK163942.

RNAseq

Samples enriched in poly(A)+ RNA were sequenced using Illumina TruSeq Stranded mRNA kit on HiSeq2000 sequencer, yielding approximately 45 million 100 base pair (bp) paired-end reads (IntegraGen, Evry).³⁴ Reads were aligned and chimeric sequences reconstructed with TopHat2³⁵ and Cufflinks V.2.2.1.³⁶ We used ElemeNT³⁷ to predict transcription start sites (TSS), Alamut Visual software (Interactive Biosoftware) to identify splicing signals on the chimeric DNA sequence, ATGpr³⁸ to identify translation initiation sites and Poly(A) Signal Miner to identify PolyA sites.³⁹ Sequences were deposited in EGA database (EGAS00001002879, EGAS00001001284 and EGAS00001003310).

Detection of viral episomal form

A specific DNase/TaqMan-based assay was adapted from protocol by Werle-Lapostolle *et al*⁴⁰ to detect AAV episomal form (detailed procedures in online supplementary materials and methods). Junctions of the circular AAV were amplified using two couples of primers surrounding the ITRs (online supplementary table 2) in 2.5% glycerol and 5% dimethyl sulfoxide. PCR products were sequenced by Sanger after ExoSAP-IT (Applied Biosystem) purification.⁴¹

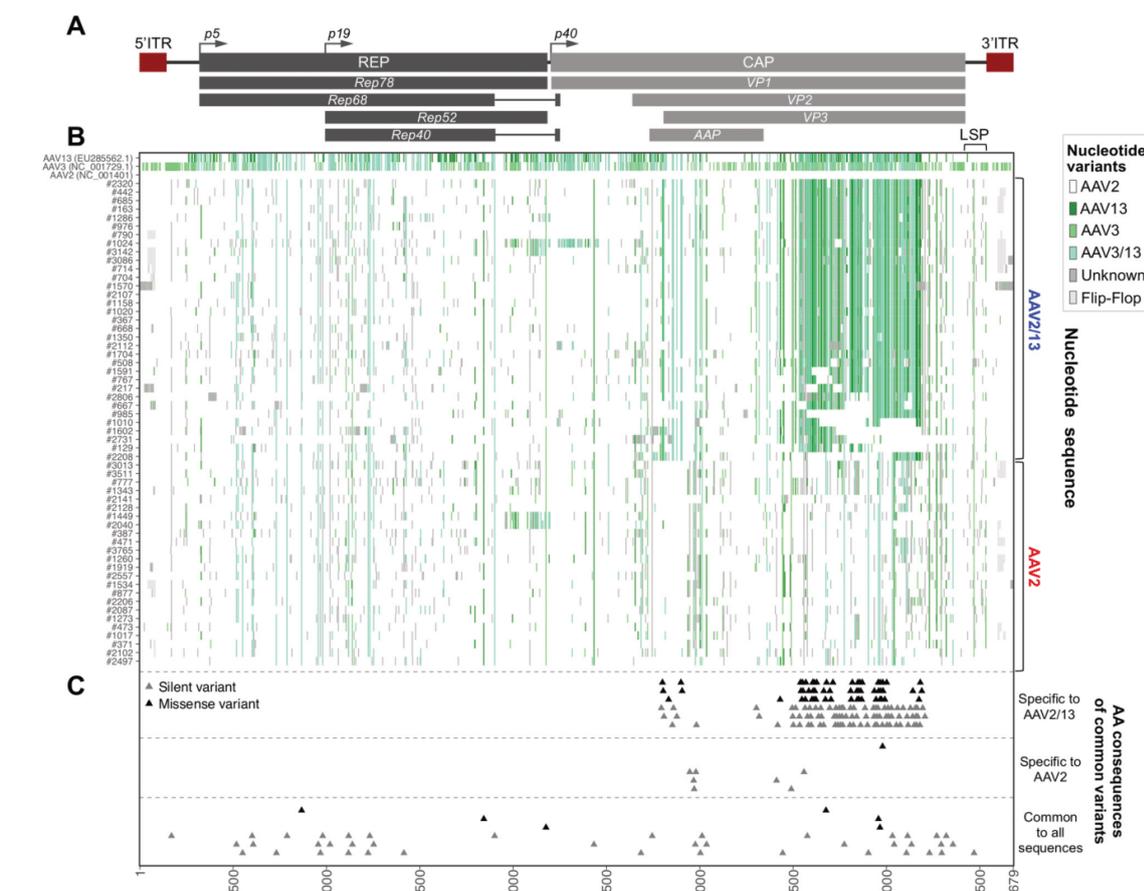


Figure 1 Adeno-associated virus (AAV) full-length sequences in 57 human liver tissues. (A) Schematic representation of AAV genome (reference NC_001401) with location of the two open reading frames encoding replication proteins (Rep78, Rep68, Rep52 and Rep40), structural proteins (VP1, VP2 and VP3) and assembly activating protein (AAP) protein. Inverted terminal repeats (ITR) are represented on the 5' and 3' ends. Promoters (p5, p19 and p40) are indicated with arrows. (B) Nucleotides sequences (4679 bp) from 57 full-length AAV isolated from human liver tissues (ID number indicated with #) multialigned with the ClustalW algorithm compared with reference sequences on the top, AAV2 (NC_001401, in white), AAV3 (NC_001729.1) and AAV13 (EU285562.1). Two distinct viral genotypes, AAV2 and AAV2/13 were identified. Colour bars indicated nucleotide divergence with the AAV2 reference genome similar to AAV3 and/or AAV13 genomes (green) or not (grey), similarities with NC_001401 are in white. Variations due to flip-flop ITR configurations compared with AAV2 reference are labelled in light grey. The liver-specific enhancer-promoter element (LSP) described by Logan *et al* is indicated.²⁹ (C) Amino acid variations compared with the AAV2 reference are indicated. The triangles indicate genome location of specific AAV2/13 (top) or AAV2 (middle) variants in the series of 57 human liver AAV isolates. Common variants shared by both genotypes are shown (bottom). Grey and black colours refer to silent and missense AAV variants, respectively; numbers correspond to wild-type AAV2 nucleotide sequence coordinates (NC_001401).

Quantitative RT-PCR

AAV mRNA and inserted target genes expressions were analysed using qRT-PCR. Specifically, we used seven AAV custom made and human catalogue TaqMan probes (online supplementary table 2) with AB7900HT PCR System (Applied Biosystem) and BioMark Real-Time PCR system. Expression data were normalised with the 2^{-ΔCt} method relative to ribosomal 18S (Hs03928990_g1). Five normal tissues were used as reference.

Site-directed mutagenesis

The role of the viral polyA signal in AAV-induced gene over-expression was investigated in two plasmids containing AAV insertions in the 3'UTR of *TNFSF10*.²⁵ QuikChange Lightning

site-directed mutagenesis kit (Agilent) was used to introduce four point mutations in the viral polyA signal (NC_001401: 4424 A>C, 4426 T>G, 4427 A>C, 4429 A>C). All mutations were verified using Sanger sequencing.

Cell culture, transfection and dual luciferase assay

HuH7, HepG2 and HuH6 cells were purchased from ATCC and cultured in Dulbecco's Modified Eagle Medium supplemented with 10% fetal bovine serum and 100 U/mL penicillin/streptomycin. Cells were tested for mycoplasma contamination. Identity was verified by exome sequencing. Cells were transfected using Lipofectamine 3000 (Life Technologies) with pmirGLO plasmid (Promega) containing wild-type *TNFSF10* 3'UTR, the 3'UTR with AAV2 insertions or scrambled AAV2 sequence downstream

Gut: first published as 10.1136/gutjnl-2019-318281 on 2 August 2019. Downloaded from <http://gut.bmj.com/> on June 15, 2020 at INSERM Consortia. Protected by copyright.

Hepatology

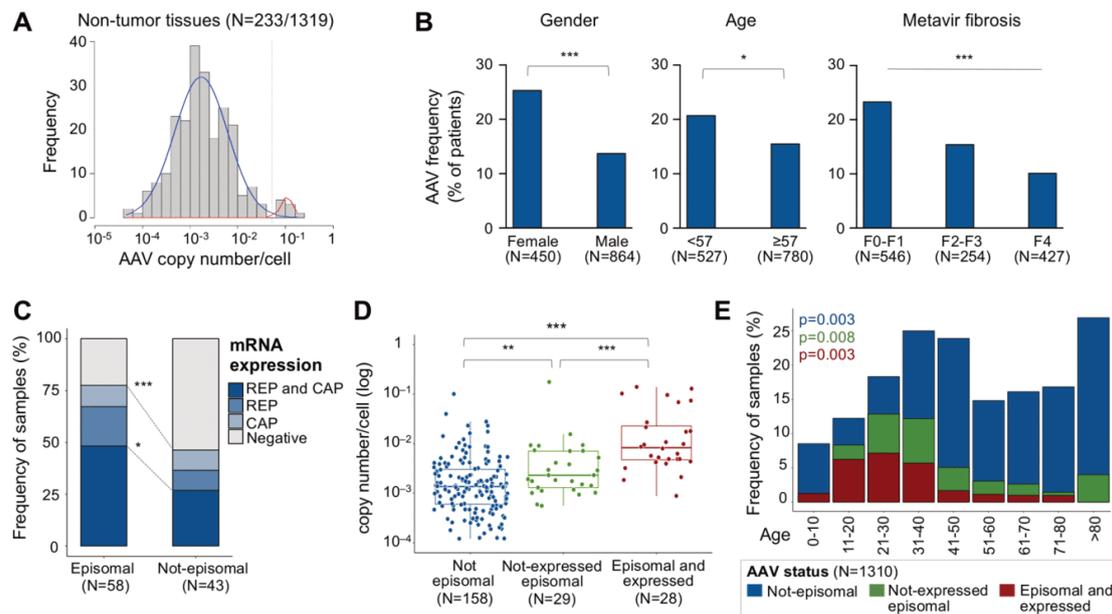


Figure 2 Adeno-associated virus (AAV) DNA in non-tumour tissues and viral episomal form. (A) Copy number/cell distribution in 233 samples. The density line defines the low and high positivity groups in blue and red, respectively. (B) Contingency analysis of AAV positive and negative patients according to gender, age and Metavir fibrosis score. Frequency of AAV-positive patients is displayed (χ^2 test with Monte Carlo simulation and χ^2 test for trend in proportions for Metavir score). (C) Frequency of RNA expression according to REP and CAP viral transcripts in patients with episomal and not-episomal AAV (χ^2 test with Monte Carlo simulation). (D) Viral copy number/cell (\log_{10}) in AAV-positive samples according to the episomal status and the transcriptional activity of the episome (Wilcoxon rank-sum test). (E) Distribution of the different viral molecular forms according to the age of the patients. * $P < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

a luciferase reporter gene. Luminescence from firefly luciferase was normalised on the corresponding renilla luciferase activity. Fold change was calculated relative to the wild-type *TNFSF10* 3'UTR construct.

Statistical analysis

Statistical analyses were performed using RStudio (V.1.0.136) and GraphPad Prism (V.6.0a). Relationship between AAV and clinical, histological features of the patients was investigated using X^2 test. P values adjustment was computed for a Monte Carlo test with 2000 permutations. Statistical significance of quantitative variable was determined by Wilcoxon rank-sum test. Association among variables was modelled by a multinomial logistic regression. Luciferase activity of transfected versus control cells was compared using Student's t-test. All tests were two-tailed and a p value < 0.05 was considered as statistically significant.

RESULTS

Identification of two major AAV genotypes in the liver

Screening of frozen liver tissues from 1319 patients with 6 Taqman probes distributed along the genome that collectively recognise all AAV genotypes 1–13 identified AAV DNA in 18% (n=233) of non-tumour liver tissues (online supplementary figure 1). For viral AAV DNA capture of all known genotypes 1–13, we selected 80 non-tumour liver samples including 68 positive samples ranging from 2×10^{-4} to 0.18 copy number/cell. After sequencing, a full-length AAV sequence was reconstructed in 57 samples and two major AAV subtypes were identified (figure 1A–B). The first subtype (n=25) is highly similar to AAV2 reference sequence (NC_001401) and to VP1 clade B

genotype isolated in human^{14 42} (online supplementary table 2). The second subtype (n=32) showed hybrid sequences including various parts of the AAV13 capsid (similar to clade C^{14 42} and c-ter in the context of an AAV2 5' part, it was named AAV2/13 (figure 1B and online supplementary figure 2). We identified along the viral genome 42 silent variants shared by both AAV subtypes, but different from the AAV2 reference NC_001401 (figure 1C). In contrast, several nucleotide variants leading to amino acid substitutions in AAV2/13 sequences were located in the hypervariable regions (HVRs) 5, 6 7 and 10 and originated from AAV13 sequence (figure 1B–C). Screening the overall series of 1319 samples with two probes specific of AAV2/13 subtype and located in the CAP2 region (online supplementary figure 1), identified 47.6% AAV2 and 52.4% AAV2/13 genotypes among 143 samples positive for the variable region.

AAV infection and episomal form

In the 233 AAV positive liver samples, quantification of the viral DNA showed a bimodal distribution: 97% of the tissues exhibited a low number of copy/cell (ranging from 4.6×10^{-5} to 0.04) and only 8 patients showed a higher quantity of AAV ranging from 0.07 to 0.18 copy/cell (figure 2A). AAV was significantly enriched in female ($p < 0.001$), young patients ($p = 0.016$) and occurred more frequently in a background of non-fibrotic liver ($p < 0.001$; figure 2B).

In 64/233 (27.5%) of the tissues positive for AAV, all the genomic AAV regions were amplified suggesting the presence of the entire viral genome. We designed a DNase/TaqMan-based assay (online supplementary figure 3A), which allowed to detect episomal AAV in 60 patients, corresponding to 26%

Gut: first published as 10.1136/gutjnl-2019-318281 on 2 August 2019. Downloaded from <http://gut.bmj.com/> on June 15, 2020 at INSERM Consortia. Protected by copyright.

of AAV positive samples and 4.6% of all patients. Using in silico analyses of the AAV capture sequencing, among the 57 cases with a complete reconstructed AAV genomic sequence, we identified 14 cases with 3'ITR–5'ITR junctions. Circularised concatemeric structures may escape from our experimental method to identify episomal form,⁴³ however, we did not identify insertion of concatemer in silico. The 3'ITR–5'ITR junctions showed various sequences presenting a double-D ITR structure, in flip or flop configuration, with a 125 bp deletion confirmed by Sanger sequencing (online supplementary figure 3C-D and 4).

AAV transcription is associated with episomal form

Then, we screened for AAV RNA expression in 101 non-tumour liver tissues positive for AAV by qRT-PCR. AAV transcript was identified in 64% of the tested liver tissues. Either AAV REP or CAP expression were enriched in liver tissues with episomal form (p<0.001) and both transcripts were more frequently associated in presence of episomal than not-episomal AAV form (p=0.022), defining a population of patients with an 'episomal-expressed AAV' (figure 2C). A higher AAV copies per cell was identified in liver tissues with episomal-expressed AAV, supporting the hypothesis of a viral active infection in these liver samples (figure 2D). Episomal AAV were also more frequent in female patients (p<0.001) and patients without cirrhosis (p<0.001; online supplementary figure 5A). Analysis

of AAV positivity in function of age showed a peak of frequency at 25% in the 30–40 years class. AAV episomal form was more frequent in young patients (aged <40 years old) reaching the highest frequency level in the twenties (figure 2E and online supplementary figure 5B). These results suggest that AAV active infection is more frequent in the second and third decade during life, while inactive not-episomal forms subsist after the primary infection.

Co-infection with AAV helper viruses

As AAV is a defective virus, we searched for the presence of potential AAV helper viruses by screening the entire cohort of 1319 liver tissues for human adenoviruses (AdV types A–F), human herpesviruses (HHV type 1, 2, 4, 5, 6, 7 and 8) and human papillomavirus type 16 (HPV16) by qRT-PCR. At least one of these viruses was detected in 43% of the patients (n=570), and only one per patient in 39% (n=520). HHV6 was the most frequent (39%), then HHV4 (Epstein-Barr virus, 6%), while HHV7 and adenovirus were only rarely detected (2% and 0.5%, respectively, figure 3A). No HPV16 and HHV type 1 (HSV1), 2 (HSV2), 5 (CMV) and 8 (KSHV) were found in our cohort of liver tissues. HHV6 was the only helper virus enriched in AAV-positive patients (37.3% vs 44.8%, p=0.039), in particular in patients with episomal or expressed-episomal forms (52.5% and 67.9%, respectively, p<0.001; figure 3B).

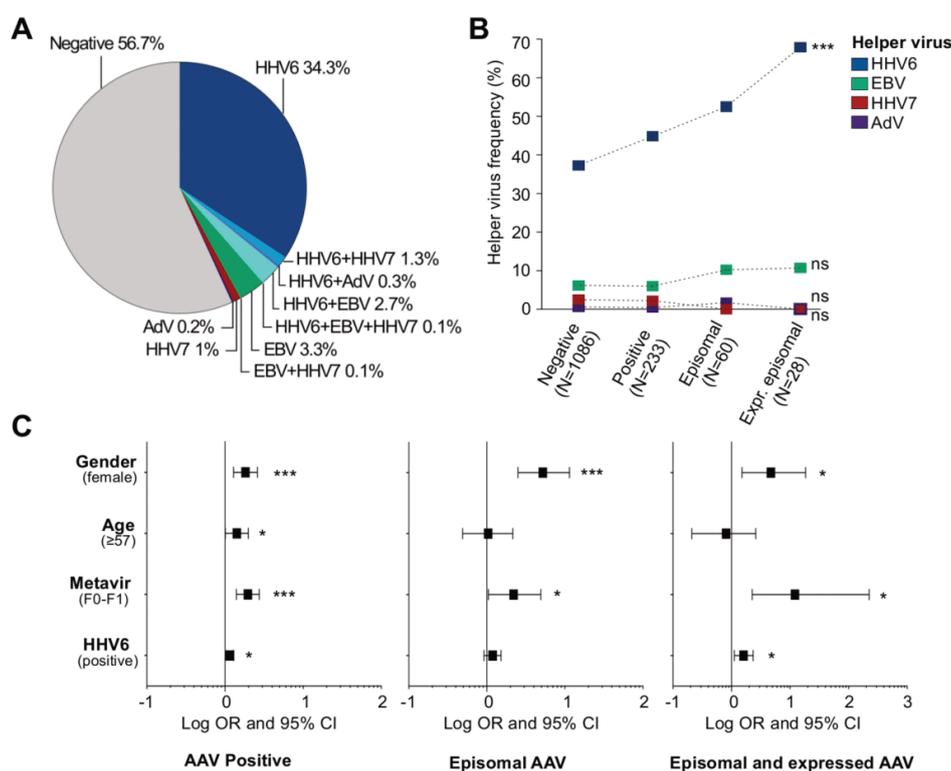


Figure 3 Helper viruses according to adeno-associated virus (AAV) status. (A) Frequency of helper viruses' infections and co-infection in non-tumour tissues (n=1319). (B) Global frequency of human herpesvirus (HHV)6, Epstein-Barr virus (EBV), HHV7 and human adenovirus (AdV) infection according to AAV presence and form (χ^2 test for trend in proportions). (C) Multivariate analysis for global AAV positivity (left) including the variables closely related to AAV presence in the univariate analysis (logistic regression). The same analysis was performed for the presence of episomal AAV (middle) and episomal and expressed form (right). *P<0.05, ***p<0.001. ns, not significant.

Hepatology

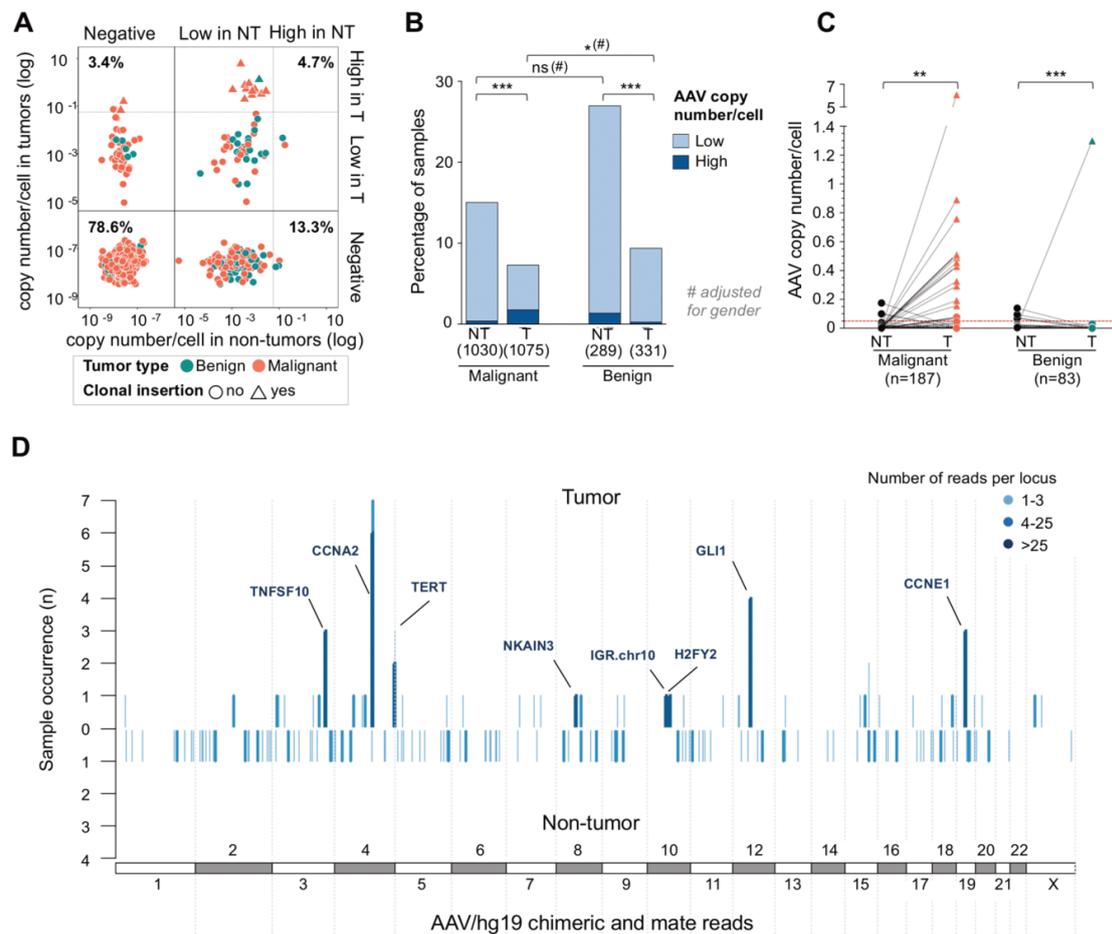


Figure 4 Adeno-associated virus (AAV) in tumour tissues and non-tumour liver counterparts. (A) Copy number/cell (\log_{10}) of paired tumour (T) and non-tumour (NT) tissues of each patient (n=1269). Solid and dashed line define, respectively, the threshold of positivity and the boundaries between high and low number of viral copies/cell. The frequency of patients with AAV in both tumour and non-tumour counterparts or only in one of them is indicated. (B) Frequency of AAV in tumour and non-tumour tissues of patients with malignant and benign tumours (χ^2 test with Monte Carlo simulation and Cochran-Mantel-Haenszel for gender adjustment). (C) AAV copy number/cell of paired tumour and non-tumour tissues of 270 AAV-positive patients grouped in malignant and benign tumour patients. Triangles represent the tumours with clonal AAV insertions (Wilcoxon rank-sum test). (D) Pan-genomic views of genomic location of the human/virus matching chimeric and mate reads in tumour (top) and non-tumour (bottom) samples. A line corresponds to a 20k-bin region, colour refers to the average number of reads counted per bin. The height of the lines corresponds to the frequency of presence of reads in the series of samples, considering 94 tumours and 82 non-tumours investigated with viral capture deepseq. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. ns, not significant.

To identify independent features associated with AAV infection in the overall cohort of patients, we performed a multivariate analysis (figure 3C). Female gender (OR=1.83, $p < 0.001$), the age (OR=1.42, $p = 0.044$), non-cirrhotic liver (OR=1.96, $p < 0.001$) and co-infection with HHV6 (OR=1.15, $p = 0.031$) were independently associated with AAV positivity. Three factors were also significantly associated with the presence of episomal and expressed AAV: female gender (OR=4.71, $p = 0.013$), non-fibrotic liver (OR=12.13, $p = 0.018$) and co-infection with HHV6 (OR=1.61, $p = 0.01$).

AAV in tumour tissues

AAV DNA positivity was less frequently identified in the tumour tissues (n=109, 8%) compared with non-tumour

liver tissues (n=233, 18%) with only 4.7% of patients presenting AAV in both tumour and non-tumour compartments (figure 4A). Twenty out of the 109 positive tumours showed a high number of AAV copies/cell ranging from 0.07 to 6.08. This value might be underestimated considering both potential contamination by normal cells and ploidy of tumour hepatocytes. The vast majority (n=83, 76%) had only one or two amplified viral regions with an enrichment for the 3'ITR region of the virus (online supplementary figure 6A-B). AAV was detected with a similar frequency in malignant and benign tumours, but with a higher number of copies/cell in malignant tumours corresponding to the clonal AAV insertion events (figure 4B-C; online supplementary table 3). Conversely, in all patients with benign tumours except one

Gut: first published as 10.1136/gutjnl-2019-318281 on 2 August 2019. Downloaded from http://gut.bmj.com/ on June 15, 2020 at INSERM Consortia. Protected by copyright.

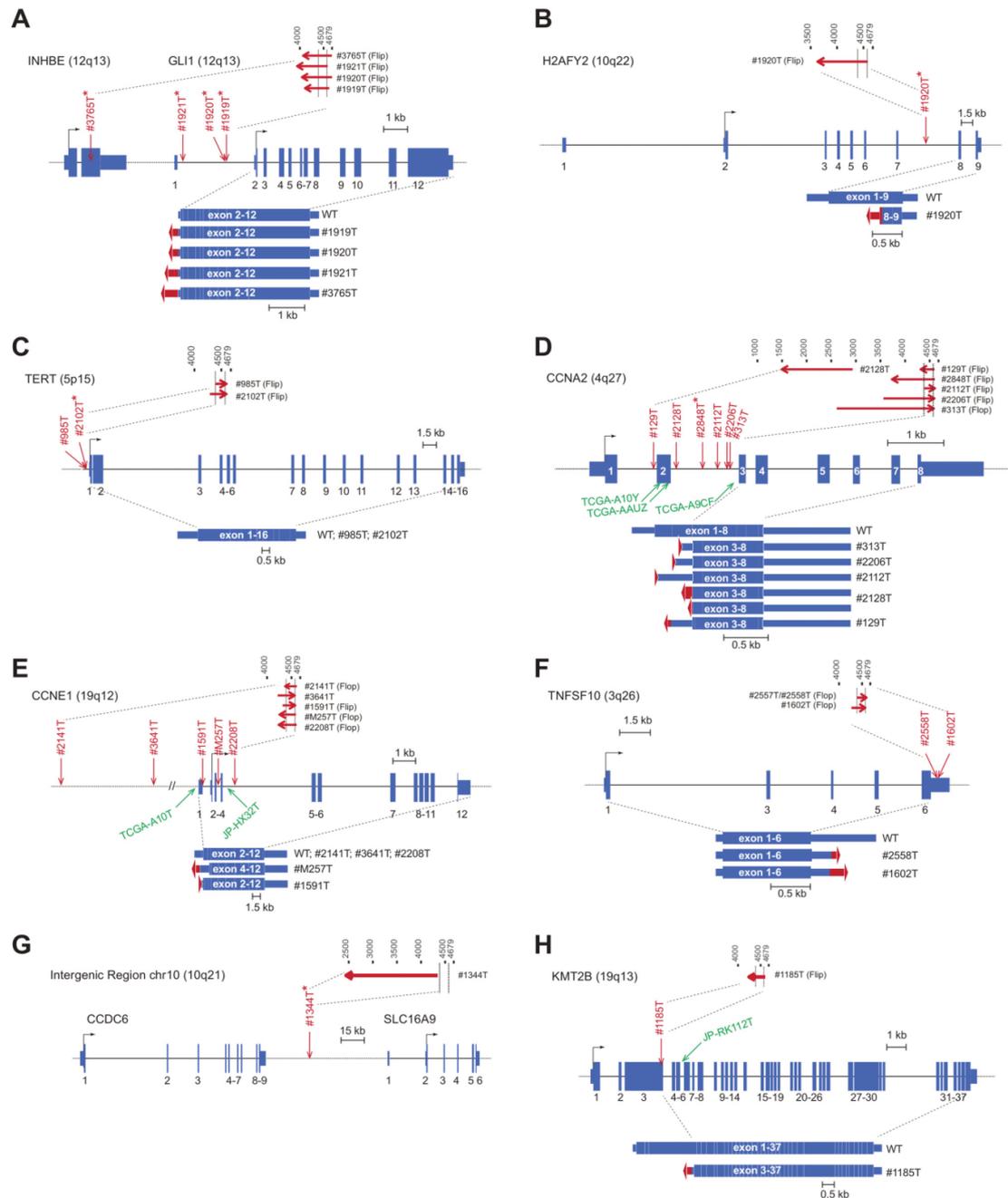


Figure 5 Adeno-associated virus (AAV) clonal integration sites and transcripts consequences in tumours. Genes structure are schematised with boxes referring to exons and lines to introns regions. Transcription start sites location is shown on 5' of the gene. Arrows indicate viral insertion sites in our series, in red, and in TCGA and ICGC tumours, in green. Asterisks refer to new inserted cases. Top lines refer to inserted AAV viral regions and arrows to 5'>3' sequence orientation. Flip or flop 3' ITR are indicated. Observed transcripts are represented at the bottom of the gene structure with fusion viral sequences in red. WT, wild type.

with focal nodular hyperplasia (FNH), AAV was more highly positive in the non-tumour counterpart than in the corresponding tumour (figure 4C). Finally, viral episomal forms

were rarely identified in tumours (n=8, 0.6%), mostly in benign tumours (four HCA and two FNH) and only two HCC (supplementary figure 3C-D and 4).

Hepatology

AAV insertion in liver tissues

We identified seven novel clonal insertions in six HCCs, in *GLI1/INHBE*, (figure 5A-B) *TERT* (figure 5C) and *CCNA2* (figure 5D). Only one clonal insertion was identified in a benign focal nodular hyperplasia, it occurred in an intergenic region of chromosome 10 (figure 5G) without consequences on the expression of the nearest genes (figures 4D and 5). Combining with AAV insertions identified in TCGA and ICGC sequenced HCC^{44 45} and previously described cases in our cohort,^{25 34} we re-analysed a total of 30 independent AAV insertions in liver tumours (online supplementary table 4). Viral insertions occurred in both directions, AAV2 and AAV2/13 subtypes were equally represented (55% vs 45% of the interpretable cases, respectively) and the minimal AAV region commonly inserted (nucleotide 4390–4570) was identified in 25 out of the 30 insertions.

Six oncogenes were recurrently activated by AAV (online supplementary figure 7). Insertions in *GLI1/INHBE* (four adenomas transformed into HCC, figure 5A-B), *TERT* (two HCC, figure 5C), *CCNE1* (seven HCC, figure 5E), *TNFSF10* (two HCC, figure 5F) and *KMT2B* (two HCC, figure 5H) led in almost all the cases to an overexpression of full-length coding region of these oncogenes by a promoter and/or an enhancer cis mechanism (figure 5). *CCNA2* was inserted in nine HCC; all insertions but one clustered in *CCNA2* intron 2, they resulted in an abnormal AAV-*CCNA2* transcript leading to a stable oncogenic truncated protein lacking the N-terminal regulatory domain (figure 5D).³⁴ The 3'UTR of *TNFSF10* showed AAV insertions in two HCC inducing *TNFSF10* overexpression with transcripts that prematurely ended at the viral polyadenylation (figure 5F). Here, using site-directed mutagenesis of both insertions, we

demonstrated that the viral polyA signal is required to ensure a strong luciferase overexpression in three different tested cell lines (online supplementary figure 8).

In the non-tumour liver tissues, no clonal AAV insertions were identified; non-clonal insertions were significantly associated with the presence of episomal AAV ($p < 0.001$), in contrast to the tumour samples. In both non-tumour and tumour tissues, non-clonal AAV insertions were randomly distributed along the genome (figure 4D and online supplementary figure 9). No specific enrichment was found in major target of AAV previously described in cell lines.^{46 47}

AAV features and tumour heterogeneity

We explored intertumour heterogeneity by analysing multi-nodules ($n=475$) from 186 patients for the presence of viral DNA, clonal insertions and episomal form. Of those, AAV DNA was detected in 25 patients (online supplementary figure 6C), including 4 patients with clonal AAV insertion in at least one nodule. Two patients with HCC displayed clonal AAV integrations in all nodules. Thanks to the next generation sequencing (NGS) data, we were able to predict the evolution of these tumours by looking at the common and private somatic mutations and copy number alterations (CNA) in each nodule. Interestingly, the two tumours from patient #2557 showed the same viral insertion in *TNFSF10*, similar gene mutations and CNA profiles, demonstrating that AAV insertion is a truck alteration occurring before intrahepatic metastasis (figure 6A). Conversely, the three tumours from patients #1919, resulting from a malignant transformation of adenoma in carcinoma, harboured three different clonal insertions all targeting *GLI1*, with different gene

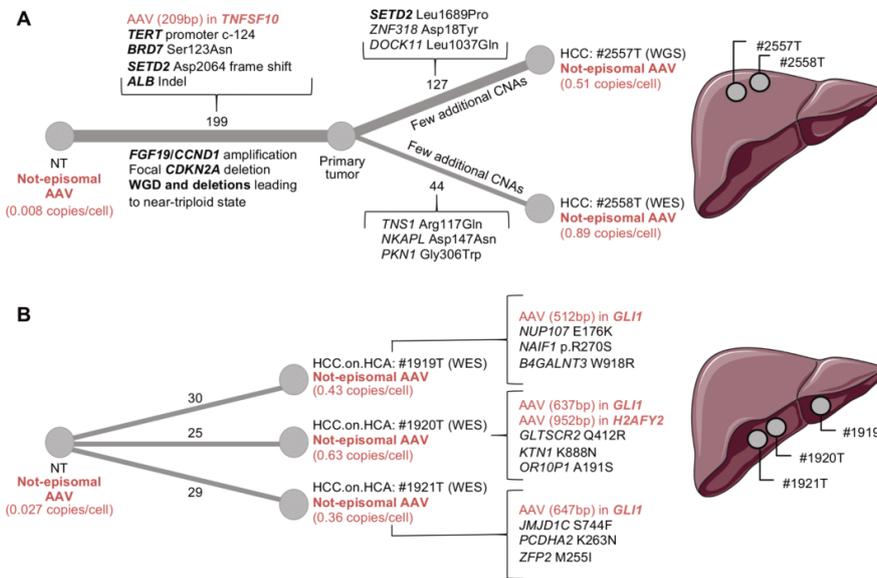


Figure 6 Tumour development in patients with multiple nodules and clonal adeno-associated virus (AAV) insertions. The relation between the tumours is determined according to gene mutation profile and copy number alteration (CNA) of each nodule. The number of shared and private alterations is indicated above each branch. The major alterations with amino acid consequences are listed; mutations in driver genes and main CNAs are in bold. The AAV status, diagnosis and sources of genomic information (WGS, WES) are specified for each nodule. The thickness of the branch indicates the number of alterations. The position of the nodules for each patient is represented on the right. (A) The two hepatocellular carcinoma (HCC) nodules of patients #2557 display the same AAV insertion in *TNFSF10* and they share 199 somatic mutations and several CNAs. This profile suggests that the nodules originate from the same primary tumour. (B) The three nodules of patient #1919 are heterogeneous for mutation profile and AAV insertions suggesting an independent origin of the tumours. NT, non-tumour.

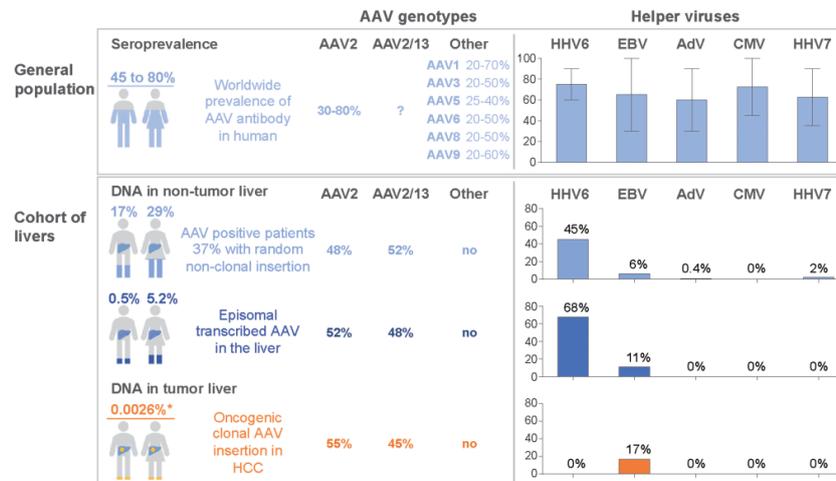


Figure 7 Adeno-associated virus (AAV) and helper viruses in the general population and in human liver. Frequency of different AAV genotypes and seroprevalence of AAV¹⁰⁻¹² and helper viruses^{50 51} in the general population are shown in the upper panel. The error bar in the histogram represents the range of helper viruses seroprevalence according to the literature. The bottom panel summarises the results found in this study, with estimated frequencies in the general population. For men and women, the global AAV frequency, the presence of episomal transcribed AAV and the prevalence of oncogenic clonal AAV insertions are indicated. *This prevalence is normalised according to the frequency of clonal AAV insertion in hepatocellular carcinoma (HCC) (2%) and the prevalence of HCC in France (0.013%). AdV, human adenovirus; EBV, Epstein Barr virus; HHV, human herpesvirus.

mutation profile and no CNA suggesting that the three nodules have an independent origin (figure 6B).

DISCUSSION

In this study, we provided a comprehensive description at large scale of the different AAV viral forms in the liver and of its oncogenic consequences, contributing to better understand the natural history of AAV infection in human.

The prevalence of AAV was observed in 21% of patients in non-tumour and/or tumour liver in agreement with the seroprevalence of antibody against AAV identified in 30%–80% of the general population.^{10-12 48} Our result showed that one out of five patients demonstrates persistent AAV DNA in the liver during life, mainly in the population of young and female patients without liver fibrosis (figure 7). However, since most of our liver tissues were sampled from patients with liver diseases, the exact prevalence of AAV DNA in the liver of healthy individuals remains to be evaluated.

Only two AAV genotypes, AAV2 and hybrids AAV2/13, were identified in our cohort equally distributed among the patients. AAV2/13 sequences were hybrids between AAV2 in the 5' part and AAV13 in the 3' corresponding to the previous clade C of the VP1 classification.¹⁴ Since only one full-length AAV sequence from clade C was publicly available,⁴² our work significantly increased the number of human AAV full-length sequences enlightening the genomic variants associated with an efficient natural AAV infection in the liver. In contrast to previous serological analysis,^{10 11 49} we did not identified other AAV genotypes in the liver, even if AAV5 and AAV8 were frequent in circulating monocytes.⁴⁸

AAV episomal form was identified in the non-tumour tissues of 4.6% of the patients, representing 26% of all AAV-positive liver samples, whereas episomal AAV has only been described in human tonsil and adenoid previously.⁹ It was frequently associated with viral mRNA expression suggesting that the

episomal AAV are also transcriptionally active in a significant proportion of the population in the liver (figure 7). Several viruses^{50 51} are able to support AAV replication in vitro, and it was commonly admitted that adenovirus is the natural AAV helper. Here, we identified HHV6 as the virus most frequently associated with episomal and transcribed AAV in the liver. This co-occurrence was previously described in healthy blood donors⁴⁸ and HHV6 is able to infect hepatocytes.⁵²⁻⁵⁴ The increased frequency of HHV6 in patients with episomal-expressed AAV form could indicate an ongoing active infection in the liver of 2.1% of the patients. In contrast, only very rare patients showed an association with adenovirus or other candidate helper viruses even in livers with episomal and expressed AAV (figure 7). All these results may suggest the role of HHV6 as the natural helper virus of AAV in the liver. However, co-infection with other helper viruses could occur at the initial acute AAV infection, followed by its clearance. Replication-competent infectious AAV has been rescued from human tonsil and adenoid tissue and lymphocytes, it remains to be searched in fresh liver tissues.^{48 55} Viral clones were isolated and their infectivity was tested in vitro in HeLa cells showing that only AAV clones with a complete double-D ITR structure were able to replicate and gave rise to infectious virus.⁵⁵ Interestingly, the analysis of the ITRs junctions of the episomal form in our series highlighted the presence of the same double-D structure supporting its role in an active AAV infection. Moreover, here a peculiar link between episomal-expressed AAV in the liver and age suggested that AAV active infection occurs during the first three decades of life and then remains latent.

Analyses of the tumour tissues confirmed the selection of clonal AAV insertion in HCC development in non-cirrhotic liver. Recurrent somatic viral integrations were identified in 2% of our HCC cohort, targeting *CCNA2* (33.3%), *CCNE1* (27.8%), *GLI1/INHBE* (11.1%), *TERT* (11.1%), *TNFSF10* (11.1%) and

Hepatology

KMT2B (5.6%). AAV insertion induced the overexpression of the target genes through multiple mechanisms that differ according to the target and the localisation of the integration. Clonal insertions upstream the TSS or within the 5' region of the gene lead to the gain of a positive regulatory mechanism such as the usage of viral enhancers and transcription factor binding sites (TFBS). Interestingly, a recent work by Logan *et al* has described a liver-specific enhancer-promoter element in wild-type AAV genome within the common inserted region in HCC tumours.²⁹ It consists of 124 nucleotides sequence that contains TFBSs for HNF1- α , HNF6 and GATA6. Noteworthy, this region is absent in many AAV vectors currently in use and should raise a biosafety flag or be deleted in the remaining. In line with our finding, this result strongly supports the mechanism of AAV-induced overexpression of the target gene. In addition, viral insertions in *CCNA2* and *TNFSF10* genes led the expression of a truncated protein or the premature ending of the transcript within the viral polyA, respectively.

AAV oncogenic integrations were identified in our cohort of European patients with HCC. They were also observed in the ICGC-Japan cohort in 3 HCC cases out of 268 HCC (1.1%),⁴⁵ in 4 out of 334 HCC (1.2%) of the TCGA cohort³⁴ and in 2 out of 289 HCC (0.7%) from Korea.⁵⁶ Interestingly, the most frequent AAV integrated oncogenes are similar to HBV, that are, *CCNA2*, *CCNE1*, *TERT* and *KMT2B*. The lower prevalence of AAV could be due to the lack of chronic liver disease associated with active AAV replication in contrast to chronic HBV infection. In the present series, we reinforced the link between AAV oncogenic insertion and the occurrence of HCC in normal liver, including recurrent AAV insertions in the malignant transformation of hepatocellular adenoma in carcinoma targeting *GLI1* that defines the activated sonic hedgehog molecular subgroup of adenoma, shHCA.⁵⁷ In the same line, AAV insertions in cyclin A2 or E1 in HCC are associated with unique chromosomal rearrangement signature and poor prognosis mainly occurring in HCC developed in normal liver.³⁴ These results underline the role of AAV insertion in the development of a specific subgroup of HCC without other aetiologies.

In conclusion, we provided a portrait of AAV infection in the liver with a description of viral genotypes, molecular forms and helper virus paving the way for a renovated interest in wild-type AAV biology. New highlights on the understanding of the oncogenic consequences of AAV integration in HCC tumours emerged from this work. However, further studies are necessary to clarify the impact of AAV infection in additional cohort of patients and the frequency of insertional mutagenesis across different countries.

Author affiliations

- ¹Centre de Recherche des Cordeliers, Sorbonne Universités, INSERM, Paris, Île-de-France, France
- ²Functional Genomics of Solid Tumor, Labex Immuno- Oncology, équipe labellisée Ligue Contre le Cancer, Université de Paris, Université Paris 13, Paris, Île-de-France, France
- ³Indian Statistical Institute, University of Kalyani, Kalyani, West Bengal, India
- ⁴Pathology Department, APHP, CHU Henri Mondor, Créteil, Île-de-France, France
- ⁵Pathology Department, APHP, Bicêtre-Paul Brousse Hospitals, Le Kremlin Bicêtre, Île-de-France, France
- ⁶Physiopathogenesis and treatment of liver diseases, INSERM, Paris, Île-de-France, France
- ⁷Pathology Department, APHP, Beaujon Hospital, Paris, Île-de-France, France
- ⁸The Research Center on Inflammation labeled, INSERM, Paris, Île-de-France
- ⁹Hepatology Department, APHP, Henri Mondor Hospital, Créteil, Île-de-France, France
- ¹⁰Molecular virology and immunology, INSERM, Institut Mondor de Recherche Biomédicale, Créteil, Île-de-France, France
- ¹¹Department of Digestive Surgery, APHP, Henri Mondor Hospital, Créteil, Île-de-France, France

- ¹²Department of Hepato-Gastroenterology and Digestive Oncology, CHU de Bordeaux, Haut-Lévêque Hospital, Bordeaux, Aquitaine, France
- ¹³Department of Digestive Surgery, Centre Médico Chirurgical Magellan, CHU de Bordeaux, Haut-Lévêque Hospital, Bordeaux, Aquitaine, France
- ¹⁴Department of Pathology, CHU de Bordeaux, Pellegrin Hospital, Bordeaux, Aquitaine, France
- ¹⁵Bordeaux Research in Translational Oncology, Université Bordeaux, Bordeaux, Aquitaine, France
- ¹⁶Department of Hepatology, Université Paris Nord, APHP, Hôpital Jean Verdier, Bondy, Île-de-France, France
- ¹⁷Department of Oncology, APHP, Hôpital Européen Georges Pompidou, Paris, Île-de-France, France

Acknowledgements The authors would like to thank surgeons, pathologists and all the clinicians who collected samples and clinical data. Sophie Prevost, Service d'anatomie pathologique, APHP, Hôpital Antoine-Béclère, Clamart, France. Anne de Muret, Service d'anatomopathologie, Centre Hospitalier Régional Universitaire de Tours, Tours, France. Eric Viber, Centre Hépatobiliaire, INSERM U785, Hôpital Paul Brousse, Villejuif, France. Philippe Merle, Department of Hepatology, Hospices Civils de Lyon, Croix-Rousse University Hospital, Lyon, France. Monique Fabre, Service Anatomie, HU-Necker Enfants Malades APHP, Paris. Nathalie Sturm, Department of Anatomie et Cytologie Pathologiques, CHU de Grenoble, Grenoble, France. Thomas Decaens, Service Hépatogastroentérologie et Tumeurs du foie, CHU de Grenoble, Grenoble, France. Sophie Michalak, Département de Pathologie cellulaire et tissulaire, CHU ANGERS. Georges-Philippe Pageau, service d'hépatogastro-entérologie Hôpital St Eloi CHU Montpellier. Jean-Michel Fabre, service de chirurgie digestive Hôpital St Eloi CHU Montpellier. Emmanuel Boleslawski, service de chirurgie digestive et transplantation. Hôpital Huriez. Chru de lille. 59037 lille cedex. Marie Christine Saint Paul, service d'Anatomie Pathologique, CHU de Nice, Nice. Dominique Wendum, Department of Pathology, Saint-Antoine Hospital, APHP, Paris, France. Olivier Rosmorduc, Department of Gastroenterology and Hepatology, Hôpital de la Pitié-Salpêtrière, APHP, Université Pierre et Marie Curie UPMC, Paris. Jean Christophe Vaillant, service de chirurgie hépatobilio- pancréatique, CHU Pitié-Salpêtrière, Université Pierre et Marie Curie UPMC, Paris. Marianne Ziol, Service d'Anatomopathologie, Hôpital Jean Verdier, Hôpitaux universitaires Paris-Seine- Saint-Denis, APHP, Bondy, France. Nathalie Ganne, Department of Hepatogastroenterology, Hôpital Jean Verdier, APHP, Bondy, France. Luigi Terraciano, Basel University Hospital, Department of Pathology, Basel, Switzerland. Vincenzo Mazzaferro, University of Milan at the Istituto Nazionale Tumori IRCCS (National Cancer Institute). Céline Bazille, Service d'Anatomie Pathologie, CHU de Caen, Caen, France.

Contributors Study design: J-CN, JZ-R, TLB, SI. Generation of experimental data: TLB, SI, CP, IM, SD. Analysis and interpretation of data: TLB, SI, CP, IM, SD, QB, SC, TZH, EL, J-CN, JZ-R. Collection of samples and related histological and clinical data: J-CN, JZ-R, JC, GM, CG, VP, GA, AL, LP, LC, PB-S, J-FB and investigators. Drafting of the manuscript: JZ-R, TLB, SI. Revision of the manuscript and approval of the final version of the manuscript: JZ-R, J-CN, TLB, SI, IM, SD, QB, SC, TZH, EL, JC, GM, CG, VP, GA, AL, LP, LC, PB-S, J-FB.

Funding This work was supported by INSERM, by INCa within the ICGC project, France Génomique, Cancéropole Ile de France (ExhauTrans project), ITMO Cancer AVIESAN (Alliance Nationale pour les Sciences de la Vie et de la Santé, National Alliance for Life Sciences & Health) within the framework of the Cancer Plan ("HTE program-HetColi network" and "Cancer et environnement program"), the Réseau national CRB Foie, Ligue Nationale contre le cancer: project équipe Labellisée, Fondation Schueller Bettencourt "coup d'élan", Prix Ligue contre le Cancer comité de Paris René et Andrée Duquesne 2018, the SIRIC CARPEM and Fondation Mérieux, Labex Oncoimmunology (investissement d'avenir) ANRS and the French Liver Biobanks network—INCa, BB-0033-00085, Hepatobio bank. QB is supported by a fellowship from the HOB doctoral school and the ministry of Education and Research, TLB is supported by an "Attractivité IDEX" fellowship from IUH and CP is supported by a doctoral fellowship funded by ANRS.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The sequencing data reported in this paper have been deposited to Genbank (accessions: KT258720-KT258730, MK139243-MK139299, MK163929-MK163942 and MK231253-MK231264) and EGA (European Genome-phenome Archive) database (RNA-seq accessions: EGAS000011002879, EGAS000011001284 and EGAS000011003310). All supplemental data and sequences are available at <http://zucmanlab.com/wp-content/uploads/2019/05/Zucman-CaptureVirusProbes.xlsx>.

ORCID iDs

- Sandrine Imbeaud <http://orcid.org/0000-0001-8439-6732>
- Julien Calderaro <http://orcid.org/0000-0002-5370-0807>
- Jean-Charles Nault <http://orcid.org/0000-0002-4875-9353>
- Jessica Zucman-Rossi <http://orcid.org/0000-0002-5687-0334>

REFERENCES

- 1 Rose JA, Berns KI, Hoggan MD, *et al.* Evidence for a single-stranded adenovirus-associated virus genome: formation of a DNA density hybrid on release of viral DNA. *Proc Natl Acad Sci U S A* 1969;64:863–9.
- 2 Atchison RW, Casto BC, Hammon WM. Adenovirus-associated defective virus particles. *Science* 1965;149:754–5.
- 3 Balakrishnan B, Jayandharan GR. Basic biology of adeno-associated virus (AAV) vectors used in gene therapy. *Curr Gene Ther* 2014;14:86–100.
- 4 Sonntag F, Schmidt K, Kleinschmidt JA. A viral assembly factor promotes AAV2 capsid formation in the nucleolus. *Proc Natl Acad Sci U S A* 2010;107:10220–5.
- 5 Yang CC, Xiao X, Zhu X, *et al.* Cellular recombination pathways and viral terminal repeat hairpin structures are sufficient for adeno-associated virus integration in vivo and in vitro. *J Virol* 1997;71:9231–47.
- 6 Nakai H, Wu X, Fuess S, *et al.* Large-scale molecular characterization of adeno-associated virus vector integration in mouse liver. *J Virol* 2005;79:3606–14.
- 7 Hüser D, Gogol-Döring A, Chen W, *et al.* Adeno-associated virus type 2 wild-type and vector-mediated genomic integration profiles of human diploid fibroblasts analyzed by third-generation PacBio DNA sequencing. *J Virol* 2014;88:11253–63.
- 8 Schultz BR, Chamberlain JS. Recombinant adeno-associated virus transduction and integration. *Mol Ther* 2008;16:1189–99.
- 9 Schnepf BC, Jensen RL, Chen CL, *et al.* Characterization of adeno-associated virus genomes isolated from human tissues. *J Virol* 2005;79:14793–803.
- 10 Boutin S, Montellhet V, Veron P, *et al.* Prevalence of serum IgG and neutralizing factors against adeno-associated virus (AAV) types 1, 2, 5, 6, 8, and 9 in the healthy population: implications for gene therapy using AAV vectors. *Hum Gene Ther* 2010;21:704–12.
- 11 Calcedo R, Vandenberghe LH, Gao G, *et al.* Worldwide epidemiology of neutralizing antibodies to adeno-associated viruses. *J Infect Dis* 2009;199:381–90.
- 12 Erles K, Seböková P, Schlehofer JR. Update on the prevalence of serum antibodies (IgG and IgM) to adeno-associated virus (AAV). *J Med Virol* 1999;59:406–11.
- 13 Gao GP, Alvira MR, Wang L, *et al.* Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. *Proc Natl Acad Sci U S A* 2002;99:11854–9.
- 14 Gao G, Vandenberghe LH, Alvira MR, *et al.* Clades of Adeno-associated viruses are widely disseminated in human tissues. *J Virol* 2004;78:6381–8.
- 15 Mori S, Wang L, Takeuchi T, *et al.* Two novel adeno-associated viruses from cynomolgus monkey: pseudotyping characterization of capsid protein. *Virology* 2004;330:375–83.
- 16 Schmidt M, Voutetakis A, Afione S, *et al.* Adeno-associated virus type 12 (AAV12): a novel AAV serotype with sialic acid- and heparan sulfate proteoglycan-independent transduction activity. *J Virol* 2008;82:1399–406.
- 17 Chandler RJ, Sands MS, Venditti CP. Recombinant adeno-associated viral integration and genotoxicity: Insights from animal models. *Hum Gene Ther* 2017;28:314–22.
- 18 Berns KI, Muzyczka N. AAV: An overview of unanswered questions. *Hum Gene Ther* 2017;28:308–13.
- 19 Donsante A, Miller DG, Li Y, *et al.* AAV vector integration sites in mouse hepatocellular carcinoma. *Science* 2007;317:477.
- 20 Chandler RJ, LaFave MC, Varshney GK, *et al.* Vector design influences hepatic genotoxicity after adeno-associated virus gene therapy. *J Clin Invest* 2015;125:870–80.
- 21 Wang PR, Xu M, Toffanin S, *et al.* Induction of hepatocellular carcinoma by in vivo gene targeting. *Proc Natl Acad Sci U S A* 2012;109:11264–9.
- 22 Hermonat PL, Plott RT, Santin AD, *et al.* Adeno-associated virus Rep78 inhibits oncogenic transformation of primary human keratinocytes by a human papillomavirus type 16-ras chimeric. *Gynecol Oncol* 1997;66:487–94.
- 23 Liu T, Cong M, Wang P, *et al.* Adeno-associated virus Rep78 protein inhibits Hepatitis B virus replication through regulation of the HBV core promoter. *Biochem Biophys Res Commun* 2009;385:106–11.
- 24 Kokorina NA, Santin AD, Li C, *et al.* Involvement of protein-DNA interaction in adeno-associated virus Rep78-mediated inhibition of HIV-1. *J Hum Virol* 1998;1:441–50.
- 25 Nault JC, Datta S, Imbeaud S, *et al.* Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat Genet* 2015;47:1187–93.
- 26 Zhao LH, Liu X, Yan HX, *et al.* Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nat Commun* 2016;7:12992.
- 27 Sung WK, Zheng H, Li S, *et al.* Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet* 2012;44:765–9.
- 28 Llovet JM, Zucman-Rossi I, Pikarsky E, *et al.* Hepatocellular carcinoma. *Nat Rev Dis Primers* 2016;2:16018.
- 29 Logan GJ, Dane AP, Hallwirth CV, *et al.* Identification of liver-specific enhancer-promoter activity in the 3' untranslated region of the wild-type AAV2 genome. *Nat Genet* 2017;49:1267–73.
- 30 Benaglia T, Chauveau D, Hunter DR, *et al.* mixtools: An R Package for Analyzing Mixture Models. *Journal of Statistical Software* 2009;32:1–29.
- 31 Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26:589–95.
- 32 Li H, Handsaker B, Wysoker A, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–9.
- 33 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–2.
- 34 Bayard Q, Meunier L, Peneau C, *et al.* Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. *Nat Commun* 2018;9:5235.
- 35 Kim D, Pertea G, Trapnell C, *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013;14:R36.
- 36 Trapnell C, Roberts A, Goff L, *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012;7:562–78.
- 37 Sloutskin A, Danino YM, Orenstein Y, *et al.* ElemeNT: a computational tool for detecting core promoter elements. *Transcription* 2015;6:41–50.
- 38 Nishikawa T, Ota T, Isogai T. Prediction whether a human cDNA sequence contains initiation codon by combining statistical information and similarity with protein sequences. *Bioinformatics* 2000;16:960–7.
- 39 Liu H, Han H, Li J, *et al.* An in-silico method for prediction of polyadenylation signals in human sequences. *Genome Inform* 2003;14:84–93.
- 40 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.
- 41 Mroske C, Rivera H, Ul-Hasan T, *et al.* A capillary electrophoresis sequencing method for the identification of mutations in the inverted terminal repeats of adeno-associated virus. *Hum Gene Ther Methods* 2012;23:128–36.
- 42 Chen CL, Jensen RL, Schnepf BC, *et al.* Molecular characterization of adeno-associated viruses infecting children. *J Virol* 2005;79:14781–92.
- 43 Pénaud-Budloo M, Le Guiner C, Nowrouzi A, *et al.* Adeno-associated virus vector genomes persist as episomal chromatin in primate muscle. *J Virol* 2008;82:7875–85.
- 44 Cancer Genome Atlas Research Network. Electronic address wbe, cancer genome atlas research N. Comprehensive and integrative genomic characterization of hepatocellular carcinoma. *Cell* 2017;169:1327–41.
- 45 Fujimoto A, Furuta M, Totoki Y, *et al.* Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet* 2016;48:500–9.
- 46 Kotin RM, Siniscalco M, Samulski RJ, *et al.* Site-specific integration by adeno-associated virus. *Proc Natl Acad Sci U S A* 1990;87:2211–5.
- 47 Samulski RJ, Zhu X, Xiao X, *et al.* Targeted integration of Adeno-Associated Virus (AAV) into human chromosome 19. *Embo J* 1991;10:3941–50.
- 48 Hüser D, Khalid D, Lutter T, *et al.* High Prevalence of Infectious Adeno-associated Virus (AAV) in human peripheral blood mononuclear cells indicative of t lymphocytes as sites of AAV persistence. *J Virol* 2017;91.
- 49 Li C, Narkbunnam N, Samulski RJ, *et al.* Neutralizing antibodies against adeno-associated virus examined prospectively in pediatric patients with hemophilia. *Gene Ther* 2012;19:288–94.
- 50 Mast TC, Kierstead L, Gupta SB, *et al.* International epidemiology of human pre-existing adenovirus (Ad) type-5, type-6, type-26 and type-36 neutralizing antibodies: correlates of high Ad5 titers and implications for potential HIV vaccine trials. *Vaccine* 2010;28:950–7.
- 51 Steininger C, Rässenti LZ, Vanura K, *et al.* Relative seroprevalence of human herpes viruses in patients with chronic lymphocytic leukaemia. *Eur J Clin Invest* 2009;39:497–506.
- 52 Ozaki Y, Tajiri H, Tanaka-Taya K, *et al.* Frequent detection of the human herpesvirus 6-specific genomes in the livers of children with various liver diseases. *J Clin Microbiol* 2001;39:2173–7.
- 53 Ishikawa K, Hasegawa K, Naritomi T, *et al.* Prevalence of herpesviridae and hepatitis virus sequences in the livers of patients with fulminant hepatitis of unknown etiology in Japan. *J Gastroenterol* 2002;37:523–30.
- 54 Cermelli C, Concari M, Carubbi F, *et al.* Growth of human herpesvirus 6 in HEPG2 cells. *Virus Res* 1996;45:75–85.
- 55 Schnepf BC, Jensen RL, Clark KR, *et al.* Infectious molecular clones of adeno-associated virus isolated directly from human tissues. *J Virol* 2009;83:1456–64.
- 56 Park KJ, Lee J, Park JH, *et al.* Adeno-associated virus 2-mediated hepatocellular carcinoma is very rare in Korean patients. *Ann Lab Med* 2016;36:469–74.
- 57 Nault JC, Couchy G, Balabaud C, *et al.* Molecular classification of hepatocellular adenoma associates with risk factors, bleeding, and malignant transformation. *Gastroenterology* 2017;152:880–94.

Gut: first published as 10.1136/gutjnl-2019-318281 on 2 August 2019. Downloaded from http://gut.bmj.com/ on June 15, 2020 at INSEHM Consortia. Protected by copyright.

Supplementary Information

Adeno-associated virus in the liver: natural history and consequences in tumor development

La Bella T, Imbeaud S. et al.

Supplementary Tables

Supplementary Table 1 : Description of the series (n=1461 patients)

Supplementary Table 2: List of probe sets and primers

Supplementary Table 3: Results of the viral DNA screening

Supplementary Table 4: List of AAV clonal inserted sites in liver tumors

Supplementary Figures

Supplementary figure 1. AAV detection using quantitative PCR and viral capture.

Supplementary figure 2. Phylogenetic tree on VP1 coding protein region.

Supplementary figure 3. Viral episomal form investigation.

Supplementary figure 4. Viral 3'ITR-5'ITR episomal junction determined by in silico analysis.

Supplementary figure 5. AAV presence in tumor tissues and comparison with non-tumor counterparts.

Supplementary figure 6. AAV presence in tumor tissues and comparison with non-tumor counterparts.

Supplementary figure 7. Impact of clonal AAV integration on the expression of the target oncogenes.

Supplementary figure 8. Viral polyadenylation signal usage in TNFSF10.

Supplementary figure 9. Description of AAV insertions in non-tumor and tumor liver tissues.

Supplementary Materials and Methods

Computational analysis of human-AAV sequences

Experimental procedure to search for episomal forms of AAV

Supplementary table 1. Description of the series (n=1461 patients)

		Available data	n (%) or median (range)	
HCC patients (n=936)	Gender (female)	934	163 (17.5%)	
	Age	930	63 (0-90)	
	Transformed HCA (HCC on HCA)	936	17 (1.8%)	
	Largest nodule diameter (>5cm)	909	425 (46.8%)	
	Number Nodules (≥2)	871	244 (28%)	
	Preoperative serum AFP (>20ng/ml)	727	293 (40.3%)	
	Risk factors	Alcohol	895	393 (43.9%)
		HBV	907	164 (18.1%)
		HCV	896	224 (25%)
		Metabolic syndrome	843	201 (23.8%)
		Without etiology	912	128 (14%)
	Metavir score	F0-F1	929	264 (28.4%)
		F2-F3	929	236 (25.4%)
		F4	929	439 (46.2%)
	G1-G6 classification	G1	596	44 (7.4%)
		G2	596	47 (7.9%)
G3		596	105 (17.6%)	
G4		596	210 (35.2%)	
G5		596	129 (21.6%)	
G6		596	61 (10.2%)	
Tumor differentiation	Edmonson I-II	869	381 (43.8%)	
	Edmonson III-IV	869	488 (56.2%)	
Vascular invasion	Macrovascular	827	127 (15.4%)	
	Microvascular	827	287 (34.7%)	
HCA and FNH patients (n=322)	Gender (female)	320	270 (84.4%)	
	Age	312	37 (2-68)	
	Borderline tumors (HCA/HCC)	322	14 (4.3%)	
	Largest nodule diameter (>5cm)	221	136 (61.5%)	
	Number Nodules (≥2)	216	100 (46.3%)	
	BMI (>25kg/m ²)	182	69 (37.9%)	
	Children (women only)	132	100 (75.8%)	
	Oral contraception (women only)	147	128 (87.1%)	
	Diabetes	174	19 (10.9%)	
	Glycogenesis	160	7 (4.4%)	
	Metavir score	F0-F1	229	210 (91.7%)
		F2-F3	229	11 (4.8%)
		F4	229	8 (3.5%)
	Molecular subgroups	HHCA	314	62 (19.7%)
		b ^{ex7,8} -HCA	314	7 (2.2%)
		b ^{ex7,8} -IHCA	314	14 (4.5%)
IHCA		314	66 (21%)	
b ^{ex3} -IHCA		314	24 (7.6%)	
b ^{ex3} -HCA		314	19 (6.1%)	
shHCA		314	10 (3.2%)	
UHCA		314	15 (4.8%)	
FNH		314	97 (30.9%)	
Histological tumor type		HB/TLCT	203	87 (42.9%)
	CCK	203	46 (22.7%)	
	FLC	203	36 (17.7%)	
	others*	203	34 (16.7%)	
	*DMN, LGDN, HGDN, angiomyolipoma, liver carcinosarcoma, liver rhabdoid tumor, solitary fibrous tumor, mesenchymal Hamartoma, embryonal sarcoma, leiomyoma, Yolk-salk-tumor, neuroendocrine carcinoma, neuroblastoma, pecoma, nephroblastoma, malignant peripheral nerve sheath tumor			

Supplementary table 2. List of probe sets and primers

Probe Name	Forward sequence 5'-3'	Reverse sequence 5'-3'	Probe sequence 5'-3'	Target	used for	
AAV2_ORF1	CGGCATTTCTGACAGCTTTG	GGGTGCTGCTCAATCA	TGGCCGAGAAGGAATGGGAGTTG	AAV	qPCR, DNase/TaqMan assay, RT-qPCR	
AAV2_2_NEW	CTTCTACGGGTGCGTAAACT	GACCTTGGCGGTCTATCTT	AGATGGTGTATCTGGTGGGAGGAG	AAV	qPCR, DNase/TaqMan assay, RT-qPCR	
AAV2_3_NEW	AGACGCGAGCTCAGTACCT	CCCTCGTTATTGCTGCCAT	AATACGATGGCTACAGGCAGTGGC	AAV	qPCR, DNase/TaqMan assay, RT-qPCR	
AAV2_4_NEW	TGACATTCGGGACCAGTCTA	TAGCTCCAGTCCACGAGTATT	TTACCCGACGACGAGTATCAAA	AAV2	qPCR, DNase/TaqMan assay, RT-qPCR	
AAV2_Itr3_NEW	CCTGACTCGTAATCTGTAATTGC	ACGTAGCCATGGAACTAGATAAG	TCGTTTCAGTTGAACCTTTGGTCTCTGC	AAV	qPCR, DNase/TaqMan assay, RT-qPCR	
AAV2_Itr3	GGTCTCTGGCTATTCTTCTTATC	GTGGCCAACCTCATCACTA	ACGTAGATAAGTAGCATGGCGGGT	AAV	qPCR, DNase/TaqMan assay, RT-qPCR	
AY6_2	CTATGGCCAGCCACAAAGA	GTCGTTGGCATTGTCTCTT	CCCCATGCATGGAAACCCTGATA	AAV2/13	qPCR, DNase/TaqMan assay, RT-qPCR	
AY6_1	CCACAGCATGTCTTCCA	ATACTTTGTAGCCGACGTCCA	AACTGGCTGCCTGGACCTTGCTA	AAV2/13	qPCR, DNase/TaqMan assay	
AAV_HidIII	CCCCAGTGACGCAGATATAAGT	CCACGTGACGAGAACATTGG	CAGACGCGAAAGCTTCCGATCA	AAV	DNase/TaqMan assay	
AAV_Xmnl	TGGAAGTCAACGGGTACGAT	ACAGCCAGATGGTGTCTCT	TCTGGATGGGCCACGAAA	AAV	DNase/TaqMan assay	
AdV_hexon1C	GCTACCCCTTCGATGATGC	ACCGTGGGGTTCTAAACTTG	AGTGGTCTTACATGCACATCTCGGG	AdV	qPCR	
AdV_hexon2C	CACCGCTAGACATGACTTTT	ACCACGTCAAAGACTTCAAACA	CATGGACGAGCCACCCTTCTTT	AdV	qPCR	
AdV_hexonACF	CGCAGTGGTCTTACATGCAC	ACCGTGGGGTTCTAAACTTG	CAGGACGCCTCGGAGTACCTGA	AdV	qPCR	
AdV_hexon1B	CCGTGCAACGACACACTACTT	TGTAAGATGATGATTTGCTCCCG	ACCACGATGTGACACCAGC	AdV	qPCR	
AdV_hexon2B	GGACATGACTTTTGAGGTGGA	GTGGCTGGTGCACCTCTGA	CCCACCTGCTTATCTTCTTTTGA	AdV	qPCR	
AdV_hexon1F	AACACGGAGCTGTCTACCAGT	GTCTGGGTCTAGCTGTCCAC	TCGATACTTCTCATGTGGAAACCAGG	AdV	qPCR	
AdV_hexon2F	TCCTTTGGACAAGCTCCCTA	GCTCCGGTTGGTACGTTTAT	CAAGTGGGCTCAGACTCCAACAATC	AdV	qPCR	
AdV_hexon1A	CATACCAGTTAATGCTGGATGCT	CATCTCTACCCGCTGATTCT	ACGACCTGACGTTCCGCTTATT	AdV	qPCR	
AdV_hexon2A	CTCCCACCATGATGTTTACC	GGCTAAGCGAGCCTTGTAG	TGCGTTTGTGCGCCGTGGAT	AdV	qPCR	
AdV_hexon1D	GGTCTGGTGCAGTTTGGC	ACCGTGGGGTTCTAAACTT	CCACGACACGTAATTCAGCCTG	AdV	qPCR	
AdV_hexon2D	GAGTTCCTCCGGAAACGAC	TTGAAGGACTGGTGTGGT	AACCTCTACGCCACATTTCTCCCA	AdV	qPCR	
AdV_hexon1E	AGTGCAACATGACCAAGGACT	AGTTGCGGAAGAAGGAGTACAT	CTGGCCACTACAACATCGGCT	AdV	qPCR	
AdV_hexon2E	CGCTTCGGAGTACCTGAGTC	TGGGGTTCTAAACTTGTCC	CGCGCCACAGACACTACTTCAGT	AdV	qPCR	
AdV_polBE	CCATCCAGGTGATAGTTTC	CACATCAACGCCATTCTCT	AAGAACTGGATCTCTGCCACCAGT	AdV	qPCR	
AdV_polD	CCCATCCAGGTGATAGTTTC	TTTATCATCAACAGCCACT	AAGAAGTGGATTTCTGCCACCAGT	AdV	qPCR	
AdV_polA	CCCATCCAGGTGATAGTTTC	GATCCACTTTTCCCAATCG	AAGGGCTCAGTGCAGGATG	AdV	qPCR	
AdV_polF	CCAAGCTGTCCGGTGTAGTAT	ATCAAACTCCACTCGTTTGTG	CGCGTTGAAAGGTGGGCATAAC	AdV	qPCR	
AdV_polC	CGGGTTGAGGACAAACTCTT	TTTACATGTAGGCTTCTACC	TCCAGTACTTGGATCGGAAACCC	AdV	qPCR	
HSV_1.1	TCCTGGCTCTCGAGTAGTT	GTTCGTGTCGGTCAAGGAGT	CGTTGGCCGTGAGCCACTTT	HSV1/2	qPCR	
HSV_1.2	TCGAGGAGTAGTGGTGTATGC	CGTITGTGTCGGTCAAGAGT	CGGTGGCCGTGATCCACTT	HSV1/2	qPCR	
HSV_2.1	TTTGACTACGACAGAAAGTTGC	CTCCGTGACATACAGGGTAT	CCCTACGTCGACATAGCCAATC	HSV1/2	qPCR	
HSV_2.2	ACTGCTGATCGACCTGTGTG	TTATATGCTGGACGAGAAGG	CCCCATATGTTGATCACAGCCAA	HSV1/2	qPCR	
HSV_3	AGGACAGCGTTTGCATTTGG	TGGGTTGTAGGCTGTCTGTC	CGGTGCGTGTGATGGTGA	HSV1/2	qPCR	
HSV_4	GAGGACATCCAGGACTTTGT	CGGGCCATGAGCTTGTAA	ACCGCCAACTGAGCAGACAC	HSV1/2	qPCR	
HHV4_1	TGGAGCGAAGGTTAGTGGTC	GGGGAAAATGTACATTTGG	TGCAGTGCCTGGTGTCTTTA	EBV/HHV4	qPCR	
HHV4_2	CTGAGGCCCTTCTCTCTTTT	CCTGAACAGCCTTGGATAGC	TTTGAGGGGTGGGGAAATATGG	EBV/HHV4	qPCR	
HHV4_3	CAGCAGTAGGCGTAAACAA	CCTACAAGGACCTGTCAAGAG	CGCAAAGGGGTTGCACGAGT	EBV/HHV4	qPCR	
HHV4_4	TAGTGTCCGGAAATAGGTCT	CCACTACCAGGAGGAGAATA	TCATGTAGGGCCCTGATGATG	EBV/HHV4	qPCR	
HHV5_1	AGGCTTTTACCTCATCATCCA	TGGGAATCCGTAACCAACA	GGAAACGCGTTCGATCTTTATCC	CMV/HHV5	qPCR	
HHV5_2	GTTAGGTGACACGGCAACACT	ATTTTATCTTCCCAACACT	TTTCCGGCACCGTCTACCTTT	CMV/HHV5	qPCR	
HHV5_3	TGGCCAAACGTTCCGAGTT	ACGTGTCCGCTTTGAAACC	TGCCAGAACACTACCAACCACCA	CMV/HHV5	qPCR	
HHV5_4	TCAAACGTGCGCTTATCTTG	AAGACGTACAGCAGCGGATT	CGAGTGCCTCGCTTTTGTCACT	CMV/HHV5	qPCR	
HHV5_5	CATCTCCGACTTGTAGGGTAGTG	CTCTCAACGCGTAACTTCTTC	CGTTAAACACCATGCTGACCGTTT	CMV/HHV5	qPCR	
HHV6_1	GATAGTATCCCTCTTCTCCCATC	ACGGTTAATCTTTGAGGACGAA	CCAGAATCGAGAAGTGGCCAG	HHV6	qPCR	
HHV6_2	CATCTCTCCGTTATGTCTTTCA	GGAGTTAATGACCACGACATT	AGCGAGAGATAGGGATGGTTGGGA	HHV6	qPCR	
HHV6_3	TCTAATAGAGCTGGTTGGATGAA	C	CGGAAATACAATACTCGTCTCT	GGCAGAGCGTTTGTGAAGACTGGC	HHV6	qPCR
HHV6_4	CCGCACTATGCTGGATTAG	ACTGGACTGTACTGTAGGAATCGT	TATAAGTTCGGGCCGTATGGGTG	HHV6	qPCR	
HHV6_5	CCGCACTATGCTGGATTAG	GGATCTGTACCGTAGGAATCGTT	TAAGTTCGGGCCGTAGGGGT	HHV6	qPCR	
HHV7_1	CATCCAGAGCGTAGACAGCA	GCCGTGGTTATCGGAAAGT	AAGACCACGACCGGAGGATCTTC	HHV7	qPCR	
HHV7_2	GTCCGTTAGAACCTCCATCAAC	TTGGCTGTACCAGGAATCAC	TCAAGGCCTGAAAAAGCGAACGT	HHV7	qPCR	
HHV7_3	GGATTAGGTAGAGTGTGGCG	CATTTAGTGGCTGTCTCTTTC	TGCGGATGTCATTCGGTATAGAGAGG	HHV7	qPCR	
HHV7_4	AGTTGGCCGATACGACACA	TCACCGGTTCTCTGTCTCTAC	CCGGAAGGCTGGATGGAATCTAGG	HHV7	qPCR	
HHV8_1	AGAGTCCGCATCACCAA	ACCCGTTGACATTTACTCTT	ACCCGTGCCACTCTATGAGATAAGCC	HHV8	qPCR	
HHV8_2	CTGGGCAAGCAGTTTTTCA	TTCGGTAGGTGAGGTCTTTGTC	TCAGGAAGACGGCCTAGAGCGATAC	HHV8	qPCR	
HHV8_3	ATCTGTCTTCCATCTTTCC	GTCCAGTATCCACCCAAA	TTGTGGCTAGCTTTCGAGGAC	HHV8	qPCR	
HHV8_4	GTCTTCCATCTTTCCAAAA	ACCGTCCCTTACAGACAGATA	TGGCCTAGCTTTCGAGGACA	HHV8	qPCR	
HMB5	GTCCAGCCTGTTGGTGAG	GGCCACAATTCAGATCTTCTTA	CTTTGAGTCAATCCAGAGCCCT	human	qPCR	
RHO	CCACACAGAAGGCAGAGAAG	TGGTGGTGAAGATGTAGAATG	TCATCGCTTCTGATCTGTCTGGG	human	qPCR	
HPV_E1		Vi03453396_s1		HPV16	qPCR	
ANKRD49		Hs06279676_en		human	qPCR	
I8S		Hs03928990_g1		human	RT-qPCR	
ACTB		Hs01060665_g1		human	RT-qPCR	
CCNA2		Hs00996788_m1		human	RT-qPCR	
CCNE1		Hs01026536_m1		human	RT-qPCR	
TERT		Hs00972656_m1		human	RT-qPCR	
GLI1		Hs00171790_m1		human	RT-qPCR	
KMT2B		Hs00207065_m1		human	RT-qPCR	
TNFSF10		Hs00921974_m1		human	RT-qPCR	
Primer name	Sequence 5'-3'			Target	used for	
AAV2_tot_6_F_bis	ACAGTACTCCACGGGACAGG			AAV	Itr3s junction PCR and sequencing	
AAV2_epi_R_2	GCTGGGACCTTAATCACAA			AAV	Itr3s junction PCR and sequencing	
AAV2_6F_NEW	CCTGACTCGTAATCTGTAATTGC			AAV	Itr3s junction PCR and sequencing	
AAV2_Itr5_F int/new	GTGGCCAACCTCATCACTAG			AAV	Itr3s junction PCR and sequencing	

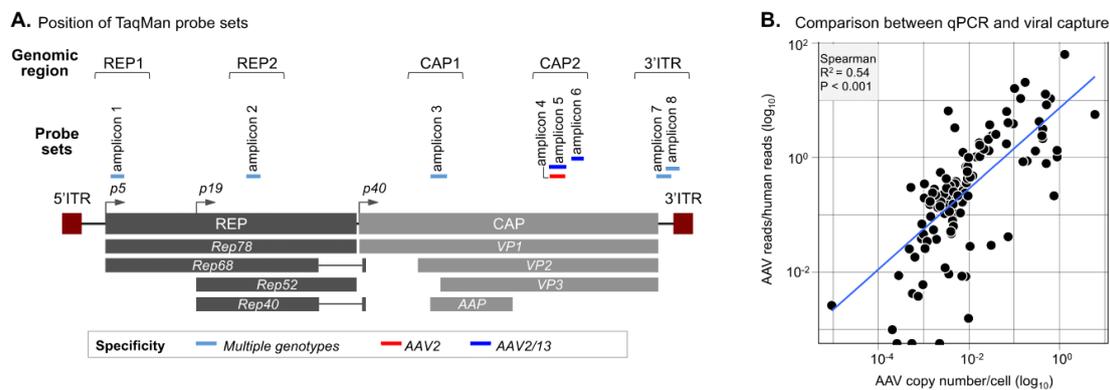
Supplementary table 3. Results of the viral DNA screening

		AAV positivity	Copy nb/cell (mean [sd])	Episomal AAV	Expressed episomal AAV	AAV subtype			HHV6 co-infection
						AAV2	AAV2 /13	Unknown	
All samples	NT	17.7% (233/1319)	0.007 [± 0.021]	4.6% (60/1304)	2.1% (28/1304)	29.2% (68/233)	32.2% (75/233)	38.6% (90/233)	7.9% (104/1318)
	T	7.8% (109/1406)	0.141 [± 0.617]	0.6% (8/1397)	0.2% (3/1397)	22.9% (25/109)	31.2% (34/109)	45.9% (50/109)	1.1% (15/1405)
Benign	NT	27% (78/289)	0.008 [± 0.021]	12% (34/284)	6% (17/284)	37.2% (29/78)	38.5% (30/78)	24.3% (19/78)	12.5% (36/289)
	T	9.4% (31/331)	0.048 [± 0.234]	1.8% (6/327)	0.9% (3/327)	41.9% (13/31)	38.7% (12/31)	19.4% (6/31)	1.8% (6/331)
Malignant	NT	15% (155/1030)	0.007 [± 0.021]	2.5% (26/1020)	1.1% (11/1020)	25.2% (39/155)	29% (45/155)	45.8% (71/155)	6.6% (68/1029)
	T	7.3% (78/1075)	0.178 [± 0.712]	0.2% (2/1070)	0% (0/1070)	15.4% (12/78)	28.2% (22/78)	56.4% (44/78)	0.8% (9/1074)
HCA	NT	25.3% (47/186)	0.007 [± 0.017]	11% (20/182)	6% (11/182)	36.2% (17/47)	42.5% (20/47)	21.3% (10/47)	12.9% (24/186)
	T	7.9% (17/215)	0.006 [± 0.013]	1.9% (4/215)	0.9% (2/215)	23.5% (4/17)	52.9% (9/17)	23.5% (4/17)	1.4% (3/215)
FNH	NT	32.1% (27/84)	0.012 [± 0.029]	14.5% (12/83)	7.2% (6/83)	40.7% (11/27)	29.6% (8/27)	29.6% (8/27)	14.3% (12/84)
	T	12.5% (12/96)	0.106 [± 0.362]	2.2% (2/92)	1.1% (1/92)	75% (9/12)	25% (3/12)	0% (0/12)	2.1% (2/96)
HCC	NT	15.4% (135/875)	0.004 [± 0.010]	2.2% (19/868)	0.8% (7/868)	24.4% (33/135)	28.9% (39/135)	46.7% (63/135)	6.5% (57/875)
	T	7.6% (69/909)	0.200 [± 0.755]	0.2% (2/906)	0% (0/906)	15.9% (11/69)	26.1% (18/69)	58% (40/69)	0.8% (7/908)
HB/TLCT	NT	9.7% (7/72)	0.020 [± 0.048]	1.4% (1/69)	1.4% (1/69)	42.9% (3/7)	28.6% (2/7)	28.6% (2/7)	2.8% (2/72)
	T	6.7% (5/75)	0.013 [± 0.017]	0% (0/73)	0% (0/73)	20% (1/5)	20% (1/5)	60% (3/5)	2.7% (2/75)
CCK	NT	16.3% (7/43)	0.004 [± 0.006]	4.7% (2/43)	2.3% (1/43)	0% (0/7)	28.6% (2/7)	71.4% (5/7)	11.6% (5/43)
	T	4.4% (2/45)	0.001 [± 0.001]	0% (0/45)	0% (0/45)	0% (0/2)	50% (1/2)	50% (1/2)	0% (0/45)
FLC	NT	19.4% (6/31)	0.056 [± 0.080]	12.9% (4/31)	3.2% (1/31)	33.3% (2/6)	50% (3/6)	16.7% (1/6)	10% (3/30)
	T	6.2% (2/32)	0.001 [± 0.001]	0% (0/32)	0% (0/32)	0% (0/2)	100% (2/2)	0% (0/2)	0% (0/32)
others	NT	14.3% (4/28)	0.002 [± 0.002]	7.1% (2/28)	3.6% (1/28)	50% (2/4)	25% (1/4)	25% (1/4)	3.6% (1/28)
	T	5.9% (2/34)	0.001 [± 0.0002]	0% (0/34)	0% (0/34)	0% (0/2)	0% (0/2)	100% (2/2)	2.9% (1/34)

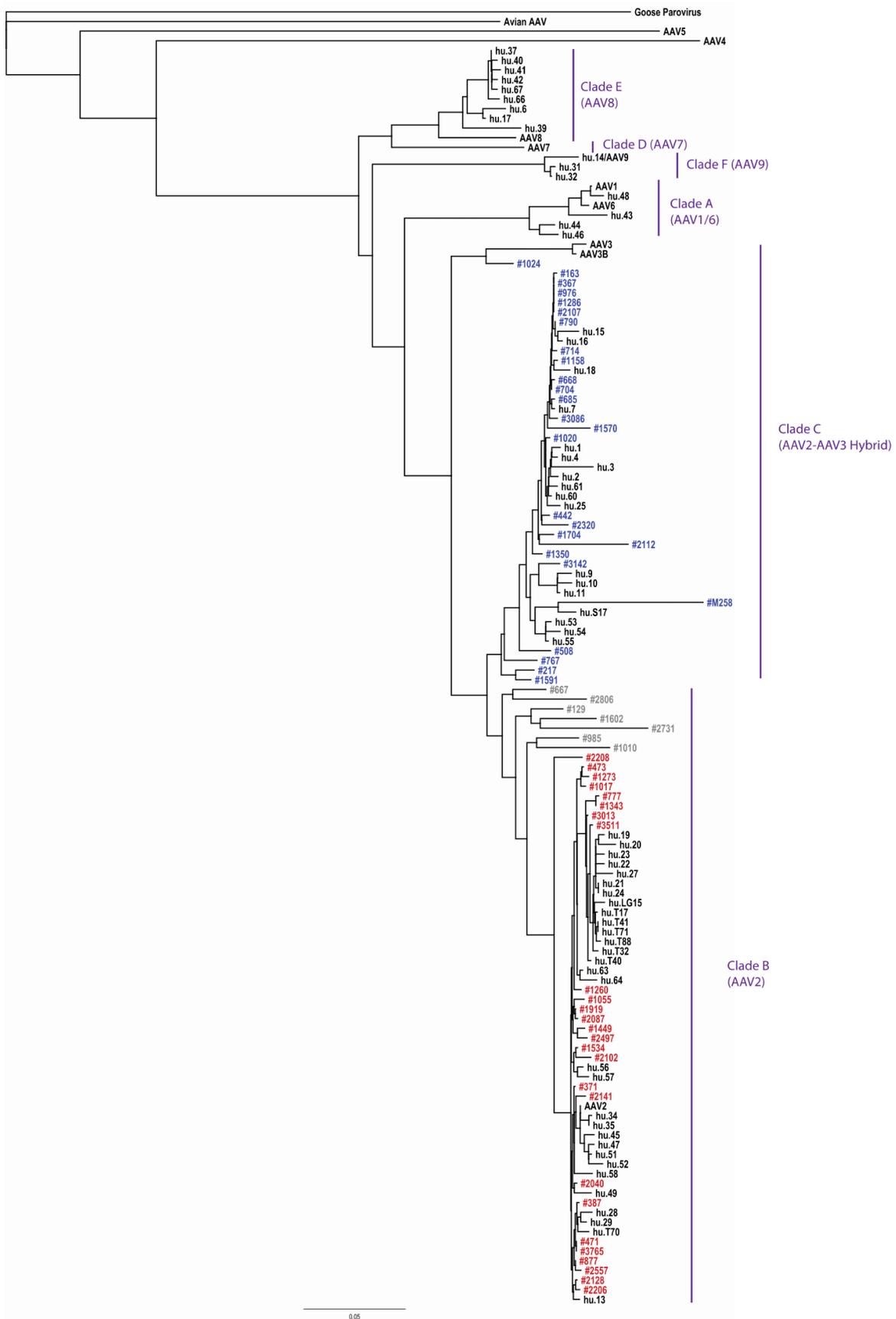
Supplementary table 4. List of AAV clonal inserted sites in liver tumors

Series	Sample	Human target gene	Virus	Breakpoint 1 on human genome (hg19)	Breakpoint 2 on human genome (hg19)	Breakpoint 1 on viral genome	Breakpoint 2 on viral genome	Orientation (5'>3')	Orientation (Flip/Flop)	Genotype	Publication
LiC1162	#2128T	CCNA2	AAV2	chr4:122,743,475	chr4:122,743,452	2934	1493	3>5	ND	AAV2	Nault <i>et al.</i> 2015
LiC1162	#313T	CCNA2	AAV2	chr4:122,742,417	chr4:122,742,409	2611	4585	5>3	Flop	AAV2	Nault <i>et al.</i> 2015
LiC1162	#2206T	CCNA2	AAV2	chr4:122,742,478	chr4:122,742,456	3560	4596	5>3	Flip	AAV2	Nault <i>et al.</i> 2015
LiC1162	#2848T	CCNA2	AAV2	chr4:122,742,948	chr4:122,742,929	4630	3730	3>5	Flip	AAV2/13	
LiC1162	#129T	CCNA2	AAV2	chr4:122,743,844	chr4:122,743,835	4597	4316	3>5	Flip	AAV2/13	Bayard <i>et al.</i> 2018
LiC1162	#2112T	CCNA2	AAV2	chr4:122,742,667	chr4:122,742,660	4389	4608	5>3	Flip	AAV2/13	Nault <i>et al.</i> 2015
TCGA	TCGA-G3-AAUZ	CCNA2	AAV2	ND	chr4:122,743,570	ND	4613	5>3	Flop	AAV2/13	Bayard <i>et al.</i> 2018
TCGA	TCGA-ZS-A9CF	CCNA2	AAV2	chr4:122,742,400	chr4:122,742,166	3407	4514	5>3	ND	AAV2	Bayard <i>et al.</i> 2018
TCGA	TCGA-BC-A10Y	CCNA2	AAV2	chr4:122,743,756	chr4:122,743,742	4064	4586	5>3	ND	AAV2/13	Bayard <i>et al.</i> 2018
LiC1162	#2557T, #2558T	TNFSF10	AAV2	chr3:172,224,150	chr3:172,302,191	4388	4597	5>3	Flop	AAV2	Nault <i>et al.</i> 2015
LiC1162	#1602T	TNFSF10	AAV2	chr3:172,224,027	chr3:172,224,026	4270	4571	5>3	Flop	AAV2/13	Nault <i>et al.</i> 2015
LiC1162	#1185T	KMT2B	AAV2	chr19:36,212,635	chr19:36,212,644	4602	4247	3>5	Flip	ND	Nault <i>et al.</i> 2015
LiC1162	#2141T	CCNE1	AAV2	chr19:30,287,313	chr19:30,287,316	4599	4379	3>5	Flop	AAV2	Nault <i>et al.</i> 2015
LiC1162	#1591T	CCNE1	AAV2	chr19:30,303,035	chr19:30,303,053	4340	4598	5>3	Flip	AAV2/13	Nault <i>et al.</i> 2015
LiC1162	#2208T	CCNE1	AAV2	chr19:30,304,532	chr19:30,304,542	4604	4237	3>5	Flop	AAV2	Nault <i>et al.</i> 2015
LiC1162	#M257T	CCNE1	AAV2	chr19:30,303,740	chr19:30,303,743	4590	4253	3>5	Flop	AAV2/13	Bayard <i>et al.</i> 2018
LiC1162	#3641T	CCNE1	AAV2	chr19:30,291,515	chr19:30,291,518	4239	4534	5>3	ND	ND	Bayard <i>et al.</i> 2018
TCGA	TCGA-BC-A10T	CCNE1	AAV2	chr19:30,302,864	ND	4626	ND	ND	ND	ND	Bayard <i>et al.</i> 2018
LiC1162	#985T	TERT	AAV2	chr5:1,295,308	chr5:1,295,291	4390	4597	5>3	Flip	AAV2/13	Nault <i>et al.</i> 2015
LiC1162	#2102T	TERT	AAV2	chr5:1,295,238	chr5:1,295,235	4288	4597	5>3	Flip	AAV2	
LiC1162	#1919T	GLI1	AAV2	chr12:57,856,202	chr12:57,856,230	4619	4108	3>5	Flip	AAV2	
LiC1162	#1920T	GLI1	AAV2	chr12:57,856,195	chr12:57,856,212	4202	4042	3>5	Flip	AAV2	
LiC1162	#1921T	GLI1	AAV2	chr12:57,854,317	chr12:57,854,407	3940	4679	3>5	Flip	AAV2	
LiC1162	#3765T	GLI1/INHBE	AAV2	chr12:57,850,368	chr12:57,850,370	4071	4679	3>5	Flip	AAV2	
LiC1162	#1920T	H2AFY2	AAV2	chr10:71,864,210	chr10:71,864,240	4579	3629	3>5	Flip	AAV2	
LiC1162	#1344T	IGR chr10	AAV2	chr10:61,519,721	chr10:61,520,465	4342	2447	3>5	ND	AAV2	
LiC1162	#2141T	NKAIN3	AAV2	chr8:63,681,461	chr8:63,681,478	4608	4570	3>5	Flop	AAV2	
ICGC-JP	HX032	CCNE1	AAV2	chr19:30,304,511	chr19:30,304,517	4600	4386	3>5	ND	ND	Fujimoto <i>et al.</i> , 2016
ICGC-JP	RK112	KMT2B	AAV2	chr19:36,213,603	chr19:36,213,955	3320	4438	5>3	ND	ND	Fujimoto <i>et al.</i> , 2016
ICGC-JP	RK236	IGR chr5	AAV2	chr5:18,813,512	chr5:18,813,398	82	2982	3>5	ND	ND	Fujimoto <i>et al.</i> , 2016

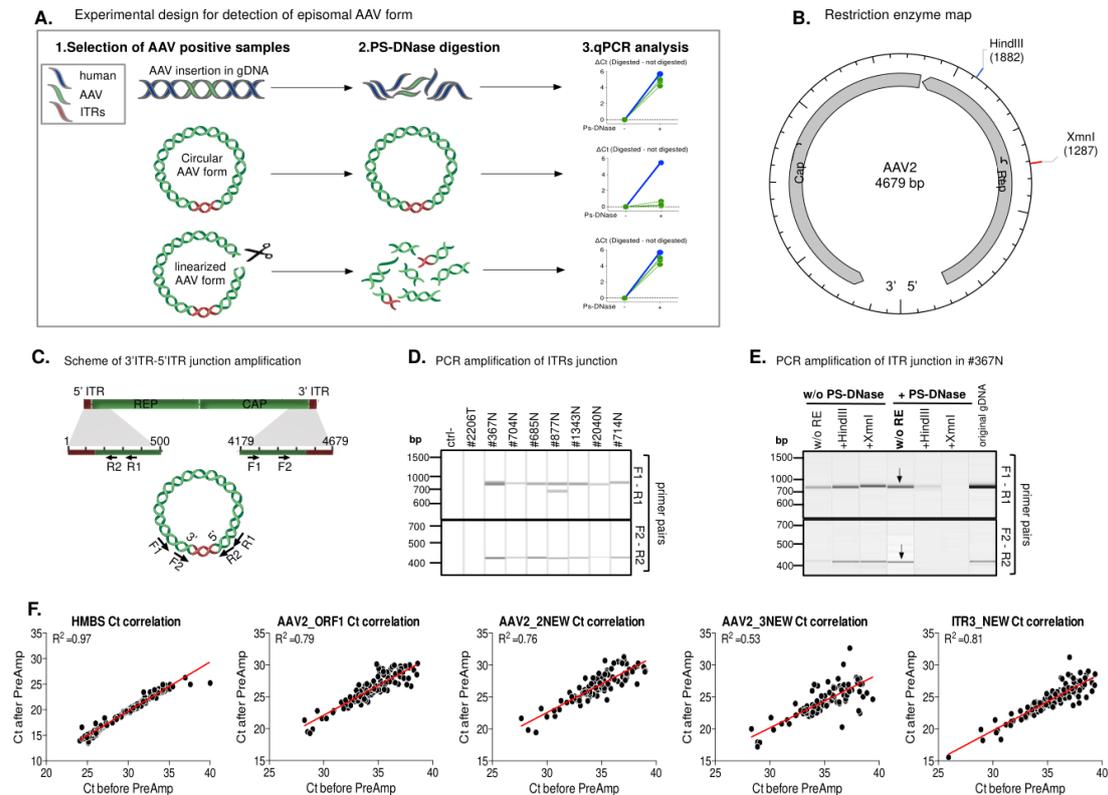
ND: not determined



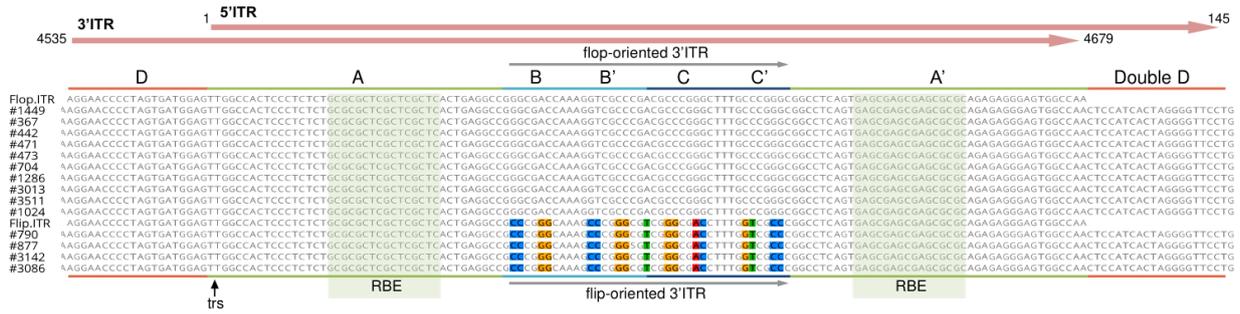
Supplementary figure 1. AAV detection using quantitative PCR and viral capture. A) Schematic representation of the 8 TaqMan probe sets used in qPCR screening mapped on AAV2 reference genome (NC_001401). The 5 genomic regions targeted by the probe sets are defined at the top of the scheme. The color of the probe sets indicates the specificity for the viral genotypes. B) Correlation between the number of AAV copies per cell determined using qPCR and the number of AAV reads in viral capture normalized on human reads.



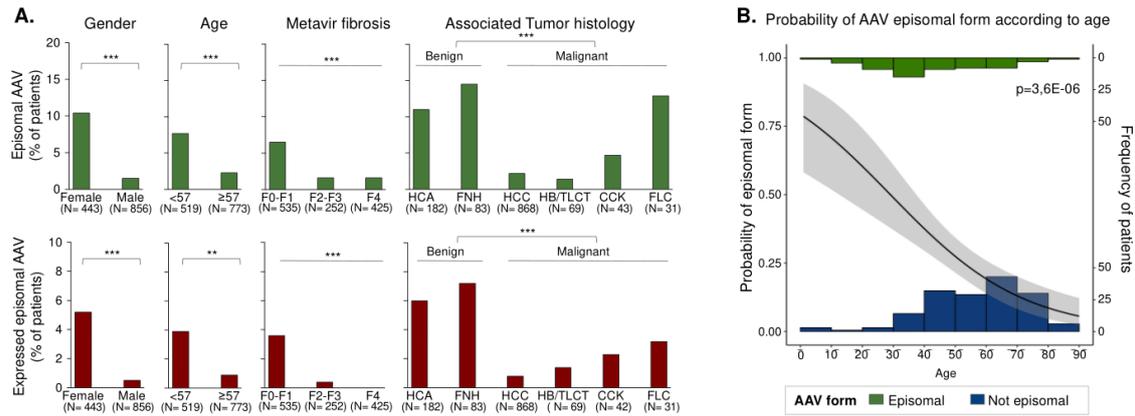
Supplementary figure 2. Phylogenetic tree on VP1 coding protein region. This was constructed with (1) in black 73 human AAVs, 1 avian AAV and a Goose Parvovirus, described in Chen et al.⁴², (2) all AAV sequences described therein that to the AAV2 group of sequences in red or (3) to the AAV2/13 group of sequences in blue. A neighbor-joining method on the basis of the Jules-Cantor model was used to derive phylogenetic distances based on 735 aa of VP1 sequence. Goose parvovirus was used as the outgroup. Previously described AAV clade nomenclature described in Gao et al.¹⁴ was adopted and organized by vertical brackets. The scale for genetic distance is indicated in the bottom left corner.



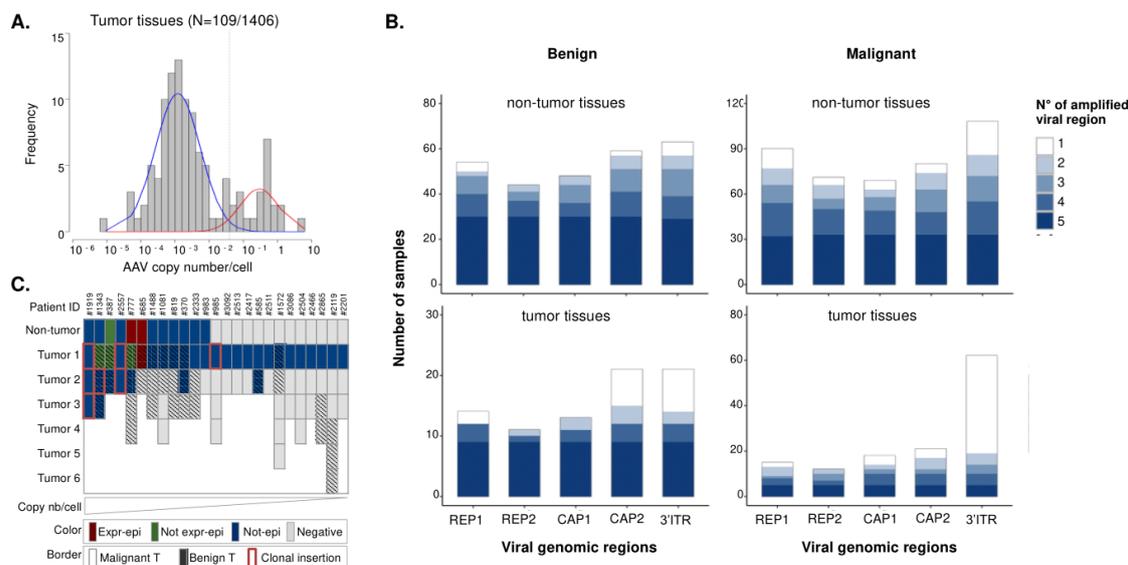
Supplementary figure 3. Viral episomal form investigation. A) Experimental design of the DNase/TaqMan assay for episomal form detection. The ΔCt (Digested – Not digested DNA) is used to define the presence or absence of episomal AAV form. An increased ΔCt indicates the digestion of AAV DNA due to the presence of an integrated not-episomal form (top), whereas a low ΔCt results from the protection of the viral DNA due to the presence of an episomal form (middle). An increased ΔCt is also observed as consequence of linearization of the episomal AAV form prior to DNase digestion (bottom). A probe set (HMBS) targeting the human genome (in blue) is used as control of DNase digestion efficiency. B) AAV map with positions of restriction enzymes (HindIII and XmnI) used to linearize viral episomal form. C) Schematic representation of the primers used to amplify the 3'ITR-5'ITR junction. Two different couples of primers were used: F1-R1 and F2-R2. D) Junction's amplification in 7 patients with confirmed viral episomal form (from line 3 to 9) using 2 different couples of primers. #2206 is used as negative control. E) PCR amplification of the 3'ITR-5'ITR junction before and after DNase digestion in patient #367. When the PCR is preceded by the linearization of episomal AAV, no amplification is detected after DNase digestion. The arrows indicate the 3'ITR-5'ITR junction in digested DNA without linearization of the episomal form. High-resolution capillary electrophoresis was performed using Qiaxcel with individual lanes of the same migration figured on the photo in D and E. F) Correlation between Ct value of HMBS and AAV probe sets before and after DNA pre-amplification.



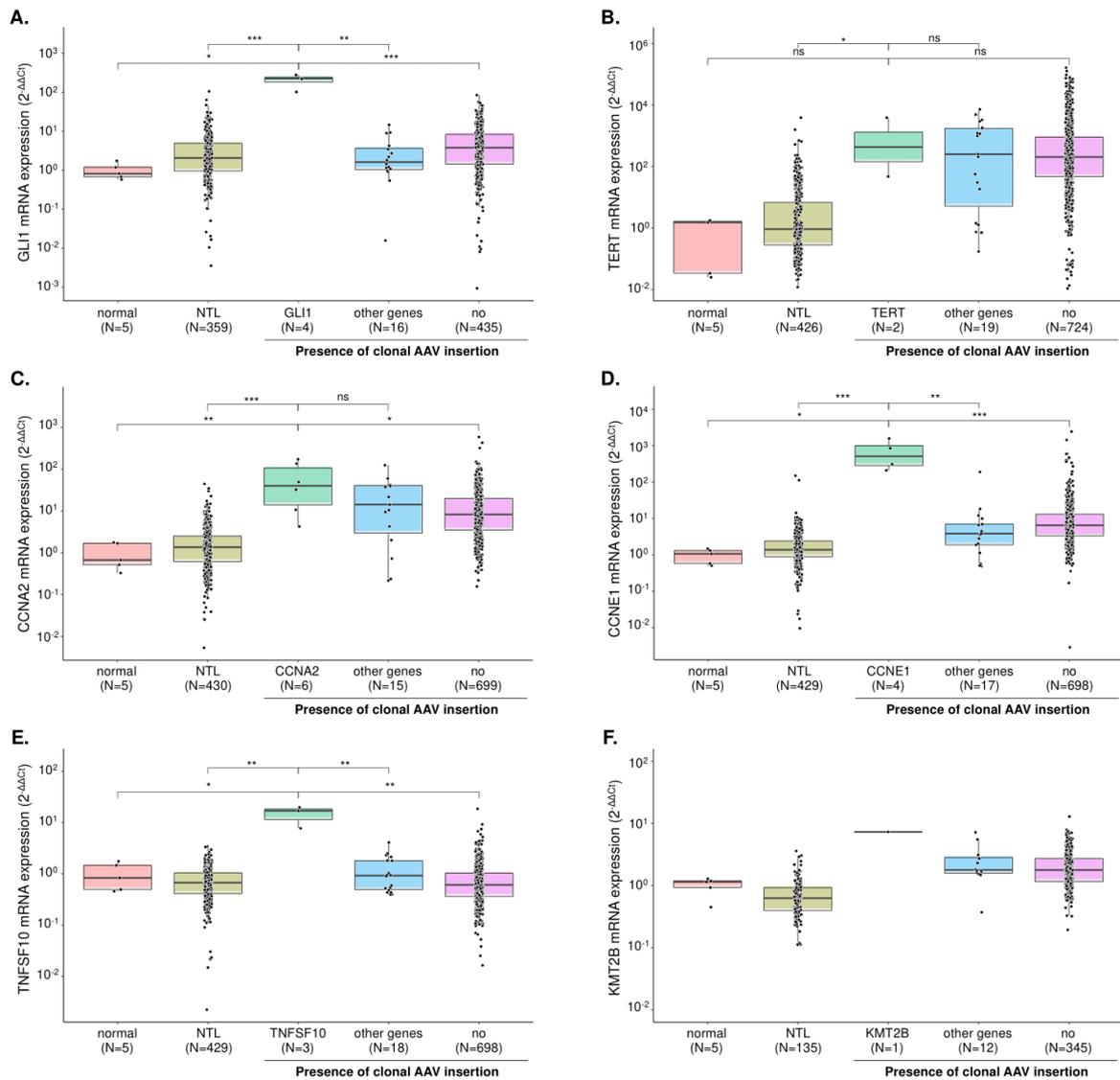
Supplementary figure 4. Viral 3' ITR-5' ITR episomal junction determined by *in silico* analysis in 14 patients. ITR is composed of seven regions: A, A', B, B', C, C' and D. The B-B' and C-C' regions exist in two palindromic configurations, "flip" (10 patients) and "flop" (4 patients) oriented DNA (grey arrows). Rep-binding element (RBE) motif is boxed; the arrow indicates the positions of the terminal resolution site (trs).



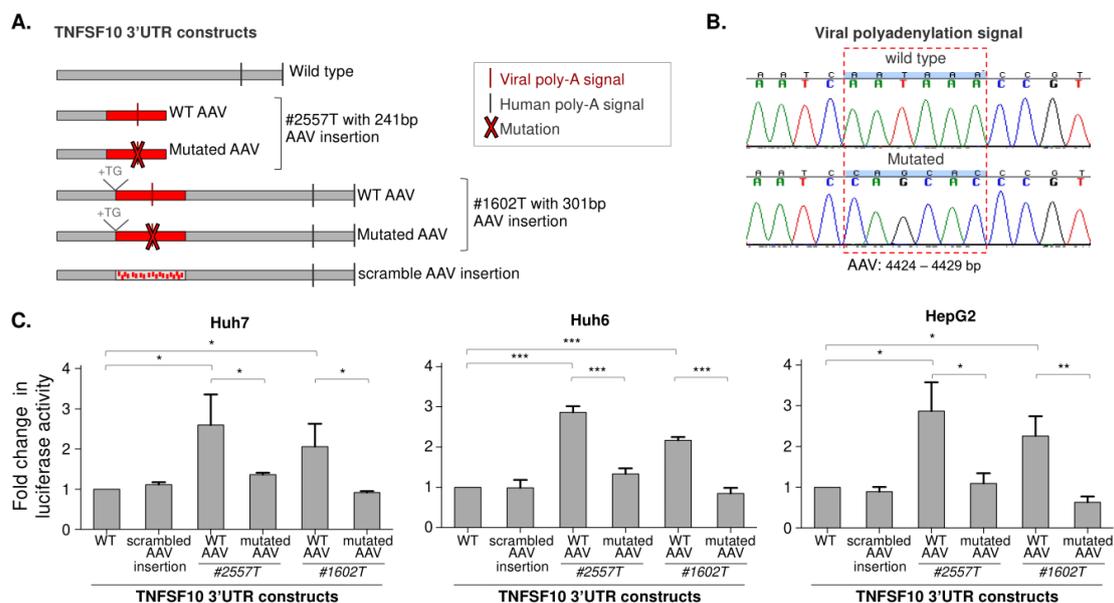
Supplementary figure 5. Correlation between AAV form and features of the patients. A) Frequency of episomal AAV (top) and episomal and expressed AAV form (bottom) according to gender, age, Metavir fibrosis score and associated tumor histology. The histological tumor types are grouped according to their malignancy. Frequency of AAV positive patients is displayed (χ^2 test with Monte Carlo simulation and χ^2 test for trend in proportions for Metavir score). ***P < 0.001; **P < 0.01; *P < 0.05. B) Frequency of episomal and not-episomal AAV form according to age of the patients. Logistic regression model was used to predict the probability to have an episomal form according to the age, where 1 indicates the highest probability and 0 the lowest. Regression curve of the probability with standard deviation is represented.



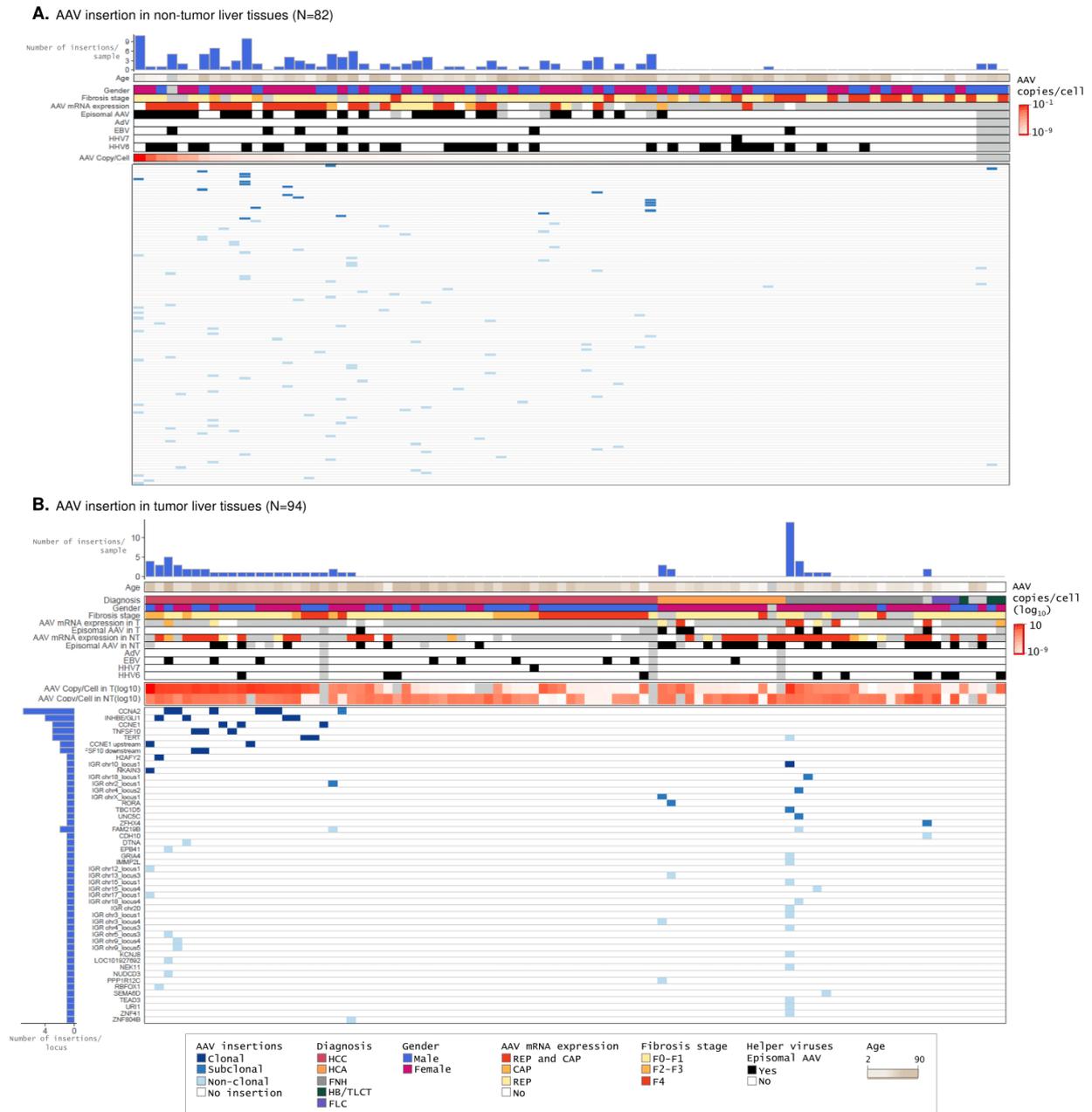
Supplementary figure 6. AAV presence in tumor tissues and comparison with non-tumor counterparts. A) Viral copy number/cell distribution in 109 tumor samples. The density line defines the low and high positivity groups in blue and red respectively. B) Number of amplified viral regions according to the 5 different genomic regions in non-tumor (top) and tumor tissues (bottom) of patients with benign (left) and malignant (right) tumors. C) Inter-tumor heterogeneity in patients with multiple nodules analyzed for AAV presence. Each column represents one patient. The 26 patients are ordered according to the copy number/cell in the non-tumor tissues. In case of negative non-tumor tissue, the order is determined by the copy number/cell in tumor. The color indicates the molecular form of the virus: episomal and expressed AAV (Expr-epi; red), not-expressed episomal AAV (Not expr-epi; green), not-episomal (Not-epi; blue) and negative (grey). The border and the pattern indicates the tumor type (malignant or benign) and the presence of clonal insertions.



Supplementary figure 7. Impact of clonal AAV integration on the expression of the target oncogenes: GLI1 (A), TERT (B), CCNA2 (C), CCNE1 (D), TNFSF10 (E) and KMT2B (F). The level of expression was assayed using qPCR in HCC with AAV insertions, within the tested gene and other target genes, in comparison to HCC without AAV insertion and non-tumor liver (NTL) tissues. Gene expression is presented relative to the expression in normal liver tissue on the y axis (\log_{10}). The black line in each boxplot corresponds to the mean values. A significant difference in expression was defined by Wilcoxon rank-sum test: *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$.



Supplementary figure 8. Viral polyadenylation signal usage in TNFSF10. A) TNFSF10 3'UTR constructs into a pmirGLO vector controlling the expression of Firefly luciferase gene under PGK promoter. Site of insertions with #2557T and #1602T AAV sequences (in red), with and without mutated viral poly-A signal, or with the scrambled viral sequence are represented. B) Sanger validation of the mutations in viral poly-A signal. Nucleotides position refers to AAV2 (NC_001401). C) Impact of AAV integration in the 3'UTR of TNFSF10 in presence of wild type and mutated viral poly-A signal was evaluated using luciferase assays in Huh7, Huh6 and HepG2 liver cell lines. The 3'UTR of TNFSF10 with either of the two AAV insertions identified in #2557T and #1602T or a scrambled sequence cloned into the pmirGLO vector was compared to vector encoding the wild-type 3'UTR (WT). Error bars, s.d. of triplicate experiments corresponding to three independent transfections for each plasmid in each cell line. t tests were performed; ***P < 0.001; **P < 0.01; *P < 0.05.



Supplementary figure 9. Description of AAV insertions in non-tumor (A) and tumor liver tissues (B). AAV insertions are represented according to the type (clonal, subclonal and non-clonal). Number of samples refers to multi-samples. Annotations related to patients features, AAV presence and helper virus infections are reported. The histograms indicate the number of AAV insertion per sample (on the top) and the number of AAV insertion per genomic locus (on the left). The latter histogram is not shown for in the non-tumor tissues graph because each locus presents only unique insertion event.

Supplementary Materials and Methods

Computational analysis of human-AAV sequences

Paired-end read sets, for which samples we have confirmed the presence of the virus, were further considered to assemble the whole viral genome using an overlap layout consensus (OLC) assembling strategy. Both circular and linear contigs were generated with the Genious Pro 11.1.5 (<http://www.geneious.com>, Biomatters Ltd, Auckland, New Zealand) using the de novo assembler with high sensitivity threshold, considering a minimum overlap of 25 bases and 80% of identity, a maximum gap size of 5 bases, less than 20% gaps per read, and maximum ambiguity of 16 bases. Consensus sequences were generated, based on the major base distribution, and confirmed on reads alignment. Sequences have been deposited in the Genbank database.

From all sequence contigs, VP1 open reading frames were used to segregate our human isolates into individual molecular forms (or clades) as described in Chen et al.¹ The coding protein region was derived from the nucleotide sequences. A neighbor-joining method² on the basis of the Jules-Cantor model was used to derive phylogenetic distances based on 735 amino acids of VP1 sequence from our 58 human isolates and 73 human AAV plus 1 avian AAV. Goose parvovirus was used as the outgroup. The whole circular and linear AAV DNA structures (4679 bp) from respectively 14 and 57 isolates were compared using multi-alignment with the ClustalW algorithm³ using default parameter. A Neighbor-Joining method² on the basis of the Tamura-Nei method⁴ was used to derive phylogenetic distances based on the nucleotides sequences. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Evolutionary analyses were conducted in Genious Pro 11.1.5 software using Treebuilder. Sequence variations with wild-type AAV2 sequences and among human AAV isolates were identified considering a minimum variant frequency of 20%. The approximate *p*-value represents the probability of a sequencing error, which is calculated using the binomial distribution, resulting in observing bases with at least the given sum of qualities.

The assessment of the circular structures of AAV on episomal junction, allowed us to define the overlapping region of 3' and 5' inverted terminal repeats (ITRs) and determine the existence of a flip or flop sequence in each human isolate.

Experimental procedure to search for episomal forms of AAV

240ng of genomic DNA were digested by incubation with 10U of Plasmid-Safe DNase (Epicentre) in a final volume of 16µL for 16h at 37°C, followed by heat inactivation of the nuclease. 20ng of DNase-digested and not-digested control DNA were pre-amplified by TaqMan PreAmp Master Mix (Applied Biosystem) using 14 preamplification PCR cycles. Real time quantitative PCR was performed on 10ng of DNA before preamplification and 5µL of preamplification product (diluted 1:10 in TE buffer) using custom made AAV probes (Supplementary Table 2) and AB 7900HT Fast Real-Time PCR System (Applied Biosystem). Expression data (Ct values) were acquired using SDS Software (v2.3). The efficiency of the nuclease digestion was determined using HMBS probe as control of the genomic DNA digestion. The difference between AAV Ct without and with PS-DNase digestion allowed to determine the presence of viral episomal form, or the digestion of the linear viral DNA.

In order to validate the results, an additional step of linearization of the viral genome was added before PS-DNase digestion. 600ng of gDNA were first digested in a double enzymatic digestion (1h using 12U of enzyme and 1h with additional 10U at 37°C) with restriction enzyme (XmnI or HindIII from

New England Biolabs) that cuts only one site within AAV sequence. A specific probe set (Supplementary Table 2) that overlapped the cutting site was used to check the efficiency of the digestion. Digested DNA was tested for the presence of episomal form with the DNase/TaqMan based assay previously described (Supplementary Figure 3).

A first set of 10 samples including 8 tissues with at least 4 amplified genomic regions and 2 controls positive only for 2 regions was tested and used to set up the protocol. In addition, a single cutter restriction enzyme was used to linearize the viral DNA prior to DNase digestion. The combination of the DNase digestion with or without the previous linearization of the viral DNA allowed to determine the presence of an episomal form according to the final TaqMan results on digested DNA. If the virus was in the episomal form, the DNase was not able to digest the viral genome and the sample was still positive for the virus in the TaqMan assay. Conversely, if the virus was in linear form, the viral DNA was digested giving a negative TaqMan results. We found the presence of episomal form in 7/8 non-tumor tissues and none of the controls was positive. In all the cases we observed a digestion of the linearized AAV DNA demonstrated by the increased Ct value.

The sample #367N was tested for 3'ITR-5'ITR junction amplification after DNase treatment with or without previous linearization of the viral genome. Interestingly, the ITR junction was protected from the DNase treatment showing that the 3'ITR-5'ITR junction belonged to the circular AAV episomal form (Supplementary Figure 3E).

Sure of the reliability of the technique, we enlarged the investigation of viral episome in all non-tumor samples with 3 or more amplified genomic regions, which represented the 58% of AAV positive non-tumor samples. A step of pre-amplification of viral DNA after DNase digestion was added in order to assess the episomal status even in case of low number of viral copies. Ct values of each probe set before and after pre-amplification were strongly correlated (Supplementary Figure 3F), therefore qPCR results after pre-amplification were used to determine the episomal status.

References Supplementary Materials and Methods

1. Chen CL, Jensen RL, Schnepf BC, et al. Molecular characterization of adeno-associated viruses infecting children. *J Virol* 2005;**79**:14781-92.
2. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987;**4**:406-25.
3. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. *Bioinformatics* 2007;**23**:2947-8.
4. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 1993;**10**:512-26.

DISCUSSION

I. Rôle de l'intégration virale dans l'infection et la carcinogénèse

Le rôle joué par les intégrations du VHB dans l'histoire naturelle de l'infection et dans la carcinogénèse reste encore très controversé. Premièrement, elles ne font pas partie du cycle de réplication viral (Valaydon and Locarnini, 2017) et ne semblent donc pas avoir de fonction directe dans l'établissement de l'infection. Deuxièmement, les intégrations ayant lieu dans les heures suivant l'infection (Tu et al., 2018) et les tumeurs se développant des années plus tard, il apparaît comme surprenant qu'elles puissent être directement à l'origine de l'initiation tumorale. La question de la fonction de ces intégrations est donc un sujet central de la recherche sur le VHB, et les travaux réalisés dans le cadre de cette thèse ont cherché à apporter des éléments de réponse.

A. Intégration ou réplication virales : deux destins incompatibles ?

Une contradiction étonnante liée aux intégrations du VHB réside dans le fait que l'ADN viral sous forme intégrée a perdu sa capacité de réplication, et l'intégration constitue donc a priori une impasse pour le virus. Notre étude a en effet confirmé que les formes intégrées dans les tissus non-tumoraux ont une taille et des extrémités correspondant à la forme virale linéaire double-brin (ou une forme intermédiaire ou épissée ayant les mêmes extrémités) (Article 1, Fig. 2e) et elles ne peuvent donc pas servir de matrice pour la transcription de l'ARNpg permettant la réplication. Cependant, l'intégration étant un mécanisme commun à l'ensemble des *hépadnavirus* et donc conservé dans l'évolution (Hu and Seeger, 2015b), il est peu probable que cela diminue les capacités globales (« fitness ») du virus. Cette observation suggère donc que les intégrations jouent un rôle dans l'histoire naturelle de l'infection, et deux hypothèses sont envisageables : les intégrations influent indirectement sur la réplication virale, ou modulent les conséquences de l'infection en agissant sur la réponse immunitaire.

Les échantillons non-tumoraux que nous avons étudiés et pour lesquels nous avons caractérisé les formes virales sont exclusivement issus de patients ayant développé un CHC et ne sont donc a priori pas représentatif des premiers stades de l'infection. En effet, selon les données cliniques, seulement 88% des patients de la série que nous

avons analysée en capture virale ont un statut négatif pour l'AgHBe, ces patients sont donc majoritairement à un stade avancé de l'infection B chronique (Article 1, Supp Table 1). Etudier et décrire les formes d'ADN viral dans ces tissus apporte néanmoins des connaissances complémentaires pouvant aider à mieux décrire la dynamique des intégrations. Dans notre étude de capture virale sur 170 échantillons non-tumoraux, nous avons notamment mis en évidence que le nombre d'intégrations était plus élevé dans les tissus hépatiques contenant de l'ADN viral réplcatif (Article 1, Fig. 1c), révélant donc l'existence d'une association entre intégration et réplication virales.

Tout d'abord, il semble attendu que la réplication virale augmente directement le processus d'intégration via une augmentation du nombre d'hépatocytes infectées par des virions contenant de l'ADNdsI et/ou via une augmentation de la quantité d'ADNdsI nouvellement synthétisé à partir de l'ARNpg. Des études *in vitro* et *in vivo* ont cependant montré que le taux d'intégration était de 10^{-3} événements par cellule et qu'il était impacté par l'inhibition de l'entrée des virions dans les cellules mais pas par la suppression de la réplication virale, indiquant que dans le modèle utilisé, les formes intégrées détectées provenaient majoritairement des virions entrés initialement et ne dépendaient pas de la réplication pendant les premières semaines après infection (Budzinska et al., 2018a; Tu et al., 2018). De plus, des travaux sur le virus DHBV chez le canard ont montré qu'une superinfection par le DHBV était impossible, suggérant un phénomène d'exclusion de la superinfection médiée par la protéine L-HBs pour l'ensemble des *hépadnavirus* même si aucune étude n'a été publiée pour le VHB à ce jour (Walters et al., 2004). Un même hépatocyte ne pourrait donc a priori ne contenir qu'une seule intégration tant qu'il contient de l'ADNccc et qu'il synthétise L-HBs.

Dans les échantillons non-tumoraux de notre série, le nombre de copies d'ADN viral par cellule est pourtant compris entre 10^{-3} et 100, suggérant donc la présence de plusieurs intégrations du génome viral au sein d'un même hépatocyte. Le fait d'avoir défini précisément la clonalité des intégrations dans notre série nous a permis de confirmer que 5 échantillons non-tumoraux sur les 21 ayant des intégrations clonales en contenaient plus d'une. Cela suggère donc que les intégrations s'accumulent au cours du temps dans les hépatocytes de la même façon que les mutations, potentiellement sur des durées longues, et ce processus est nécessairement favorisé par la persistance d'une infection active dans le foie. Il serait alors intéressant de corrélérer le nombre d'intégrations au cours du temps chez des patients pendant les

premières phases après infection. Dans notre cohorte, les patients ayant le plus grand nombre d'intégrations dans leur foie non-tumoral sont en réalité parmi les patients les plus jeunes, mais étant donné qu'ils ont pu être infectés dès leur naissance, il est possible qu'ils aient néanmoins subi un temps de réplication active du virus très long.

D'un autre côté, une revue sur les intégrations virales a récemment suggéré que les séquences intégrées pourraient favoriser la réplication virale via la synthèse de la protéine HBx. En effet, comme cela a été décrit dans la partie III.A.3 de l'Introduction, HBx est nécessaire à l'initiation et à la maintenance de la réplication virale. L'expression de protéines HBx a été détectée dans le foie de plus de 50% de patients infectés par le VHB ayant une cirrhose, bien que cette expression soit faible par rapport à celles des autres protéines virales (Slagle and Bouchard, 2018; Su et al., 1998; Wei et al., 2010). Des protéines transcrites à partir de séquences intégrées pourraient donc a priori jouer un rôle dans la persistance de l'infection. Néanmoins, un précédent projet du laboratoire a mis en évidence que 33% des échantillons non-tumoraux de patients infectés par le VHB contiennent des mutations inactivatrices du gène HBx (Amaddeo et al., 2015), et notre étude a montré que dans 28% des tissus non-tumoraux exprimant un transcrite HBx, ce dernier est tronqué. Dans notre série, le niveau d'expression de l'ARN HBx est comparable dans les échantillons contenant des intégrations sous-clonales ou clonales en absence de réplication et dans les échantillons ayant de l'ADN répliatif pour une sonde située en 5' du gène HBx mais est bien inférieur pour une sonde située en 3', indiquant que les formes intégrées permettent bien la synthèse de transcrits HBx en absence de réplication mais en partie tronqués (Figure 29).

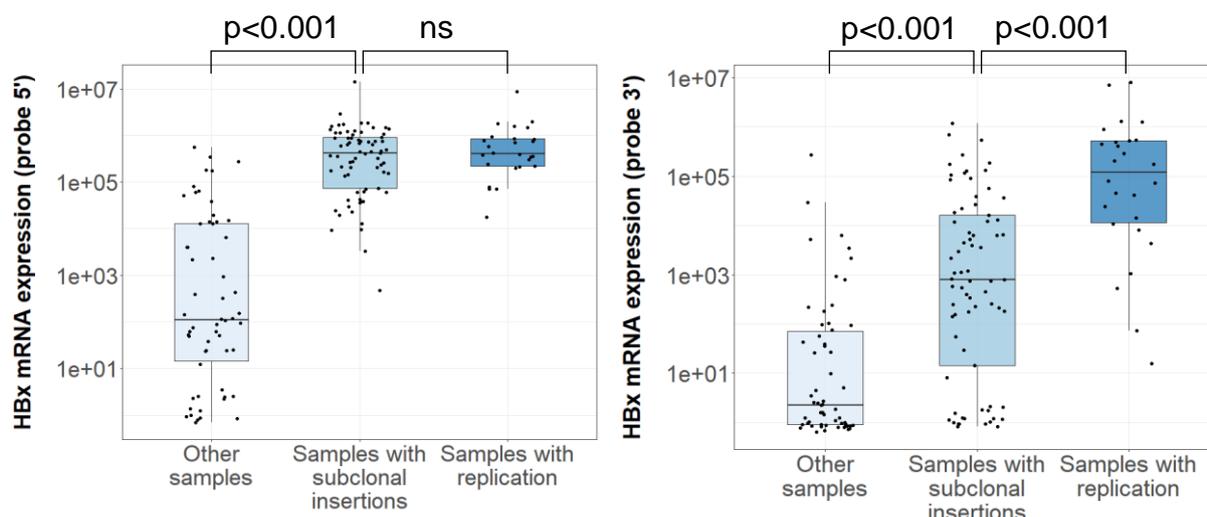


Figure 29 : Niveau d'expression des transcrits HBx en 5' (gauche) ou 3' (droite) dans 157 tissus non-tumoraux de notre série analysée en Capture.

Bien qu'une étude récente ait montré que les variants d'HBx peuvent avoir conservé leur fonctionnalité (Rivière et al., 2019), des travaux plus approfondis sur l'expression protéique d'HBx dans des tissus non-tumoraux sont nécessaires pour conclure sur son réel impact sur la réplication. La présence d'une forme virale répliquative restant minoritaire parmi les échantillons non-tumoraux de notre série contenant des formes intégrées et ayant un transcrit HBx non tronqué (28%, 26 sur 93), l'effet potentiel causé par la protéine sur la réplication n'est dans tous les cas pas suffisant pour l'initier.

Comprendre le lien entre intégration et réplication virales représente un enjeu primordial pour le traitement de l'infection par le VHB et la surveillance de la survenue d'un CHC. En effet, les traitements antiviraux actuels ne permettent pas l'élimination totale du virus dans le foie des patients mais suppriment ou réduisent très fortement ses capacités répliquatives. Si le nombre d'intégration est directement lié à la durée pendant laquelle le virus est répliquatif, il est impératif de traiter les patients le plus précocement possible pour limiter intégrations et conséquences oncogéniques.

B. Les séquences intégrées du VHB face à la réponse immunitaire

La deuxième hypothèse concernant le rôle joué par les intégrations virales est liée à l'interaction entre virus et réponse immunitaire. Les séquences intégrées pourraient en effet contribuer à l'atténuation de la réponse immunitaire en induisant une forte expression d'AgHBs. La synthèse d'AgHBs par des séquences intégrées du virus en absence de réplication a déjà été identifiée dans un modèle d'infection chronique chez le chimpanzé (Wooddell et al., 2017) et dans des lignées cellulaires infectées par le virus HDV (Freitas et al., 2014). Par ailleurs, de nombreuses études ont prouvé qu'un niveau élevé d'AgHBs peut avoir un rôle immunosuppresseur sur l'immunité innée et adaptative contre le VHB (Kondo et al., 2013; Ma et al., 2019; Tsai et al., 2018). Les différents mécanismes sont décrits ci-dessous et illustrés sur la Figure 30 :

- (1-2) l'épuisement des réponses des cellules T et des cellules B spécifiques à l'AgHBs faisant suite à une stimulation continue par les AgHBs, qui se caractérisent par une surexpression de *PD-1*, de *CTLA-4* et de *TIM3*
- (3-4) l'atténuation de l'immunité innée via l'inhibition de la signalisation médiée par les TLR et l'induction de l'interleukine-10 (cytokine anti-inflammatoire) dans les cellules endothéliales sinusoidales (LSEC) et les cellules de Kupffer (KC).

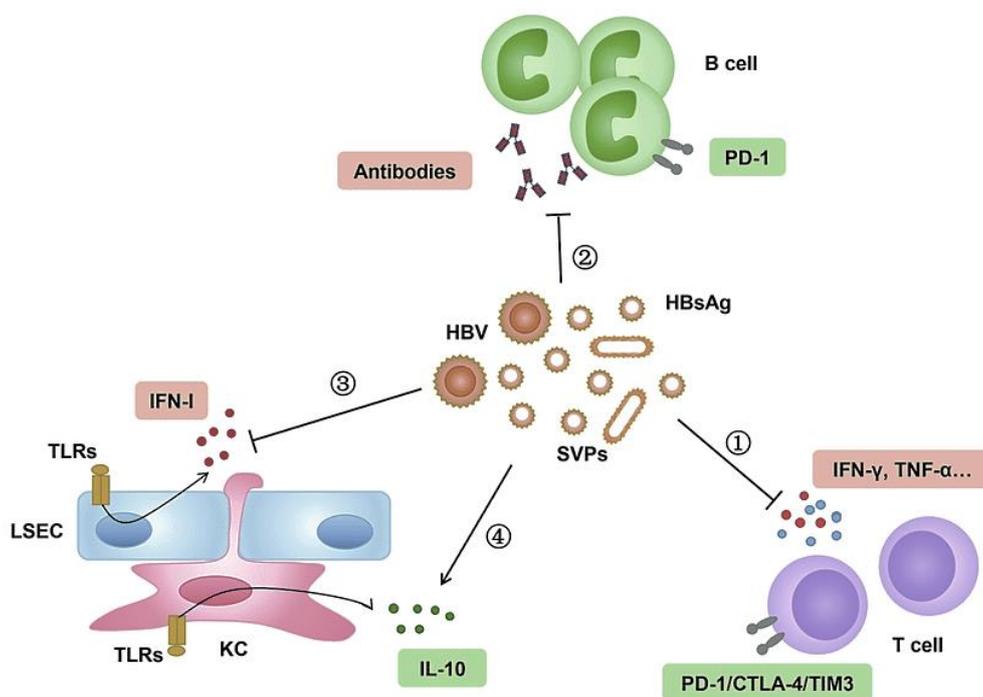


Figure 30 : Mécanismes de suppression des réponses immunitaires innées et adaptatives par les AgHBVs. (Ma et al., 2019).

La présence de séquences intégrées dans les tissus non-tumoraux générerait donc ainsi une production d'AgHBVs plus importante qui impacterait la réponse immunitaire. Notre étude a confirmé l'expression de transcrits préS2-S complets dans la grande majorité des échantillons de notre série (74%), bien que le niveau d'expression de l'ARN préS2-S générant l'AgHBVs soit un peu moins élevé dans les échantillons contenant des intégrations sous-clonales ou clonales en absence de réplication que dans les échantillons ayant de l'ADN réplcatif (Figure 31).

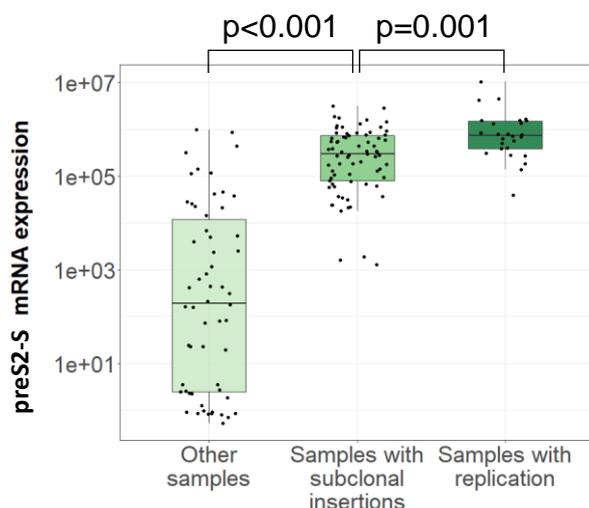


Figure 31 : Niveau d'expression des ARNm de la région préS2 dans 157 tissus non-tumoraux de notre série analysée en Capture.

Le laboratoire a également montré précédemment la présence de mutations dans le gène HBs (Amaddeo et al., 2015) et ces mutations pourraient potentiellement favoriser un échappement immunitaire (Lazarevic, 2014), même si cette hypothèse reste à démontrer. Une autre analyse intéressante serait d'analyser le profil immunologique des échantillons de notre série en fonction de la présence ou de l'expression du transcrite préS2-S. Une analyse préliminaire basée sur les score immunologique (méthode MCP-counter (microenvironment cell population) adaptée au RNA-Seq ; Becht et al., 2016) montre que les tissus exprimant des transcrits du VHB semblent en effet contenir moins de cellules B et T mais l'analyse est limitée par le nombre d'échantillons (n=26) et il serait important de prendre en compte les caractéristiques histologiques du foie non-tumoral.

Enfin, de par leur structure, les intégrations du VHB peuvent générer des AgHBs immunosuppresseifs mais pas d'AgHBe qui sont ciblés par les anticorps anti-HBe après séroconversion. Les séquences intégrées donneraient ainsi un avantage sélectif aux cellules contenant des séquences intégrées par rapport aux hépatocytes ayant de l'ADN viral répliquatif en échappant à la réaction immunitaire face au VHB.

La présence d'insertions clonales dans 11% des tissus non-tumoraux de notre série nous a amenés à approcher cette notion d'avantage sélectif. Ces échantillons contenaient tous un degré significatif d'atteinte hépatique (Score METAVIR F2-F4), et l'analyse des gènes dérégulés dans ces tissus a montré une diminution de l'expression des gènes liés à la réponse inflammatoire ainsi que des gènes liés à la voie de signalisation TNF α /NF κ B (Article 1, Suppl Fig. 3c). Ces résultats suggèrent donc que la pression de sélection exercée par la réponse immunitaire envers les cellules infectées favorise la prolifération locale de cellules échappant au système immunitaire. Les sites d'insertion clonaux dans le génome humain des hépatocytes non-tumoraux étant très variables et différents de ceux observés dans les tumeurs, ces résultats suggèrent que les intégrations dans ces échantillons n'ont pas de conséquences fonctionnelles directes impliquées dans la sélection, en ce qui concerne l'altération du génome humain. On peut cependant remarquer que 88% d'entre eux (14/16) contenaient des transcrits préS1 pouvant générer des AgHBs et 19% (3/16) des transcrits preC pouvant générer des AgHBe avec la double-mutation du Promoteur Basal du Core (PBC). Il est donc envisageable que les séquences intégrées jouent un

rôle indirect dans l'expansion clonale de ces hépatocytes en influant sur la réponse immunitaire via la surexpression d'AgHBs et l'absence d'AgHBe non mutés.

Au final, les intégrations du VHB dans les tissus non-tumoraux peuvent atténuer la réponse immunitaire de façon active, et la sélection clonale de cellules présentant une intégration dans le foie non-tumoral témoigne indirectement de cet échappement immunitaire, pouvant être favorisé par l'intégration.

C. Intégration et réplication du VHB par rapport à la carcinogénèse

L'identification d'intégrations virales dès la phase aiguë de l'infection par le VHB et l'augmentation de leur nombre au cours de l'ensemble des différentes phases de l'infection chronique a souligné le caractère commun de ce processus dans l'histoire naturelle de l'infection et a confirmé l'accumulation de ces intégrations au cours du temps dans les tissus hépatiques (Kimbi, 2005; Mason et al., 2016; Yang et al., 2018). Ces intégrations pouvant participer activement ou passivement à une expansion clonale, il serait possible qu'elles soient également « driver » dans la carcinogénèse.

L'ensemble des études d'intégration du VHB a mis en évidence des profils différents entre les intégrations dans les tissus tumoraux et les tissus non-tumoraux (Furuta et al., 2018; Sung et al., 2012; Yoo et al., 2017; Zhao et al., 2016). Curieusement, nous n'avons pas trouvé de forte association entre le nombre d'intégrations identifiées dans les tumeurs et les caractéristiques des génomes viraux dans le foie non-tumoral adjacent, suggérant que tumeurs et tissus non-tumoraux suivent des processus d'évolution différents et potentiellement indépendants. De plus, dans notre série analysée en capture, aucun gène contenant des insertions clonales dans les N n'est retrouvé parmi les gènes contenant des insertions clonales dans les T, indiquant que les processus d'expansion clonale sont différents et reposent sur des mécanismes a priori exclusifs dans ces deux types de tissus. Le lien entre intégration, réplication et expansion clonale ayant été discuté pour les tissus non-tumoraux dans les paragraphes précédents, la même question se pose pour les tissus tumoraux.

Une majorité des CHC liés à une infection par le VHB contiennent des intégrations (entre 51% et 100% en fonction des études ; les chiffres sont détaillés dans la Table 3

de la partie III.B.2 de l'Introduction). Notre étude ayant défini précisément la clonalité de ces événements, nous avons pu identifier la présence d'intégrations clonales dans 82% des 177 CHC de notre série et sans association particulière avec le degré d'atteintes hépatiques (Article 1, Fig. 2b). Les cellules infectées diminuant au cours de l'infection chronique du VHB (Mason et al., 2016), cette forte proportion de CHC contenant des insertions clonales chez les patients infectés suggère que ces intégrations jouent un rôle direct ou indirect dans la carcinogénèse.

Par ailleurs, nous avons identifié des formes répliquatives dans seulement 7% des échantillons, et sans nécessairement détecter de formes répliquatives dans le foie non-tumoral associé (Article 1, Suppl Fig. 4a). Deux éléments semblent ainsi indiquer que la réplication virale n'est pas directement impliquée dans le développement tumoral :

- Les différents variants du VHB ayant été décrits comme favorisant la réplication virale et la persistance de l'infection (partie II.C.2 de l'Introduction) sont contre-sélectionnés dans les tumeurs (Article 1, Suppl Fig. 5f)
- Les génotypes identifiés diffèrent entre le génome et le transcriptome dans des tumeurs contenant des formes répliquatives, suggérant que les séquences clonales intégrées et les séquences répliquatives ne proviennent pas de la même souche virale, les intégrations se seraient produites précocement alors que la réplication pourrait être issue d'une nouvelle infection plus récente

L'initiation tumorale ferait donc suite à l'apparition d'altérations « driver » dans des hépatocytes contenant des intégrations en absence de réplication. Il faut cependant préciser que la persistance de la réplication virale augmentant le degré d'atteintes hépatiques, cela impacte nécessairement de façon indirecte la carcinogénèse en accélérant l'accumulation d'altérations (mutations et intégrations).

Au final, étant donné le nombre d'événements d'intégration identifiés dans les tumeurs, la majorité d'entre eux sont a priori « passager » mais très schématiquement, le rôle joué par les intégrations dans la carcinogénèse peut être décrit selon deux scénarii :

- L'intégration du VHB à un endroit précis du génome induit ou accentue une altération fortement oncogénique directement à l'origine de la transformation maligne et déclenche la survenue d'un CHC agressif sur foie non-cirrhotique

- L'accumulation d'intégrations et de mutations dans un contexte d'inflammation chronique accentué par une réplication active du VHB donne un avantage sélectif à l'origine de la progression tumorale

Ces scénarii vont alors être discutés dans la partie suivante.

II. Conséquences oncogéniques des intégrations et diversité des CHC liés au VHB

Comme cela a été détaillé dans la partie III.B.3 de l'Introduction, trois mécanismes oncogéniques liés aux intégrations du VHB ont été décrits :

- La mutagénèse insertionnelle altérant des gènes associés au CHC
- L'induction d'une instabilité chromosomique
- L'expression de protéines virales à partir des séquences intégrées

La grande majorité des études portant sur les intégrations du VHB à partir de séquençage nouvelle génération s'est concentrée sur le premier mécanisme en décrivant les conséquences fonctionnelles des insertions virales à des sites récurrents tels que dans le promoteur de *TERT* (Li et al., 2019; Sze et al., 2020) ou dans le gène *KMT2B* (Dong et al., 2015; Saigo et al., 2008). Dans le cadre de cette thèse, nous avons confirmé l'impact joué par la mutagénèse insertionnelle du VHB sur la carcinogénèse mais nous avons également abordé le lien entre les intégrations du virus et l'instabilité chromosomique, qui n'avait été que brièvement décrit dans la littérature (Sung et al., 2012; Totoki et al., 2014; Zapatka et al., 2020). Nous n'avons cependant pas analysé les conséquences des intégrations sur les protéines virales, qui nécessiteraient de réaliser une analyse plus spécifique des échantillons tumoraux de notre série par la suite.

A. Intégrations du VHB et instabilité chromosomique

Une étude génomique des CHC liés au VHB réalisée au laboratoire a montré qu'une caractéristique importante de ces tumeurs est leur forte instabilité chromosomique. En

effet, au sein de la classification transcriptomique G1-G6, les CHC liés au VHB font essentiellement partie des groupes G1-G3 rassemblant les CHC prolifératifs (Amaddeo et al., 2015; Boyault et al., 2007; Rebouissou and Nault, 2020). Cette instabilité est directement associée à la présence d'altérations du gène *TP53* dans ces CHC, un gène suppresseur de tumeur pouvant déclencher l'arrêt du cycle cellulaire, l'apoptose et la sénescence en réponse à un stress cellulaire (Zucman-Rossi et al., 2015). En effet, parmi les six grandes voies dérégulées dans les CHC, c'est la voie de signalisation cellulaire p53 qui est la plus fréquemment altérée dans les CHC liés au VHB, principalement via des mutations inactivatrices du gène *TP53* dans 41% de ces tumeurs (Amaddeo et al., 2015; Schulze et al., 2015).

Cette association entre l'infection par le VHB et l'altération de *TP53* est généralement expliquée par le nombre important de patients infectés par le VHB qui ont été exposés à l'Aflatoxine B1 (AFB1), une toxine qui induit notamment des mutations du codon 249 de *TP53* (Gouas et al., 2009). En caractérisant les mutations par séquençage haut-débit (génomique entier ou exome entier) et l'exposition à l'AFB1 par analyse des signatures mutationnelles (Letouzé et al., 2017) dans notre cohorte de 265 patients (130 étant infectés par le VHB), nous avons pu vérifier partiellement cette corrélation. En effet, parmi les 24 CHC de patients (dont 22 d'origine africaine) pour lesquelles nous avons pu identifier la signature 24 caractéristique de l'AFB1, seulement 11 avaient la mutation R249S, et réciproquement, 12 tumeurs portaient cette mutation sans avoir une signature (dont 5 provenant de patients d'origine européenne). En revanche, tous les patients ayant la mutation 249S avaient une infection par le VHB, et plus globalement il a déjà été montré que des mutations inactivatrices de *TP53* sont plus fréquentes chez les patients infectés par le VHB (Amaddeo et al., 2015). L'association entre l'infection virale et l'altération de *TP53* est donc forte mais ne s'explique pas exclusivement par une exposition à l'AFB1.

Notre étude a mis en évidence une autre altération récurrente du gène *TP53* dans les CHC liés au VHB, consistant en une intégration virale sur le chromosome 17p associée à une perte d'une région chromosomique contenant le locus de *TP53* et allant du point de jonction avec la séquence virale jusqu'au télomère du bras 17p. Cet événement a été spécifiquement identifié dans 15 tumeurs au sein de notre série de 122 CHC liés au VHB (Article 1, Fig. 3b-3c). Il constitue donc la deuxième altération la plus fréquente associée à une intégration du virus, après les intégrations dans le promoteur de *TERT*

dont nous parlerons par la suite. Cependant, la délétion d'une région du chromosome 17p contenant *TP53* est une altération fréquente (retrouvée dans 50% des CHC de notre série) qui n'est pas significativement associée à l'infection par le VHB, bien qu'il s'agisse de l'étiologie largement majoritaire parmi les CHC présentant ce type d'altération (56%). Cette altération récurrente liée à l'intégration du VHB reflète donc dans ces tumeurs un réarrangement structural qui n'est pas spécifique de l'infection virale. Mais les CHC liés au VHB ayant par ailleurs plus fréquemment des mutations de *TP53*, ils sont globalement plus susceptibles d'acquérir deux hits à ce locus, entraînant alors l'inactivation totale du gène ou la synthèse de protéines mutantes exclusivement. En effet, dans notre série de 265 CHC, la présence d'altérations « double hit » est significativement associée à l'infection du VHB ($p=0.009$). Enfin, nous avons remarqué que les 4 patients de notre série ayant été exposés à l'Aflatoxine B1 et ayant une délétion d'une copie du gène *TP53* liée à une intégration faisaient partie des patients les plus jeunes de la cohorte (entre 32 et 42 ans), soulignant le caractère précoce des altérations et l'agressivité de ces tumeurs *TP53*-altérées basée sur l'apparition d'une forte instabilité.

Un autre type d'altérations générant une instabilité chromosomique dans les tissus hépatiques et favorisant ainsi la carcinogénèse est l'altération des gènes *CCNA2* ou *CCNE1*. Ces deux gènes codent pour des protéines cyclines jouant un rôle clé dans la régulation du cycle cellulaire et des altérations dans ces gènes ont été décrites dans le développement de plusieurs tumeurs, comme le cancer du sein, le cancer du poumon non à petites cellules et le cancer colorectal (Deshpande et al., 2005). En analysant spécifiquement ces deux gènes dans une série de 751 CHC issus de 3 différentes cohortes, nous avons mis en évidence la présence d'altérations localisées à ces loci dans un groupe spécifique de CHC, appelé CHC-CCN et représentant 7% de l'ensemble des CHC. L'activation de ces gènes peut être induite par différents types d'altérations : un réarrangement structural ou une intégration virale du VHB ou du virus AAV2. Bien que les CHC-CCN forment un sous-groupe homogène de tumeurs (Article 2, Fig. 3A), les mécanismes d'activation sont différents suivant le gène considéré :

- Les altérations de *CCNA2* sont localisées dans l'intron 2 et induisent toutes la synthèse d'une protéine tronquée à son extrémité N-terminale, qui ne possède pas le domaine de régulation de la protéine normale (Article 2, Fig. 1).

DISCUSSION

- Les altérations de *CCNE1* sont localisées en amont du gène et ont pour conséquences de positionner une séquence Enhancer d'un génome viral ou venant d'une région distante du génome humain dans la région régulatrice de *CCNE1* (Article 2, Fig. 2).

Ces mécanismes entraînent une surexpression de ces protéines, et une dérégulation du cycle cellulaire induisant un stress répliatif à l'origine de l'instabilité chromosomique. Cette étude nous a permis de mettre en évidence que les intégrations du VHB dans l'un de ces deux gènes avaient des conséquences fonctionnelles qui n'étaient pas spécifiques au virus mais communes à d'autres types d'altérations localisées aux mêmes loci. Pour *CCNA2* notamment, bien que l'intégration du VHB puisse générer la synthèse de protéines de fusion HBs-*CCNA2* (Figure 32), les similitudes entre les transcriptomes issues des différentes altérations laissent penser que c'est principalement la synthèse de la protéine tronquée qui a un impact sur la dérégulation du cycle cellulaire et qui induit la signature de réarrangements observée.

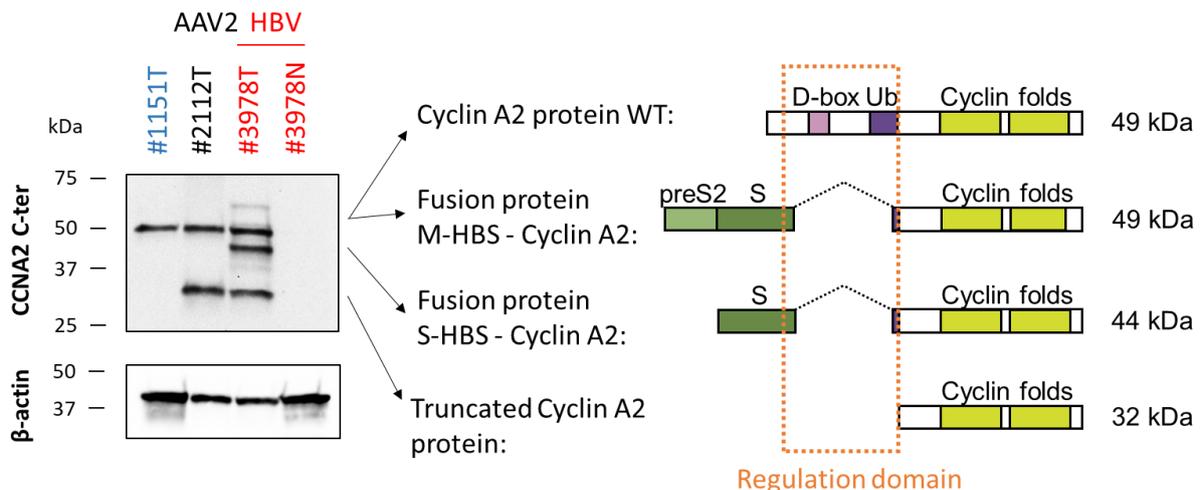


Figure 32 : Prédiction des transcrits et analyse par western-blot du gène *CCNA2* dans le CHC #3798T (intégration du VHB dans l'intron 2).

En ce qui concerne *CCNE1*, l'intégration du VHB induit la surexpression du gène via la séquence Enhancer du génome viral qui est présente dans la séquence intégrée, mais de nouveau, le mécanisme semble identique que pour les autres altérations.

Dans les CHC-CCN comme dans les CHC *TP53*-altérés (avec un double-hit sur *TP53*), l'intégration du VHB est donc un moyen d'altération qui favorise l'initiation de la

cancérogénèse mais sans spécificité particulière liée au virus lui-même. Les conséquences observées aux niveaux transcriptomique et protéique sont en effet similaires dans ces tumeurs quel que soit le type d'altération qui en est à l'origine. De plus, ces CHC ont comme point commun d'être des tumeurs dont le génome est très instable et ils présentent donc un nombre plus élevé de réarrangements chromosomiques et d'altérations du nombre de copies. Les altérations des cyclines et de *TP53* impliquant des mécanismes différents, elles ne sont pas exclusives et parmi les 19 CHC-CCN de *notre* série de 265 CHC, 4 tumeurs possèdent également un double-hit sur le gène *TP53* (dont une tumeur liée au VHB). Ces deux mécanismes oncogéniques forts ont la particularité de déclencher un processus de carcinogénèse chez des patients majoritairement non-cirrhotiques.

B. Caractérisation et spécificité des réarrangements

Il est intéressant de constater que les deux types d'altérations présentés dans le paragraphe précédent (de *TP53* ou des gènes *CCNA2/CCNE1*) induisent deux signatures de réarrangements différentes. Un projet du laboratoire ayant analysé le génome de 35 CHC par séquençage de génome entier a mis en évidence que le type et le nombre de réarrangements structuraux pouvaient varier considérablement entre les tumeurs (Letouzé et al., 2017). La classification de ces réarrangements en 38 catégories a permis de montrer l'existence de 6 signatures de réarrangements différentes notées RS1 à RS6, et cette étude a été approfondie par la suite en ajoutant 10 nouvelles tumeurs (Article 2, Fig. 4A). Les signatures RS1 et RS2 sont toutes deux caractérisées par des duplications en tandem mais de tailles différentes : RS1 est enrichie en petites duplications (<100kb) alors que RS2 regroupe principalement de grandes duplications (>100kb). Ces deux signatures sont justement associées aux deux sous-groupes de tumeurs exposés précédemment : le nombre d'événements attribués à la signature RS1 est en effet très élevé dans les CHC-CCN (Article 2, Fig. 4B) et la signature RS2 a été associée à la présence de mutations *TP53* non seulement parmi les CHC analysés (Letouzé et al., 2017) mais également parmi une série de 560 tumeurs du sein (Nik-Zainal et al., 2016). Les signatures RS1 et RS2 ont aussi la particularité d'être enrichies en variants structuraux situés dans des régions

répliquées précocement et fortement transcrites (Letouzé et al., 2017), des régions où l'on retrouve également un plus grand nombre d'intégrations du VHB (Article 1, Suppl Fig. 2a). Les CHC liés au VHB ayant des altérations des gènes *CCNA2*, *CCNE1* ou *TP53* possèdent d'ailleurs un nombre d'intégrations significativement plus élevé par rapport aux autres CHC. Cela suggère donc que l'instabilité chromosomique générée dans ces tumeurs entraîne par la suite une multiplication de réarrangements pouvant être couplés à des intégrations virales.

Une question qui découle de cette observation est de savoir si dans les tumeurs liées au VHB, le virus induit directement des réarrangements en interférant avec la machinerie cellulaire lors de la réplication ou s'il s'intègre de manière opportuniste lorsqu'un réarrangement se produit spontanément dans une cellule infectée en perturbant les mécanismes de réparation. Certains oncovirus comme les virus herpes et le papillomavirus humain peuvent générer des cassures double-brins dans l'ADN génomique afin de favoriser le recrutement de la machinerie cellulaire nécessaire à leur réplication (Gaglia and Munger, 2018). Concernant le VHB, des études effectuées dans des modèles *in vitro* suggèrent plutôt l'association inverse, c'est-à-dire que les intégrations du VHB se produisent préférentiellement à des sites de cassures double-brins de l'ADN génomique qui sont induits par des mécanismes extérieurs (Bill and Summers, 2004; Dandri, 2002). Les modèles utilisés étant basé sur des transfections de l'ADN viral, il serait cependant intéressant de confirmer cette observation dans un modèle d'infection réelle, comme par exemple en générant avec le système CRISPR-Cas9 des cassures dans l'ADN de cellules pouvant être infectés par des virions du VHB. Dans notre série de 265 tumeurs, nous pouvons remarquer que le nombre de variants structuraux n'est pas significativement plus élevé dans les CHC de patients infectés par le VHB que dans ceux de patients non-infectés, au sein des sous-groupes CHC-CCN et CHC *TP53*-altérés, suggérant que le virus n'est probablement pas à l'origine de ces réarrangements. Il est néanmoins probable que la séquence virale intégrée modifie ou accentue les conséquences du réarrangement au milieu duquel elle se trouve et que l'intégration contribue ainsi au développement tumoral. Les CHC comportant une intégration dans *CCNE1* montrent bien l'impact direct que peut avoir la séquence intégrée puisque c'est la présence de la région Enhancer du génome viral qui permet la surexpression du gène et donc déclenche l'initiation tumorale.

Une particularité des réarrangements structuraux associés aux intégrations est leur enrichissement dans les régions centromériques et télomériques. Ces régions sont impliquées dans des réarrangements complexes se produisant lors de la division cellulaire, comme les cycles BSB (Breakage-Fusion-Bridge) suite à une cassure télomérique ou les iso-chromosomes résultant d'une fission centromérique (Figure 33).

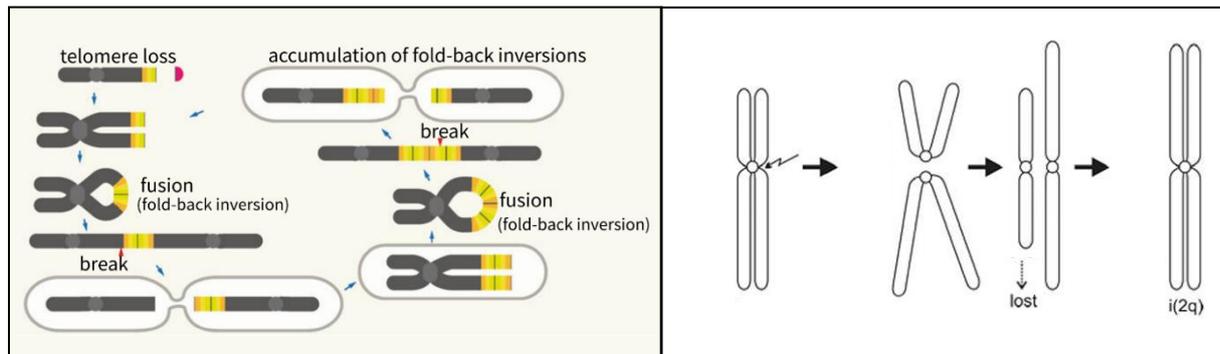


Figure 33 : Exemples de réarrangements chromosomiques complexes : cycle BSB (gauche, Yi and Ju, 2018) et iso-chromosome (droite, Gardner et al., 2011).

Grâce au séquençage long-read, nous avons notamment pu identifier dans une tumeur de notre série la présence d'un iso-chromosome iso(8q) dans la tumeur #1151T (Article 1, Suppl Fig. 6a) associée à une intégration virale dans le centromère. Cette anomalie caractérisée par le gain du bras 8q et la perte du bras 8p est en réalité fréquemment observée dans les CHC (dans environ 50% des cas ; Figure 34).

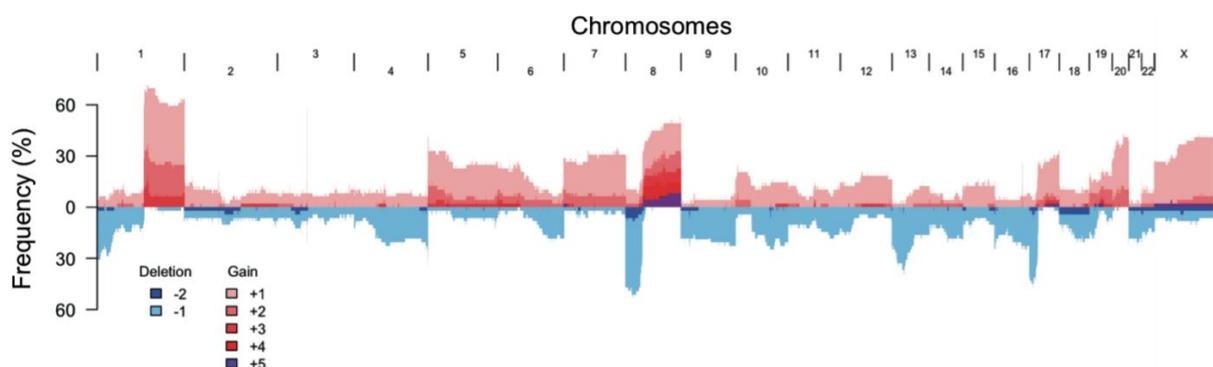


Figure 34 : Fréquence des gains et délétions dans 35 CHC (Letouzé et al., 2017)

Cela semble de nouveau renforcer l'hypothèse que les intégrations du VHB sont associées à des altérations chromosomiques qui sont retrouvées dans l'ensemble des CHC et qui sont donc sélectionnées fonctionnellement de par leurs conséquences oncogéniques sans nécessairement avoir de lien direct avec la présence du virus.

C. Activation de *TERT* par la séquence intégrée du VHB

Le promoteur de *TERT* est le seul locus à être altéré par les réarrangements structuraux induits dans les CHC-CCN à une fréquence plus élevée que celle attendue, indiquant une sélection positive de cette altération dans les tumeurs du sous-groupe (Article 2, Fig. 2E). Dans notre série de CHC liés au VHB analysée par capture virale, nous avons identifié 5 tumeurs ayant une intégration dans *CCNA2* ou *CCNE1* (dont 2 décrites dans l'Article 2) et toutes ont une altération du promoteur de *TERT* : 2 ont une translocation vers un autre chromosome et 3 ont une autre intégration du VHB.

Les intégrations du VHB dans le promoteur de *TERT* ont comme point commun avec les intégrations dans *CCNE1* d'induire une surexpression du gène liée à la présence d'une séquence Enhancer du génome viral, et notamment la région Enhancer I qui induit une activation plus forte que la région Enhancer II (Article 1, Fig. 4c). Cette observation avait déjà été décrite dans la littérature dans deux études utilisant des modèles cellulaires qui ont montré que la région Enhancer I permettait le recrutement de facteurs de transcription induisant l'activation de *TERT* (Li et al., 2019; Sze et al., 2020). Les facteurs de transcription mis en évidence diffèrent néanmoins dans les deux études : la première publication observe le recrutement des récepteurs des hormones sexuelles (Li et al., 2019) et la deuxième étude du facteur de transcription ELF4 (E74-like ETS transcription factor 4) (Sze et al., 2020). Nous avons regardé le niveau d'expression de ces facteurs dans notre série de tumeurs à partir des données de RNA-Seq, et observé que le récepteur des androgènes (AR) a un niveau d'expression significativement plus élevé dans les CHC ayant une altération du promoteur de *TERT*, ce qui pourrait suggérer une sélection fonctionnelle d'hépatocytes ayant une forte expression d'AR induisant une forte expression de *TERT*. Cependant, l'expression de *TERT* n'est pas associée au sexe masculin dans notre série, quel que soit le type d'altération. De plus, nous n'avons pas noté de différence d'expression liée à l'altération du promoteur de *TERT* pour les gènes du récepteur des œstrogènes *ESR1* et *ELF4*. D'autres investigations seraient nécessaires pour confirmer ou infirmer le rôle des facteurs de transcription dans l'activation de *TERT*, et peut-être un jour envisager la privation d'androgène comme thérapie possible du CHC.

Par ailleurs, l'activation du gène en cis par la séquence virale peut également être amplifiée lorsque l'intégration est associée à un réarrangement structural induisant un gain de cette région chromosomique. Cependant, l'expression de *TERT* étant déjà très élevée dans les tumeurs ayant une intégration du VHB, elle n'augmente pas de manière significative lorsque l'intégration est associée à un gain du locus du gène.

L'altération de *TERT* est un événement qualifié de « gatekeeper » dans le développement tumoral (Rebouissou and Nault, 2020). En effet, elle est nécessaire pour permettre la réactivation de la télomérase et ainsi empêcher le raccourcissement des télomères qui déclencherait la sénescence des hépatocytes lors de leur prolifération. Comme cela a été décrit dans la partie I.C.1 de l'Introduction, des mutations du promoteur de *TERT* ont été détectées dès les étapes précoces de la carcinogénèse sur foie cirrhotique : dans 6% des nodules dysplasiques de bas grade et 19% des nodules de haut grade (Nault et al., 2014). Cependant, dans les CHC-CCN, l'événement le plus précoce est a priori l'altération du gène *CCNA2* ou *CCNE1* qui à l'origine de l'instabilité qui va induire par la suite des réarrangements dans le promoteur de *TERT* (pouvant être associé à une autre intégration du VHB), et ces CHC ont été principalement identifiés chez des patients sans cirrhose hépatique. L'altération du promoteur de *TERT* serait donc plutôt dans ces tumeurs un deuxième hit. Le processus de carcinogénèse chez ces patients semble donc plus proche de celui observé dans la transformation maligne des adénomes hépatocellulaires (AHC). En effet, des mutations du promoteur de *TERT* ont été observées dans 44% des AHC transformés mais pas dans les AHC sans dysplasie, indiquant que l'altération de *TERT* est alors la dernière étape déclenchant la transformation maligne (Nault et al., 2013). Dans le processus de carcinogénèse sur foie sain en général, l'altération du promoteur de *TERT* interviendrait donc plutôt dans un deuxième temps.

En ce qui concerne les CHC TP53-altérés, ils ne sont pas significativement enrichis en altérations du promoteur de *TERT*. En réalité, 18% (22 sur 121) des tumeurs liées au VHB de notre série analysée en capture et en RNA-Seq ne présentent aucune altération de *TERT*. 41 CHC (34%) ont une intégration du VHB dans le promoteur de *TERT*, 48 CHC (40%) ont une mutation du promoteur en -124, -146 ou -57, et 10 (8%) ont un gain (dont 5 pour lesquels un variant structural a été identifié). Un projet du laboratoire a analysé spécifiquement les altérations de *TERT* dans une série de 256 CHC provenant de 210 patients avec différentes étiologies, et a identifié un

de *TERT* substantielle malgré l'absence d'altération. Il s'agit de tumeurs développées chez des patients majoritairement jeunes et de sexe féminin, ayant une origine africaine et sans fibrose (score METAVIR F0-F1). Nous avons cherché des altérations dans les gènes « driver » associés au CHC et nous avons identifié un enrichissement en mutations *AXIN1* (7 tumeurs sur 22). Ce groupe de CHC regroupe également des tumeurs présentant des mutations *TP53* (9 tumeurs dont 3 ont également une mutation *AXIN1*), et des tumeurs présentant des intégrations du VHB dans le gène *KMT2B* (2 tumeurs n'ayant aucune autre altération « driver »).

Le point commun principale entre ces CHC est que d'un point de vue transcriptomique, ils appartiennent au groupe G1 comme cela apparaît sur la Figure 35 et ils se caractérisent par une forte surexpression de gènes cibles exprimés par les cellules progénitrices hépatiques et impliqués dans la vie embryonnaire et fœtale, comme *EPCAM* ou *IGF2* (Article 1, Fig. 4d). Les tumeurs ayant des mutations *BAP1* ont été décrites comme ayant un profil transcriptomique similaire sans présenter d'altération de *TERT* (Hirsch et al., 2019; Ningarhari et al., Non publié), mais aucun CHC lié au VHB ne présentant d'altération du gène *BAP1*, cela suggère deux mécanismes différents. Il est possible que le profil « progéniteur » de la tumeur reflète la nature de la cellule d'origine qui a subi la transformation maligne, car les cellules progénitrices peuvent maintenir leurs télomères sans altération du promoteur (Lin et al., 2018). Alors que les tumeurs mutées *BAP1*, qui présentent des caractéristiques proches des tumeurs fibrolamellaires, seraient issues d'une dé-différentiation d'un hépatocyte mature suite à l'inactivation de *BAP1* (Hirsch et al., 2019), les tumeurs liées au VHB du même sous-groupe transcriptomique proviendraient de tumeurs initiées chez des patients jeunes où l'infection aurait potentiellement pu bloquer la maturation de cellules hépatiques immatures. Par ailleurs, ce sous-groupe G1 contenant également les CHC ayant des intégrations dans le gène *KMT2B*, codant pour une protéine modifiant la méthylation des histones, il est possible que des mécanismes épigénétiques soient impliqués.

Enfin, mieux caractériser ce sous-groupe de CHC nous paraît d'autant plus un enjeu important dans la compréhension de la carcinogénèse liée au VHB que le groupe transcriptomique G1 est associée dans notre série à une sérologie AgHBs négative et donc à de potentielles infections B occultes : 11% (5/44) et 57% (4/7) des CHC sont du groupe G1 parmi les patients AgHBs-positif et AgHBs-négatif respectivement.

III. Similitudes et spécificités du VHB parmi les oncovirus

Trois mécanismes principaux impliqués dans la carcinogénèse viro-induite ont été décrits dans la partie III.A.2 de l'Introduction : l'expression de protéines virales oncogéniques, l'établissement d'une inflammation chronique et l'induction d'altérations génomiques de la cellule hôte dont font partie les intégrations virales. Parmi les six virus inclus dans le groupe 1 du classement du Centre International de Recherche sur le Cancer (CIRC) répertoriant les agents carcinogènes pour l'humain, seuls le VHB et le VPH ont la capacité de s'intégrer dans le génome de la cellule qu'ils infectent. Mais plus récemment, des séquences intégrées de deux nouveaux virus ont été découvertes : le polyomavirus à cellules de Merkel intégré dans des carcinomes à cellules de Merkel (Feng et al., 2008), et le virus AAV2 intégré dans des CHC (Nault et al., 2015). Cela pose donc des questions sur les points communs et les spécificités existants entre les mécanismes d'intégrations de ces 4 virus et notamment entre les rôles que peuvent jouer les différentes séquences intégrées dans la carcinogénèse.

A. Intégrations virales et carcinogénèse : VHB versus VPH

Le VPH et le VHB sont les deux virus à l'origine du plus grand nombre de cancers viro-induits à l'échelle mondiale : en 2018, le CIRC a recensé 690 000 cas de cancers liés au VPH (principalement des cancers du col de l'utérus) et 360 000 cas de CHC liés au VHB (de Martel et al., 2020). Le mécanisme d'intégration du VPH a donc également été investigué dans de nombreux travaux et notre étude sur le VHB nous a alors permis de mettre en lumière les similitudes et différences entre les deux virus.

Une étude du TCGA a identifié des intégrations du VPH dans 83% des échantillons tumoraux analysés (The Cancer Genome Atlas Research Network, 2017), ce qui est du même ordre de grandeur que pour le VHB dans les CHC (82% dans notre série de 177 tumeurs en capture virale et 84% dans l'étude du TCGA sur 44 CHC liés au VHB ; Ally et al., 2017), mais le nombre d'événements détecté par tumeur est globalement plus faible (~1,5 et ~6,1 intégrations par tumeur pour le VPH et le VHB respectivement, entre les deux analyses du TCGA à partir de données RNA-Seq). Les conséquences oncogéniques des intégrations présentent également des différences notables.

- Expression des protéines virales :

Comme pour le VHB, l'intégration du VPH ne fait pas partie du cycle de réplication virale et les séquences intégrées ne possèdent pas de capacité de réplication. Mais alors que dans les CHC liés au VHB, le rôle oncogénique joué par HBx est très discuté (étant donné son faible niveau d'expression et la présence de fréquentes mutations inactivatrices du gène dans les séquences intégrées), une conséquence majeure des intégrations du VPH consiste en l'expression des oncogènes viraux E6 et E7 combinée à l'absence d'expression du gène E2 qui réprime normalement E6/E7. Cela favorise donc la prolifération cellulaire, dérégule le contrôle du cycle cellulaire et provoque une instabilité génétique progressive. La persistance de certains cancers liés au VPH a en effet été décrite comme dépendante de l'expression de ces oncogènes (Goodwin et al., 2000). Cependant, des études récentes ont montré que le développement tumoral pouvait également avoir lieu sans expression des oncogènes E6/E7, suggérant l'implication d'autres mécanismes (Hu et al., 2015).

- Mutagenèse insertionnelle :

De la même façon que pour le VHB, les intégrations du VPH peuvent altérer la structure ou l'expression de gènes cellulaires impliqués dans la carcinogénèse : en augmentant l'expression d'oncogènes via un promoteur viral ou en entraînant une perte de fonction d'un gène suppresseur de tumeurs. Cependant, alors que le promoteur de *TERT* représente un important « hotspot » contenant une intégration virale dans un tiers des CHC de notre série, aucun site récurrent d'intégration n'a été identifié dans les cancers liés au VPH (Akagi et al., 2014; Zapatka et al., 2020). L'intégration du VHB induisant la présence d'une région Enhancer virale dans le promoteur de *TERT* et ainsi l'expression du gène est donc un mécanisme spécifique au VHB permettant la réactivation de la télomérase. Dans les cancers liés au VPH, cette réactivation est induite par l'interaction de la protéine virale E6 avec le promoteur de *TERT* et avec le complexe de la télomérase directement (Liu et al., 2009).

- Instabilité chromosomique :

Une forte association entre la localisation de points d'intégrations du VPH et de variations structurales génomiques a été mise en évidence dans des cancers liés au VPH présentant des séquences virales intégrées, de façon similaire aux intégrations du VHB dans notre série (Akagi et al., 2014; Holmes et al., 2016). Un motif particulier

de réarrangements a été identifié et repose sur un mécanisme de « boucle » (looping ou rolling-circle) générant des amplifications de large régions chromosomiques, qui se produit lors de la réplication de l'ADN à partir de fourches de réplication localisées dans la séquence virale (Figure 36) (Akagi et al., 2014; Yi and Ju, 2018).

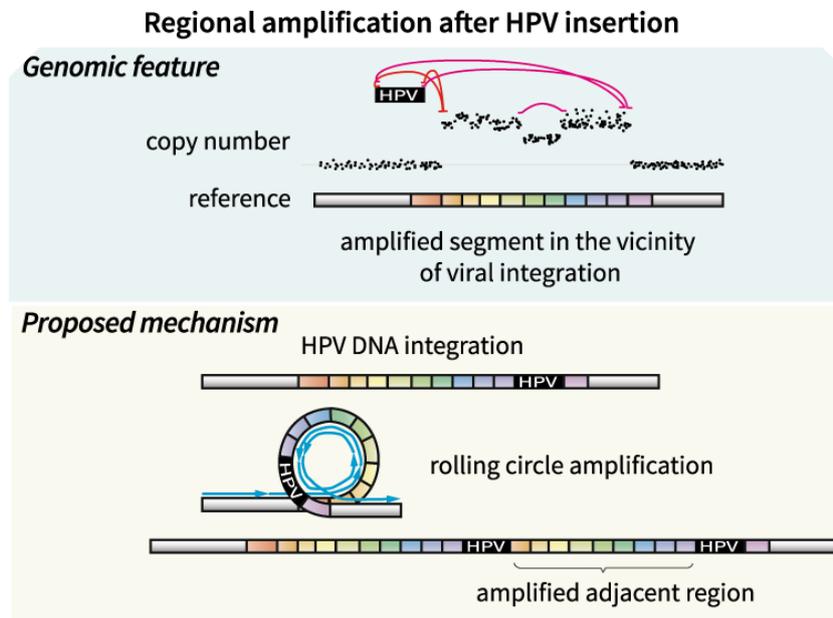


Figure 36 : Profil génomique et schéma du mécanisme d'amplification en « rolling circle » lié à l'intégration d'ADN du VPH (Yi and Ju, 2018)

Nous n'avons pas identifié spécifiquement ce mécanisme dans les 3 tumeurs séquencées en long-read, mais une analyse plus précise avec un plus grand nombre d'échantillons serait nécessaire pour détecter ce réarrangement particulier. Les variants structuraux associés à des intégrations virales que nous avons pu mettre en évidence dans des CHC liés au VHB impliquent très fréquemment des réarrangements à l'intérieur même de la séquence virale intégrée. En effet, un événement de « duplication-inversée » (Article 1, Fig. 3d) nécessite une inversion dans la séquence virale pour obtenir généralement une séquence intégrée dupliquée en miroir.

A notre connaissance, des réarrangements internes au génome du VPH n'ont pour l'instant pas été décrits dans les séquences intégrées, et cette caractéristique d'instabilité supplémentaire serait donc potentiellement spécifique au VHB. Dans notre série de CHC liés au VHB, nous avons identifié un nombre considérable de réarrangements au sein d'une même tumeur (la tumeur #1987T présente 43 réarrangements pour 13 intégrations clonales ; Figure 37).

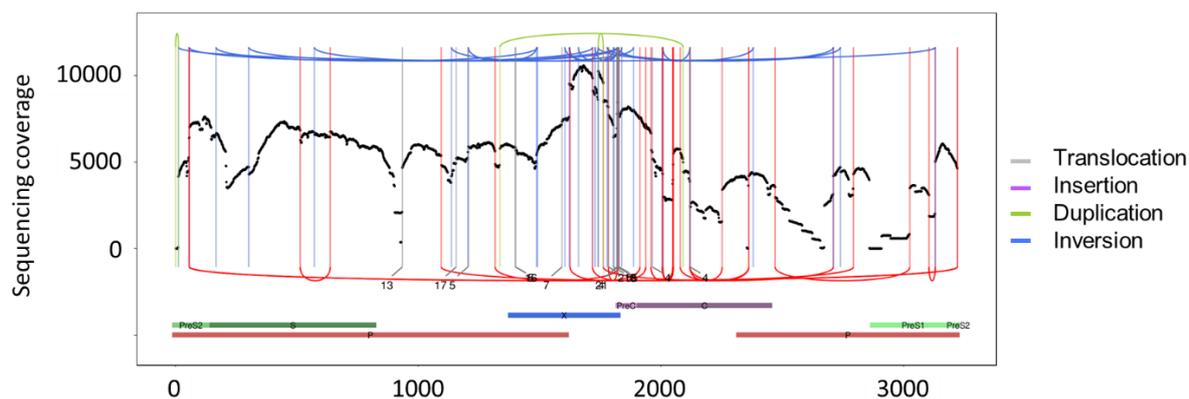


Figure 37 : Profil des réarrangements structuraux internes aux génomes du VHB présents dans la tumeur #1987T

La caractérisation précise de ces réarrangements est donc nécessairement rendue complexe par le nombre d'intégrations clonales présentes au sein d'un même échantillon. Mais concernant la dynamique de ces réarrangements, deux arguments nous laissent penser qu'ils ont lieu pendant ou après l'intégration. Tout d'abord, étant donné la très petite taille du génome viral (3,2kb), il est peu probable que ces réarrangements aient eu lieu suite à des cassures double-brins aléatoires dans l'ADNccc ou ADNdsl présent dans les cellules infectées, de la même façon que cela arrive spontanément dans l'ADN génomique. De plus, le petit nombre d'échantillons analysés par séquençage long-read nous a permis d'observer que chaque réarrangement n'est présent que dans une seule séquence intégrée et n'apparaît pas dans l'ensemble des intégrations de la tumeur. Au final, il serait intéressant d'étudier plus précisément ces variants structuraux viraux pour déterminer s'il existe une corrélation entre les types de réarrangements observés dans le génome viral et dans le génome humain (en lien avec les six signatures de réarrangements structuraux identifiées dans les CHC ; Letouzé et al., 2017) et pour éventuellement identifier des positions du génome viral plus susceptibles de générer ces variants (ce qui pourrait induire la synthèse de transcrits altérés récurrents des gènes viraux).

B. Hépatocarcinogénèse et virus : VHB, VHC et aussi VHD et AAV

Pendant longtemps, l'hépatocarcinogénèse viro-induite a été exclusivement associée aux virus de l'Hépatite B et de l'Hépatite C. Cependant, comme décrit dans la partie II.C.3 de l'Introduction, 5% des personnes infectés par le VHB sont co-infectés par le

VHD dans le monde et des études cliniques ont montré que le risque de survenue du CHC était plus élevé chez les patients co-infectés que chez les patients infectés par le VHB seul (Mentha et al., 2019; World Health Organization, 2017). Cependant, aucune étude génomique décrivant le profil mutationnel et transcriptomique des tumeurs liées à des co-infections VHB-VHD n'a été publiée à l'heure actuelle (Puigvehí et al., 2019).

Dans notre cohorte de patients ayant développé des CHC liés au VHB, nous avons identifié la présence d'ARN viraux du VHD dans 10% des tumeurs (16/158) et 9% des tissus non-tumoraux adjacents (15/159), chez 23 patients. D'un point de vue clinique, la co-infection VHB/VHD était associée significativement dans notre série à un âge jeune, et à la présence de cirrhose. Par rapport aux facteurs viraux, aucun patient ayant une co-infection VHB/VHD ne présentait de réplication virale du VHB dans le foie non-tumoral ou dans la tumeur, confirmant ainsi une dominance du VHD sur le VHB (Mentha et al., 2019). De plus, les tissus non-tumoraux de patients co-infectés présentaient un faible nombre de copies du VHB par cellule et un faible nombre d'intégrations (Article 1, Fig. 1d), mais nous n'avons pas retrouvé ces associations dans la tumeur. Cela suggère que ces patients ont a priori d'abord été infectés par le VHB puis par le VHD qui a bloqué la réplication virale du VHB mais après son intégration dans le génome humain. D'un point de vue moléculaire, nous n'avons pas identifié d'altérations récurrentes parmi les tumeurs VHB/VHD de notre série et elles n'appartiennent pas à un groupe transcriptomique spécifique dans la classification G1-G6. Leur développement semble principalement lié à la présence d'une inflammation chronique induite par la réplication active du VHD, et la co-infection augmenterait donc le risque de survenue d'un CHC en accélérant la progression vers la cirrhose. De ce fait, le VHD pourrait être considéré comme un virus oncogénique via son action inflammatoire sur le microenvironnement hépatique, au même titre que le VHC. Ces tumeurs nécessiteraient cependant une étude spécifique pour être mieux caractérisées. Parmi les patients co-infectés de notre série, 3 avaient une sérologie AgHBs négative et donc une potentielle infection occulte, ce qui montre l'intérêt fort que pourrait avoir une caractérisation précise de ces tumeurs.

Enfin, une partie de cette thèse a consisté à étudier l'infection par le virus AAV2 comme nouveau facteur de risque du CHC. Malgré la forte prévalence d'AAV2 dans la population générale (entre 30 et 80% en fonction des génotypes ; Boutin et al., 2010), l'intégration clonale de séquences d'AAV2 est un événement rare qui a été identifié

dans 2% des CHC d'une série constituée au laboratoire. Néanmoins, AAV2 a comme point commun avec le VHB de pouvoir induire un développement tumoral en absence de cirrhose et donc via un événement oncogénique fort lié aux intégrations.

Les intégrations clonales d'AAV2 dans les CHC ont été identifiées dans des gènes également ciblés par le VHB : *TERT*, *KMT2B* et *CCNE1* (Article 3, Fig. 4D – Fig. 5), suggérant que les séquences intégrées des deux virus induisent des conséquences similaires, aboutissant à une sélection fonctionnelle des hépatocytes contenant des intégrations dans ces gènes. Dans le sous-groupe des CHC-CCN, l'activation des gènes *CCNA2* et *CCNE1* est dépendante de la localisation de l'intégration et de la présence d'une séquence Enhancer pouvant provenir du génome du VHB ou d'AAV2, ou d'une région du génome humain suite à un réarrangement structural. Ces gènes sont donc étroitement liés au processus d'hépatocarcinogénèse quelle que soit l'étiologie sous-jacente. Le nombre total d'intégrations d'AAV2 dans le génome humain est cependant plus faible que pour le VHB et cela peut s'expliquer par le fait que le nombre de copies du virus par cellule dans le foie non-tumoral est plus faible pour AAV2 que pour le VHB (10^{-5} à 0,04 et 10^{-3} à 100 pour AAV2 et HBV respectivement ; Article 3, Fig.1A). Le processus d'intégration est donc probablement beaucoup plus fréquent pour le VHB. De plus, plusieurs génomes viraux différents sont généralement détectés dans une tumeur liée au VHB (plusieurs intégrations clonales et présence éventuelle d'une forme répliquative), alors qu'aucun CHC ayant une intégration clonale d'AAV2 ne contient de formes épisomales du virus. La carcinogénèse liée à AAV2 est donc a priori uniquement liée aux conséquences de la mutagénèse insertionnelle.

Une dernière spécificité importante d'AAV2 est que nous avons identifié la présence de formes épisomales virales dans des tissus de tumeurs bénignes (adénomes hépatocellulaires et hyperplasies nodulaires focales) et cela plus fréquemment que dans les tissus de tumeurs malignes, suggérant que la présence d'une infection active d'AAV est associée au développement de tumeurs bénignes. Une intégration clonale a cependant été identifiée dans une hyperplasie nodulaire focale mais dans une région intergénique sans conséquence au niveau transcriptomique.

Au final, la caractérisation de ces nouveaux oncovirus a permis de mieux comprendre le rôle joué par les infections virales dans les mécanismes d'hépatocarcinogénèse indirects comme l'inflammation et directs comme la mutagénèse insertionnelle.

CONCLUSIONS ET PERSPECTIVES

Nos travaux ont permis d'apporter un nouvel éclairage sur les mécanismes d'intégrations du virus de l'Hépatite B et sur les conséquences oncogéniques de ces événements. A partir d'une analyse intégrée de données provenant d'échantillons hépatiques d'une cohorte de patients ayant développé un carcinome hépatocellulaire lié à une infection B chronique, nous avons pu caractériser précisément les formes virales présentes dans les tissus hépatiques tumoraux et non-tumoraux.

En particulier, nous avons mis en évidence un lien direct entre réplication et intégrations virales qui peut avoir un impact clinique important pour le traitement précoce des infections par le VHB et la surveillance des patients. Plus globalement, nous avons décrit les facteurs principaux liés aux intégrations virales dans le foie non-tumoral et les processus impliqués dans la sélection clonale des hépatocytes infectés, qui peut résulter ou non d'une transformation maligne. Enfin, nous avons montré pour la première fois que l'association directe entre des intégrations du VHB et des réarrangements structuraux peut avoir des conséquences en *cis* ou en *trans* favorisant l'initiation et le développement tumoral.

Etant donné la diversité des CHC liés au VHB, beaucoup de questions restent néanmoins ouvertes pour avoir un aperçu complet de la carcinogénèse induite par ce virus. La richesse de notre série de données pourrait alors permettre par la suite d'investiguer plus en détails certains sous-groupes de tumeurs spécifiques. L'identification du sous-groupe des CHC-CCN regroupant 7% des CHC a en effet montré l'importance de réaliser une caractérisation précise des altérations moléculaires dans les tumeurs pour comprendre les voies de signalisation altérées et en déduire de potentielles opportunités thérapeutiques personnalisées. Définir les CHC-CCN comme des tumeurs particulièrement prolifératives et liées à un stress réplicatif permet d'envisager pour les patients des thérapies ciblant principalement les cellules en divisions cellulaires (chimiothérapies) ou les voies de réponse au stress réplicatif (inhibiteurs de la voie ATR). Il est donc primordial de poursuivre cette description moléculaire des CHC liés au VHB, et il serait notamment intéressant d'étudier plus précisément les CHC liés à une co-infection VHB/VHD (14% des CHC notre série), les CHC ayant un profil « progéniteur » appartenant au groupe

transcriptomique G1 mais dont les altérations sont encore mal définies (14%), ou bien les CHC développés chez des patients ayant une sérologie AgHBs négative (8%), ces groupes n'étant pas exclusifs entre eux.

Ce dernier sous-groupe constitue un enjeu particulièrement important pour le suivi des patients, car à l'heure actuelle la séroconversion de l'AgHBs est considérée comme l'objectif de « guérison fonctionnelle » à atteindre. Le risque de développer un CHC chez les patients « guéris » est faible mais reste non nulle, et la caractérisation des intégrations et des formes virales présentes chez ces patients pourrait potentiellement permettre d'affiner le calcul de ce risque, pour diagnostiquer le plus précocement possible un éventuel CHC. Une perspective de ce travail repose ainsi sur l'analyse du VHB dans les plasmas de certains patients disponibles au laboratoire, pour comparer leurs profils d'intégrations avec les tissus hépatiques et identifier d'éventuels biomarqueurs permettant une surveillance plus adaptée des patients (Li et al., 2020)

Pour conclure, les infections virales sont en plein cœur des considérations actuelles mais le Virus de l'Hépatite B est resté longtemps dans l'ombre de maladies infectieuses plus célèbres comme la tuberculose, le paludisme ou le SIDA. Il concerne pourtant 3,5% de la population mondiale et provoque un million de morts chaque année. Au final, en comparaison au nouveau coronavirus à l'origine de la pandémie actuelle, le VHB n'a jamais beaucoup attiré l'attention internationale et probablement parce qu'il n'existe plus que dans les pays en développement. Malgré l'existence d'un vaccin efficace et de traitements supprimant la réplication virale, cette « épidémie silencieuse » augmente donc encore aujourd'hui, notamment en Afrique (Graber-Stiehl, 2018; Lemoine and Thursz, 2017), avec une mortalité liée dans 39% des cas au développement d'un CHC. La totalité des publications sur les intégrations du VHB sont basées sur des cohortes de patients d'origine asiatique (à l'exception de quelques patients européens). Nos travaux ont donc aussi cherché à remédier à une absence de données qui était contradictoire avec la répartition actuelle et l'évolution de l'infection à l'échelle mondiale, en étant la première étude caractérisant les intégrations du VHB dans une cohorte de patients dont 30% étaient d'origine africaine.

BIBLIOGRAPHIE

- Abou-Alfa, G.K., Meyer, T., Cheng, A.-L., El-Khoueiry, A.B., Rimassa, L., Ryoo, B.-Y., Cicin, I., Merle, P., Chen, Y., Park, J.-W., Blanc, J.-F., Bolondi, L., Klümpen, H.-J., Chan, S.L., Zagonel, V., Pressiani, T., Ryu, M.-H., Venook, A.P., Hessel, C., Borgman-Hagey, A.E., Schwab, G., Kelley, R.K., 2018. Cabozantinib in Patients with Advanced and Progressing Hepatocellular Carcinoma. *N. Engl. J. Med.* 379, 54–63. <https://doi.org/10.1056/NEJMoa1717002>
- Abul-Husn, N.S., Cheng, X., Li, A.H., Xin, Y., Schurmann, C., Stevis, P., Liu, Y., Kozlitina, J., Stender, S., Wood, G.C., Stepanchick, A.N., Still, M.D., McCarthy, S., O'Dushlaine, C., Packer, J.S., Balasubramanian, S., Gosalia, N., Esopi, D., Kim, S.Y., Mukherjee, S., Lopez, A.E., Fuller, E.D., Penn, J., Chu, X., Luo, J.Z., Mirshahi, U.L., Carey, D.J., Still, C.D., Feldman, M.D., Small, A., Damrauer, S.M., Rader, D.J., Zambrowicz, B., Olson, W., Murphy, A.J., Borecki, I.B., Shuldiner, A.R., Reid, J.G., Overton, J.D., Yancopoulos, G.D., Hobbs, H.H., Cohen, J.C., Gottesman, O., Teslovich, T.M., Baras, A., Mirshahi, T., Gromada, J., Dewey, F.E., 2018. A Protein-Truncating *HSD17B13* Variant and Protection from Chronic Liver Disease. *N. Engl. J. Med.* 378, 1096–1106. <https://doi.org/10.1056/NEJMoa1712191>
- Ahn, S., Jang, S.J., Shim, J.H., Kim, D., Hong, S., Sung, C.O., Baek, D., Haq, F., Ansari, A.A., Lee, S.Y., Chun, S., Choi, S., Choi, H., Kim, J., Kim, S., Hwang, S., Lee, Y., Lee, J., Jung, W., Jang, H.Y., Yang, E., Sung, W., Lee, N.P., Mao, M., Lee, C., Zucman-Rossi, J., Yu, E., Lee, H.C., Kong, G., 2014. Genomic portrait of resectable hepatocellular carcinomas: Implications of *RB1* and *FGF19* aberrations for patient stratification. *Hepatology* 60, 1972–1982. <https://doi.org/10.1002/hep.27198>
- Akagi, K., Li, J., Broutian, T.R., Padilla-Nash, H., Xiao, W., Jiang, B., Rocco, J.W., Teknos, T.N., Kumar, B., Wangsa, D., He, D., Ried, T., Symer, D.E., Gillison, M.L., 2014. Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res.* 24, 185–199. <https://doi.org/10.1101/gr.164806.113>
- Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., Islam, S.M.A., Lopez-Bigas, N., Klimczak, L.J., McPherson, J.R., Morganella, S., Sabarinathan, R., Wheeler, D.A., Mustonen, V., Getz, G., Rozen, S.G., Stratton, M.R., 2020. The repertoire of mutational signatures in human cancer. *Nature* 578, 94–101. <https://doi.org/10.1038/s41586-020-1943-3>
- Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A.V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.-L., Boyault, S., Burkhardt, B., Butler, A.P., Caldas, C., Davies, H.R., Desmedt, C., Eils, R., Eyfjörd, J.E., Foekens, J.A., Greaves, M., Hosoda, F., Hutter, B., Illicic, T., Imbeaud, S., Imielinski, M., Jäger, N., Jones, D.T.W., Jones, D., Knappskog, S., Kool, M., Lakhani, S.R., López-Otín, C., Martin, S., Munshi, N.C., Nakamura, H., Northcott, P.A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J.V., Puente, X.S., Raine, K., Ramakrishna, M., Richardson, A.L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T.N., Span, P.N., Teague, J.W., Totoki, Y., Tutt, A.N.J., Valdés-Mas, R., van Buuren, M.M., van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L.R., Zucman-Rossi, J., Andrew Futreal, P., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S.M., Siebert, R., Campo, E., Shibata, T., Pfister, S.M., Campbell, P.J., Stratton, M.R., 2013. Signatures of mutational processes in human cancer. *Nature* 500, 415–421. <https://doi.org/10.1038/nature12477>
- Allweiss, L., Dandri, M., 2016. Experimental in vitro and in vivo models for the study of human hepatitis B virus infection. *J. Hepatol.* 64, S17–S31. <https://doi.org/10.1016/j.jhep.2016.02.012>
- Allweiss, L., Strick-Marchand, H., 2020. In-vitro and in-vivo models for hepatitis B cure research: *Curr. Opin. HIV AIDS* 15, 173–179. <https://doi.org/10.1097/COH.0000000000000616>

- Ally, A., Balasundaram, M., Carlsen, R., Chuah, E., Clarke, A., Dhalla, N., Holt, R.A., Jones, S.J.M., Lee, D., Ma, Y., Marra, M.A., Mayo, M., Moore, R.A., Mungall, A.J., Schein, J.E., Sipahimalani, P., Tam, A., Thiessen, N., Cheung, D., Wong, T., Brooks, D., Robertson, A.G., Bowlby, R., Mungall, K., Sadeghi, S., Xi, L., Covington, K., Shinbrot, E., Wheeler, D.A., Gibbs, R.A., Donehower, L.A., Wang, L., Bowen, J., Gastier-Foster, J.M., Gerken, M., Helsel, C., Leraas, K.M., Lichtenberg, T.M., Ramirez, N.C., Wise, L., Zmuda, E., Gabriel, S.B., Meyerson, M., Cibulskis, C., Murray, B.A., Shih, J., Beroukhi, R., Cherniack, A.D., Schumacher, S.E., Saksena, G., Peadarallu, C.S., Chin, L., Getz, G., Noble, M., Zhang, Hailei, Heiman, D., Cho, J., Gehlenborg, N., Saksena, G., Voet, D., Lin, P., Frazer, S., Defreitas, T., Meier, S., Lawrence, M., Kim, J., Creighton, C.J., Muzny, D., Doddapaneni, H., Hu, J., Wang, M., Morton, D., Korchina, V., Han, Y., Dinh, H., Lewis, L., Bellair, M., Liu, X., Santibanez, J., Glenn, R., Lee, S., Hale, W., Parker, J.S., Wilkerson, M.D., Hayes, D.N., Reynolds, S.M., Shmulevich, I., Zhang, W., Liu, Y., Iype, L., Makhlof, H., Torbenson, M.S., Kakar, S., Yeh, M.M., Jain, D., Kleiner, D.E., Jain, D., Dhanasekaran, R., El-Serag, H.B., Yim, S.Y., Weinstein, J.N., Mishra, L., Zhang, Jianping, Akbani, R., Ling, S., Ju, Z., Su, X., Hegde, A.M., Mills, G.B., Lu, Y., Chen, J., Lee, J.-S., Sohn, B.H., Shim, J.J., Tong, P., Aburatani, H., Yamamoto, S., Tatsuno, K., Li, W., Xia, Z., Stransky, N., Seiser, E., Innocenti, F., Gao, J., Kundra, R., Zhang, Hongxin, Heins, Z., Ochoa, A., Sander, C., Ladanyi, M., Shen, R., Arora, A., Sanchez-Vega, F., Schultz, N., Kasaian, K., Radenbaugh, A., Bissig, K.-D., Moore, D.D., Totoki, Y., Nakamura, H., Shibata, T., Yau, C., Graim, K., Stuart, J., Haussler, D., Slagle, B.L., Ojesina, A.I., Katsonis, P., Koire, A., Lichtarge, O., Hsu, T.-K., Ferguson, M.L., Demchok, J.A., Felau, I., Sheth, M., Tarnuzzer, R., Wang, Z., Yang, L., Zenklusen, J.C., Zhang, Jiashan, Hutter, C.M., Sofia, H.J., Verhaak, R.G.W., Zheng, S., Lang, F., Chudamani, S., Liu, J., Lolla, L., Wu, Y., Naresh, R., Pihl, T., Sun, C., Wan, Y., Benz, C., Perou, A.H., Thorne, L.B., Boice, L., Huang, M., Rathmell, W.K., Noushmehr, H., Saggiaro, F.P., Tirapelli, D.P. da C., Junior, C.G.C., Mente, E.D., Silva, O. de C., Trevisan, F.A., Kang, K.J., Ahn, K.S., Giama, N.H., Moser, C.D., Giordano, T.J., Vinco, M., Welling, T.H., Crain, D., Curley, E., Gardner, J., Mallery, D., Morris, S., Paulauskis, J., Penny, R., Shelton, C., Shelton, T., Kelley, R., Park, J.-W., Chandan, V.S., Roberts, L.R., Bathe, O.F., Hagedorn, C.H., Auman, J.T., O'Brien, D.R., Kocher, J.-P.A., Jones, C.D., Mieczkowski, P.A., Perou, C.M., Skelly, T., Tan, D., Veluvolu, U., Balu, S., Bodenheimer, T., Hoyle, A.P., Jefferys, S.R., Meng, S., Mose, L.E., Shi, Y., Simons, J.V., Soloway, M.G., Roach, J., Hoadley, K.A., Baylin, S.B., Shen, H., Hinoue, T., Bootwalla, M.S., Van Den Berg, D.J., Weisenberger, D.J., Lai, P.H., Holbrook, A., Berrios, M., Laird, P.W., 2017a. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* 169, 1327-1341.e23. <https://doi.org/10.1016/j.cell.2017.05.046>
- Amaddeo, G., Cao, Q., Ladeiro, Y., Imbeaud, S., Nault, J.-C., Jaoui, D., Gaston Mathe, Y., Laurent, C., Laurent, A., Bioulac-Sage, P., Calderaro, J., Zucman-Rossi, J., 2015. Integration of tumour and viral genomic characterisations in HBV-related hepatocellular carcinomas. *Gut* 64, 820–829. <https://doi.org/10.1136/gutjnl-2013-306228>
- Bai, F., Yano, Y., Fukumoto, T., Takebe, A., Tanaka, M., Kuramitsu, K., Anggorowati, N., Rinonce, H.T., Widasari, D.I., Saito, M., Hirano, H., Hayakumo, T., Seo, Y., Azuma, T., Ku, Y., Hayashi, Y., 2013. Quantification of Pregenomic RNA and Covalently Closed Circular DNA in Hepatitis B Virus-Related Hepatocellular Carcinoma. *Int. J. Hepatol.* 2013, 1–9. <https://doi.org/10.1155/2013/849290>
- Bangham, C.R., Ratner, L., 2015. How does HTLV-1 cause adult T-cell leukaemia/lymphoma (ATL)? *Curr. Opin. Virol.* 14, 93–100. <https://doi.org/10.1016/j.coviro.2015.09.004>
- Bayard, Q., Meunier, L., Peneau, C., Renault, V., Shinde, J., Nault, J.-C., Mami, I., Couchy, G., Amaddeo, G., Tubacher, E., Bacq, D., Meyer, V., La Bella, T., Debaillon-Vesque, A., Bioulac-Sage, P., Seror, O., Blanc, J.-F., Calderaro, J., Deleuze, J.-F., Imbeaud, S., Zucman-Rossi, J., Letouzé, E., 2018. Cyclin A2/E1 activation defines a hepatocellular carcinoma subclass with a rearrangement signature of replication stress. *Nat. Commun.* 9, 5235. <https://doi.org/10.1038/s41467-018-07552-9>
- Baydoun, H.H., Bai, X.T., Shelton, S., Nicot, C., 2012. HTLV-I Tax Increases Genetic Instability by Inducing DNA Double Strand Breaks during DNA Replication and Switching Repair to NHEJ. *PLoS ONE* 7, e42226. <https://doi.org/10.1371/journal.pone.0042226>

- Bayliss, J., Lim, L., Thompson, A.J.V., Desmond, P., Angus, P., Locarnini, S., Revill, P.A., 2013. Hepatitis B virus splicing is enhanced prior to development of hepatocellular carcinoma. *J. Hepatol.* 59, 1022–1028. <https://doi.org/10.1016/j.jhep.2013.06.018>
- Becht, E., Giraldo, N.A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., Selves, J., Laurent-Puig, P., Sautès-Fridman, C., Fridman, W.H., de Reyniès, A., 2016. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17, 218. <https://doi.org/10.1186/s13059-016-1070-5>
- Benaglia, T., Chauveau, D., Hunter, D.R., Young, D., 2009. **mixtools** : An R Package for Analyzing Finite Mixture Models. *J. Stat. Softw.* 32. <https://doi.org/10.18637/jss.v032.i06>
- Benhenda, S., Cougot, D., Buendia, M.-A., Neuveut, C., 2009. Chapter 4 Hepatitis B Virus X Protein, in: *Advances in Cancer Research*. Elsevier, pp. 75–109. [https://doi.org/10.1016/S0065-230X\(09\)03004-8](https://doi.org/10.1016/S0065-230X(09)03004-8)
- Beroukhi, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., Mc Henry, K.T., Pinchback, R.M., Ligon, A.H., Cho, Y.-J., Haery, L., Greulich, H., Reich, M., Winckler, W., Lawrence, M.S., Weir, B.A., Tanaka, K.E., Chiang, D.Y., Bass, A.J., Loo, A., Hoffman, C., Prensner, J., Liefeld, T., Gao, Q., Yecies, D., Signoretti, S., Maher, E., Kaye, F.J., Sasaki, H., Tepper, J.E., Fletcher, J.A., Taberero, J., Baselga, J., Tsao, M.-S., Demichelis, F., Rubin, M.A., Janne, P.A., Daly, M.J., Nucera, C., Levine, R.L., Ebert, B.L., Gabriel, S., Rustgi, A.K., Antonescu, C.R., Ladanyi, M., Letai, A., Garraway, L.A., Loda, M., Beer, D.G., True, L.D., Okamoto, A., Pomeroy, S.L., Singer, S., Golub, T.R., Lander, E.S., Getz, G., Sellers, W.R., Meyerson, M., 2010. The landscape of somatic copy-number alteration across human cancers. *Nature* 463, 899–905. <https://doi.org/10.1038/nature08822>
- Bill, C.A., Summers, J., 2004. Genomic DNA double-strand breaks are targets for hepadnaviral DNA integration. *Proc. Natl. Acad. Sci.* 101, 11135–11140. <https://doi.org/10.1073/pnas.0403925101>
- Block, T.M., Alter, H.J., London, W.T., Bray, M., 2016. A historical perspective on the discovery and elucidation of the hepatitis B virus. *Antiviral Res.* 131, 109–123. <https://doi.org/10.1016/j.antiviral.2016.04.012>
- Blumberg, B.S., 2002. *Hepatitis B. The Hunt for a Killer Virus*, Pinceton University Press. ed.
- Borozan, I., Brewer, D.S., Iskar, M., Grundhoff, A., Alawi, M., Desai, N., Sülthmann, H., Moch, H., Cooper, C.S., Eils, R., Ferretti, V., Lichter, P., 2020. The landscape of viral associations in human cancers. *Nat. Genet.* 52, 320–330. <https://doi.org/10.1038/s41588-019-0558-9>
- Bouchard, M.J., Navas-Martin, S., 2011. Hepatitis B and C virus hepatocarcinogenesis: Lessons learned and future challenges. *Cancer Lett.* 305, 123–143. <https://doi.org/10.1016/j.canlet.2010.11.014>
- Bou-Nader, M., Caruso, S., Donne, R., Celton-Morizur, S., Calderaro, J., Gentric, G., Cadoux, M., L'Hermitte, A., Klein, C., Guilbert, T., Albuquerque, M., Couchy, G., Paradis, V., Couty, J.-P., Zucman-Rossi, J., Desdouets, C., 2020. Polyploidy spectrum: a new marker in HCC classification. *Gut* 69, 355–364. <https://doi.org/10.1136/gutjnl-2018-318021>
- Boutin, S., Monteilhet, V., Veron, P., Leborgne, C., Benveniste, O., Montus, M.F., Masurier, C., 2010. Prevalence of Serum IgG and Neutralizing Factors Against Adeno-Associated Virus (AAV) Types 1, 2, 5, 6, 8, and 9 in the Healthy Population: Implications for Gene Therapy Using AAV Vectors. *Hum. Gene Ther.* 21, 704–712. <https://doi.org/10.1089/hum.2009.182>
- Bouvard, V., Baan, R., Straif, K., Grosse, Y., Secretan, B., Ghissassi, F.E., Benbrahim-Tallaa, L., Guha, N., Freeman, C., Galichet, L., Coglianò, V., 2009. A review of human carcinogens—Part B: biological agents. *Lancet Oncol.* 10, 321–322. [https://doi.org/10.1016/S1470-2045\(09\)70096-8](https://doi.org/10.1016/S1470-2045(09)70096-8)

BIBLIOGRAPHIE

- Boyault, S., Rickman, D.S., de Reyniès, A., Balabaud, C., Rebouissou, S., Jeannot, E., Hérault, A., Saric, J., Belghiti, J., Franco, D., Bioulac-Sage, P., Laurent-Puig, P., Zucman-Rossi, J., 2007. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* 45, 42–52. <https://doi.org/10.1002/hep.21467>
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* 68, 394–424. <https://doi.org/10.3322/caac.21492>
- Bréchet, C., Gozuacik, D., Murakami, Y., Paterlini-Bréchet, P., 2000. Molecular bases for the development of hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC). *Semin. Cancer Biol.* 10, 211–231. <https://doi.org/10.1006/scbi.2000.0321>
- Brechot, C., Pourcel, C., Louise, A., Rain, B., Tiollais, P., 1980. Presence of integrated hepatitis B virus DNA sequences in cellular DNA of human hepatocellular carcinoma. *Nature* 286, 533–535. <https://doi.org/10.1038/286533a0>
- Bruix, J., Qin, S., Merle, P., Granito, A., Huang, Y.-H., Bodoky, G., Pracht, M., Yokosuka, O., Rosmorduc, O., Breder, V., Gerolami, R., Masi, G., Ross, P.J., Song, T., Bronowicki, J.-P., Ollivier-Hourmand, I., Kudo, M., Cheng, A.-L., Llovet, J.M., Finn, R.S., LeBerre, M.-A., Baumhauer, A., Meinhardt, G., Han, G., 2017. Regorafenib for patients with hepatocellular carcinoma who progressed on sorafenib treatment (RESORCE): a randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet* 389, 56–66. [https://doi.org/10.1016/S0140-6736\(16\)32453-9](https://doi.org/10.1016/S0140-6736(16)32453-9)
- Budzinska, Magdalena A., Shackel, N.A., Urban, S., Tu, T., 2018a. Cellular Genomic Sites of Hepatitis B Virus DNA Integration. *Genes* 9, 365. <https://doi.org/10.3390/genes9070365>
- Budzinska, Magdalena Agnieszka, Shackel, N.A., Urban, S., Tu, T., 2018. Sequence analysis of integrated hepatitis B virus DNA during HBeAg-seroconversion. *Emerg. Microbes Infect.* 7, 1–12. <https://doi.org/10.1038/s41426-018-0145-7>
- Budzinska, Magdalena A., Urban, S., Tu, T., 2018b. Sequence analysis of integrated hepatitis B virus DNA during HBeAg-seroconversion. *Emerg. Microbes Infect.* 7, 1–12. <https://doi.org/10.1038/s41426-018-0145-7>
- Calado, R.T., Young, N.S., 2009. Telomere Diseases. *N. Engl. J. Med.* 361, 2353–2365. <https://doi.org/10.1056/NEJMra0903373>
- Calderaro, J., Ziol, M., Paradis, V., Zucman-Rossi, J., 2019. Molecular and histological correlations in liver cancer. *J. Hepatol.* 71, 616–630. <https://doi.org/10.1016/j.jhep.2019.06.001>
- Caruso, S., Calatayud, A.-L., Pilet, J., La Bella, T., Rekik, S., Imbeaud, S., Letouzé, E., Meunier, L., Bayard, Q., Rohr-Udilova, N., Péneau, C., Grasl-Kraupp, B., de Koning, L., Ouine, B., Bioulac-Sage, P., Couchy, G., Calderaro, J., Nault, J.-C., Zucman-Rossi, J., Rebouissou, S., 2019. Analysis of Liver Cancer Cell Lines Identifies Agents With Likely Efficacy Against Hepatocellular Carcinoma and Markers of Response. *Gastroenterology* 157, 760–776. <https://doi.org/10.1053/j.gastro.2019.05.001>
- Chen, C., Yang, Y., Lin, M., Lee, C., Tsan, Y., Lai, M., Yang, H., Ho, W., Chen, P., The Health Data Analysis in Taiwan (hDATA) Research Group, 2018. Herbal medicine containing aristolochic acid and the risk of hepatocellular carcinoma in patients with hepatitis B virus infection. *Int. J. Cancer* 143, 1578–1587. <https://doi.org/10.1002/ijc.31544>
- Chen, M., Sallberg, M., Hughes, J., Jones, J., Guidotti, L.G., Chisari, F.V., Billaud, J.-N., Milich, D.R., 2005. Immune Tolerance Split between Hepatitis B Virus Precore and Core Proteins. *J. Virol.* 79, 3016–3027. <https://doi.org/10.1128/JVI.79.5.3016-3027.2005>

- Chen, Y., Williams, V., Filippova, M., Filippov, V., Duerksen-Hughes, P., 2014. Viral Carcinogenesis: Factors Inducing DNA Damage and Virus Integration. *Cancers* 6, 2155–2186. <https://doi.org/10.3390/cancers6042155>
- Chiu, Y.-T., Wong, J.K.L., Choi, S.-W., Sze, K.M.F., Ho, D.W.H., Chan, L.-K., Lee, J.M.F., Man, K., Cherny, S., Yang, W.-L., Wong, C.-M., Sham, P.-C., Ng, I.O.L., 2016. Novel pre-mRNA splicing of intronically integrated HBV generates oncogenic chimera in hepatocellular carcinoma. *J. Hepatol.* 64, 1256–1264. <https://doi.org/10.1016/j.jhep.2016.02.005>
- Coppola, N., Onorato, L., Iodice, V., Starace, M., Minichini, C., Farella, N., Liorre, G., Filippini, P., Sagnelli, E., de Stefano, G., 2016. Occult HBV infection in HCC and cirrhotic tissue of HBsAg-negative patients: a virological and clinical study. *Oncotarget* 7, 62706–62714. <https://doi.org/10.18632/oncotarget.10909>
- Croagh, C.M., 2014. Natural history of chronic hepatitis B: Phases in a complex relationship. *World J. Gastroenterol.* 20, 10395. <https://doi.org/10.3748/wjg.v20.i30.10395>
- Dandri, M., 2002. Increase in de novo HBV DNA integrations in response to oxidative DNA damage or inhibition of poly(ADP-ribosyl)ation. *Hepatology* 35, 217–223. <https://doi.org/10.1053/jhep.2002.30203>
- Dane, D., Cameron, C., Briggs, M., 1970. Virus-like particles in serum of patients with Australia-antigen-associated hepatitis. *Lancet*.
- Davison, S.A., Strasser, S.I., 2014. Ordering and interpreting hepatitis B serology. *BMJ* 348, g2522–g2522. <https://doi.org/10.1136/bmj.g2522>
- De Flora, S., La Maestra, S., 2015. Epidemiology of cancers of infectious origin and prevention strategies.
- de Martel, C., Georges, D., Bray, F., Ferlay, J., Clifford, G.M., 2020. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. *Lancet Glob. Health* 8, e180–e190. [https://doi.org/10.1016/S2214-109X\(19\)30488-7](https://doi.org/10.1016/S2214-109X(19)30488-7)
- Debacker, K., Kooy, R.F., 2007. Fragile sites and human disease. *Hum. Mol. Genet.* 16, R150–R158. <https://doi.org/10.1093/hmg/ddm136>
- Decorsière, A., Mueller, H., van Breugel, P.C., Abdul, F., Gerossier, L., Beran, R.K., Livingston, C.M., Niu, C., Fletcher, S.P., Hantz, O., Strubin, M., 2016. Hepatitis B virus X protein identifies the Smc5/6 complex as a host restriction factor. *Nature* 531, 386–389. <https://doi.org/10.1038/nature17170>
- Dény, P., Zoulim, F., 2010. Hepatitis B virus: From diagnosis to treatment. *Pathol. Biol.* 58, 245–253. <https://doi.org/10.1016/j.patbio.2010.05.002>
- Deshpande, A., Sicinski, P., Hinds, P.W., 2005. Cyclins and cdks in development and cancer: a perspective. *Oncogene* 24, 2909–2915. <https://doi.org/10.1038/sj.onc.1208618>
- Di Tommaso, L., Sangiovanni, A., Borzio, M., Park, Y.N., Farinati, F., Roncalli, M., 2013. Advanced precancerous lesions in the liver. *Best Pract. Res. Clin. Gastroenterol.* 27, 269–284. <https://doi.org/10.1016/j.bpg.2013.03.015>
- Dill, J.A., Camus, A.C., Leary, J.H., Di Giallonardo, F., Holmes, E.C., Ng, T.F.F., 2016. Distinct Viral Lineages from Fish and Amphibians Reveal the Complex Evolutionary History of Hepadnaviruses. *J. Virol.* 90, 7920–7933. <https://doi.org/10.1128/JVI.00832-16>
- Ding, D., Lou, X., Hua, D., Yu, W., Li, Lisha, Wang, J., Gao, F., Zhao, N., Ren, G., Li, Lanjuan, Lin, B., 2012. Recurrent Targeted Genes of Hepatitis B Virus in the Liver Cancer Genomes Identified by a Next-Generation Sequencing–Based Approach. *PLoS Genet.* 8, e1003065. <https://doi.org/10.1371/journal.pgen.1003065>

BIBLIOGRAPHIE

- Dong, H., Zhang, L., Qian, Z., Zhu, X., Zhu, G., Chen, Y., Xie, X., Ye, Q., Zang, J., Ren, Z., Ji, Q., 2015. Identification of HBV-MLL4 Integration and Its Molecular Basis in Chinese Hepatocellular Carcinoma. *PLOS ONE* 10, e0123175. <https://doi.org/10.1371/journal.pone.0123175>
- Duan, M., Hao, J., Cui, S., Worthley, D.L., Zhang, S., Wang, Z., Shi, J., Liu, L., Wang, X., Ke, A., Cao, Y., Xi, R., Zhang, X., Zhou, J., Fan, J., Li, C., Gao, Q., 2018. Diverse modes of clonal evolution in HBV-related hepatocellular carcinoma revealed by single-cell genome sequencing. *Cell Res.* 28, 359–373. <https://doi.org/10.1038/cr.2018.11>
- Ducancelle, A., Pivert, A., Bertrais, S., Boursier, J., Balan, V., Veillon, P., le Guillou-Guillemette, H., Thibault, V., Castelain, S., Roquebert, B., Coste-Burel, M., Mackiewicz, V., Schvoerer, E., Larrat, S., Maylin, S., Alain, S., Loustaud-Ratti, V., Gordien, E., Gozlan, J., Brodard, V., Chevaliez, S., Calès, P., Lunel-Fabiani, F., 2016. Different precore/core mutations of hepatitis B interact with, limit, or favor liver fibrosis severity: HBV precore mutants. *J. Gastroenterol. Hepatol.* 31, 1750–1756. <https://doi.org/10.1111/jgh.13338>
- Duriez, M., Mandouri, Y., Lekbaby, B., Wang, H., Schnuriger, A., Redelsperger, F., Guerrero, C.I., Lefevre, M., Fauveau, V., Ahodantin, J., Quetier, I., Chhuon, C., Gourari, S., Boissonnas, A., Gill, U., Kennedy, P., Debzi, N., Sitterlin, D., Maini, M.K., Kremsdorf, D., Soussan, P., 2017. Alternative splicing of hepatitis B virus: A novel virus/host interaction altering liver immunity. *J. Hepatol.* 67, 687–699. <https://doi.org/10.1016/j.jhep.2017.05.025>
- El-Khoueiry, A.B., Sangro, B., Yau, T., Crocenzi, T.S., Kudo, M., Hsu, C., Kim, T.-Y., Choo, S.-P., Trojan, J., Welling, T.H., Meyer, T., Kang, Y.-K., Yeo, W., Chopra, A., Anderson, J., dela Cruz, C., Lang, L., Neely, J., Tang, H., Dastani, H.B., Melero, I., 2017. Nivolumab in patients with advanced hepatocellular carcinoma (CheckMate 040): an open-label, non-comparative, phase 1/2 dose escalation and expansion trial. *The Lancet* 389, 2492–2502. [https://doi.org/10.1016/S0140-6736\(17\)31046-2](https://doi.org/10.1016/S0140-6736(17)31046-2)
- Fallot, G., Neuveut, C., Buendia, M.-A., 2012. Diverse roles of hepatitis B virus in liver cancer. *Curr. Opin. Virol.* 2, 467–473. <https://doi.org/10.1016/j.coviro.2012.05.008>
- Fanning, G.C., Zoulim, F., Hou, J., Bertoletti, A., 2019. Therapeutic strategies for hepatitis B virus infection: towards a cure. *Nat. Rev. Drug Discov.* 18, 827–844. <https://doi.org/10.1038/s41573-019-0037-0>
- Farazi, P.A., DePinho, R.A., 2006. Hepatocellular carcinoma pathogenesis: from genes to environment. *Nat. Rev. Cancer* 6, 674–687. <https://doi.org/10.1038/nrc1934>
- Fattovich, G., Stroffolini, T., Zagni, I., Donato, F., 2004. Hepatocellular carcinoma in cirrhosis: Incidence and risk factors. *Gastroenterology* 127, S35–S50. <https://doi.org/10.1053/j.gastro.2004.09.014>
- Feng, H., Shuda, M., Chang, Y., Moore, P.S., 2008. Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma. *Science* 319, 1096–1100. <https://doi.org/10.1126/science.1152586>
- Finn, R.S., Qin, S., Ikeda, M., Galle, P.R., Ducreux, M., Kim, T.-Y., Kudo, M., Breder, V., Merle, P., Kaseb, A.O., Li, D., Verret, W., Xu, D.-Z., Hernandez, S., Liu, J., Huang, C., Mulla, S., Wang, Y., Lim, H.Y., Zhu, A.X., Cheng, A.-L., 2020. Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma. *N. Engl. J. Med.* 382, 1894–1905. <https://doi.org/10.1056/NEJMoa1915745>
- Freitas, N., Cunha, C., Menne, S., Gudima, S.O., 2014. Envelope Proteins Derived from Naturally Integrated Hepatitis B Virus DNA Support Assembly and Release of Infectious Hepatitis Delta Virus Particles. *J. Virol.* 88, 5742–5754. <https://doi.org/10.1128/JVI.00430-14>
- Fruman, D.A., Rommel, C., 2014. PI3K and cancer: lessons, challenges and opportunities. *Nat. Rev. Drug Discov.* 13, 140–156. <https://doi.org/10.1038/nrd4204>

- Fujimoto, A., Furuta, M., Totoki, Y., Tsunoda, T., Kato, M., Shiraishi, Y., Tanaka, H., Taniguchi, H., Kawakami, Y., Ueno, M., Gotoh, K., Ariizumi, S., Wardell, C.P., Hayami, S., Nakamura, T., Aikata, H., Arihiro, K., Boroevich, K.A., Abe, T., Nakano, K., Maejima, K., Sasaki-Oku, A., Ohsawa, A., Shibuya, T., Nakamura, H., Hama, N., Hosoda, F., Arai, Y., Ohashi, S., Urushidate, T., Nagae, G., Yamamoto, S., Ueda, H., Tatsuno, K., Ojima, H., Hiraoka, N., Okusaka, T., Kubo, M., Marubashi, S., Yamada, T., Hirano, S., Yamamoto, M., Ohdan, H., Shimada, K., Ishikawa, O., Yamaue, H., Chayama, K., Miyano, S., Aburatani, H., Shibata, T., Nakagawa, H., 2016. Whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat. Genet.* 48, 500–509. <https://doi.org/10.1038/ng.3547>
- Fujimoto, A., Totoki, Y., Abe, T., Boroevich, K.A., Hosoda, F., Nguyen, H.H., Aoki, M., Hosono, N., Kubo, M., Miya, F., Arai, Y., Takahashi, H., Shirakihara, T., Nagasaki, M., Shibuya, T., Nakano, K., Watanabe-Makino, K., Tanaka, H., Nakamura, H., Kusuda, J., Ojima, H., Shimada, K., Okusaka, T., Ueno, M., Shigekawa, Y., Kawakami, Y., Arihiro, K., Ohdan, H., Gotoh, K., Ishikawa, O., Ariizumi, S., Yamamoto, M., Yamada, T., Chayama, K., Kosuge, T., Yamaue, H., Kamatani, N., Miyano, S., Nakagawa, H., Nakamura, Y., Tsunoda, T., Shibata, T., Nakagawa, H., 2012. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* 44, 760–764. <https://doi.org/10.1038/ng.2291>
- Furuta, M., Tanaka, H., Shiraishi, Y., Uchida, T., Imamura, M., Fujimoto, A., Fujita, M., Sasaki-Oku, A., Maejima, K., Nakano, K., Kawakami, Y., Arihiro, K., Aikata, H., Ueno, M., Hayami, S., Ariizumi, S.-I., Yamamoto, M., Gotoh, K., Ohdan, H., Yamaue, H., Miyano, S., Chayama, K., Nakagawa, H., 2018. Characterization of HBV integration patterns and timing in liver cancer and HBV-infected livers. *Oncotarget* 9. <https://doi.org/10.18632/oncotarget.25308>
- Gaglia, M.M., Munger, K., 2018. More than just oncogenes: mechanisms of tumorigenesis by human viruses. *Curr. Opin. Virol.* 32, 48–59. <https://doi.org/10.1016/j.coviro.2018.09.003>
- Galle, P.R., Forner, A., Llovet, J.M., Mazzaferro, V., Piscaglia, F., Raoul, J.-L., Schirmacher, P., Vilgrain, V., 2018. EASL Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J. Hepatol.* 69, 182–236. <https://doi.org/10.1016/j.jhep.2018.03.019>
- Ganesan, M., Poluektova, L.Y., Kharbanda, K.K., Osna, N.A., 2019. Human immunodeficiency virus and hepatotropic viruses co-morbidities as the inducers of liver injury progression. *World J. Gastroenterol.* 25, 398–410. <https://doi.org/10.3748/wjg.v25.i4.398>
- Ganne-Carrié, N., Nahon, P., 2019. Hepatocellular carcinoma in the setting of alcohol-related liver disease. *J. Hepatol.* 70, 284–293. <https://doi.org/10.1016/j.jhep.2018.10.008>
- Gao, G., Vandenberghe, L.H., Alvira, M.R., Lu, Y., Calcedo, R., Zhou, X., Wilson, J.M., 2004. Clades of Adeno-Associated Viruses Are Widely Disseminated in Human Tissues. *J. Virol.* 78, 6381–6388. <https://doi.org/10.1128/JVI.78.12.6381-6388.2004>
- Gardner, R., Sutherland, G., Shaffer, L., 2011. *Chromosome Abnormalities and Genetic Counseling: Fourth Edition.*
- Glebe, D., 2007. Viral and cellular determinants involved in hepadnaviral entry. *World J. Gastroenterol.* 13, 22. <https://doi.org/10.3748/wjg.v13.i1.22>
- Global Burden of Disease Liver Cancer Collaboration, Akinyemiju, T., Abera, S., Ahmed, M., Alam, N., Alemayohu, M.A., Allen, C., Al-Raddadi, R., Alvis-Guzman, N., Amoako, Y., Artaman, A., Ayele, T.A., Barac, A., Bensenor, I., Berhane, A., Bhutta, Z., Castillo-Rivas, J., Chittheer, A., Choi, J.-Y., Cowie, B., Dandona, L., Dandona, R., Dey, S., Dicker, D., Phuc, H., Ekwueme, D.U., Zaki, M.E.S., Fischer, F., Fürst, T., Hancock, J., Hay, S.I., Hotez, P., Jee, S.H., Kasaeian, A., Khader, Y., Khang, Y.-H., Kumar, G.A., Kutz, M., Larson, H., Lopez, A., Lunevicius, R., Malekzadeh, R., McAlinden, C., Meier, T., Mendoza, W., Mokdad, A., Moradi-Lakeh, M., Nagel, G., Nguyen, Q., Nguyen, G., Ogbo, F., Patton, G., Pereira, D.M., Pourmalek, F., Qorbani, M., Radfar, A., Roshandel, G., Salomon, J.A., Sanabria, J., Sartorius,

BIBLIOGRAPHIE

- B., Satpathy, M., Sawhney, M., Sepanlou, S., Shackelford, K., Shore, H., Sun, J., Mengistu, D.T., Topór-Madry, R., Tran, B., Ukwaja, K.N., Vlassov, V., Vollset, S.E., Vos, T., Wakayo, T., Weiderpass, E., Werdecker, A., Yonemoto, N., Younis, M., Yu, C., Zaidi, Z., Zhu, L., Murray, C.J.L., Naghavi, M., Fitzmaurice, C., 2017. The Burden of Primary Liver Cancer and Underlying Etiologies From 1990 to 2015 at the Global, Regional, and National Level: Results From the Global Burden of Disease Study 2015. *JAMA Oncol.* 3, 1683. <https://doi.org/10.1001/jamaoncol.2017.3055>
- Gonçalves, M.A., 2005. [No title found]. *Viol. J.* 2, 43. <https://doi.org/10.1186/1743-422X-2-43>
- Goodwin, E.C., Yang, E., Lee, C.-J., Lee, H.-W., DiMaio, D., Hwang, E.-S., 2000. Rapid induction of senescence in human cervical carcinoma cells. *Proc. Natl. Acad. Sci.* 97, 10978–10983. <https://doi.org/10.1073/pnas.97.20.10978>
- Gouas, D., Shi, H., Hainaut, P., 2009. The aflatoxin-induced TP53 mutation at codon 249 (R249S): Biomarker of exposure, early detection and target for therapy. *Cancer Lett.* 286, 29–37. <https://doi.org/10.1016/j.canlet.2009.02.057>
- Graber-Stiehl, I., 2018. Hepatitis now kills more people worldwide than HIV, tuberculosis or malaria. Tackling the hepatitis B virus in Africa is key to fighting back. 4. <https://doi.org/10.1038/d41586-018-07592-7>
- Guichard, C., Amaddeo, G., Imbeaud, S., Ladeiro, Y., Pelletier, L., Maad, I.B., Calderaro, J., Bioulac-Sage, P., Letexier, M., Degos, F., Clément, B., Balabaud, C., Chevet, E., Laurent, A., Couchy, G., Letouzé, E., Calvo, F., Zucman-Rossi, J., 2012. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat. Genet.* 44, 694–698. <https://doi.org/10.1038/ng.2256>
- Hadziyannis, E., Laras, A., 2018. Viral Biomarkers in Chronic HBeAg Negative HBV Infection. *Genes* 9, 469. <https://doi.org/10.3390/genes9100469>
- Halgand, B., Desterke, C., Rivièrè, L., Fallot, G., Sebagh, M., Calderaro, J., Bioulac-Sage, P., Neuveut, C., Buendía, M.-A., Samuel, D., Féray, C., 2018. Hepatitis B Virus Pregenomic RNA in Hepatocellular Carcinoma: A Nosological and Prognostic Determinant: Halgand et al. *Hepatology* 67, 86–96. <https://doi.org/10.1002/hep.29463>
- Harden, M.E., Munger, K., 2017. Human papillomavirus molecular biology. *Mutat. Res. Mutat. Res.* 772, 3–12. <https://doi.org/10.1016/j.mrrev.2016.07.002>
- Hepatitis B Foundation, 2020. Hepatitis B Foundation. Drug watch.
- Hernandez-Gea, V., Friedman, S.L., 2011. Pathogenesis of Liver Fibrosis. *Annu. Rev. Pathol. Mech. Dis.* 6, 425–456. <https://doi.org/10.1146/annurev-pathol-011110-130246>
- Hirsch, T.Z., Negulescu, A., Gupta, B., Caruso, S., Noblet, B., Couchy, G., Bayard, Q., Meunier, L., Morcrette, G., Scoazec, J.-Y., Blanc, J.-F., Amaddeo, G., Nault, J.-C., Bioulac-Sage, P., Ziol, M., Beaufrère, A., Paradis, V., Calderaro, J., Imbeaud, S., Zucman-Rossi, J., 2019. BAP1 mutations define a homogeneous subgroup of hepatocellular carcinoma with fibrolamellar-like features and activated PKA. *J. Hepatol.* S0168827819307184. <https://doi.org/10.1016/j.jhep.2019.12.006>
- Holmes, A., Lameiras, S., Jeannot, E., Marie, Y., Castera, L., Sastre-Garau, X., Nicolas, A., 2016. Mechanistic signatures of HPV insertions in cervical carcinomas. *Npj Genomic Med.* 1, 16004. <https://doi.org/10.1038/npjgenmed.2016.4>
- Hosaka, T., Suzuki, F., Kobayashi, Masahiro, Seko, Y., Kawamura, Y., Sezaki, H., Akuta, N., Suzuki, Y., Saitoh, S., Arase, Y., Ikeda, K., Kobayashi, Mariko, Kumada, H., 2013. Long-term entecavir treatment reduces hepatocellular carcinoma incidence in patients with hepatitis B virus infection. *Hepatology* 58, 98–107. <https://doi.org/10.1002/hep.26180>

- Hoshida, Y., 2009. Risk of recurrence in hepatitis B-related hepatocellular carcinoma: Impact of viral load in late recurrence. *J. Hepatol.* 51, 842–844. <https://doi.org/10.1016/j.jhep.2009.08.003>
- Hoshida, Y., Nijman, S.M.B., Kobayashi, M., Chan, J.A., Brunet, J.-P., Chiang, D.Y., Villanueva, A., Newell, P., Ikeda, K., Hashimoto, M., Watanabe, G., Gabriel, S., Friedman, S.L., Kumada, H., Llovet, J.M., Golub, T.R., 2009. Integrative Transcriptome Analysis Reveals Common Molecular Subclasses of Human Hepatocellular Carcinoma. *Cancer Res.* 69, 7385–7392. <https://doi.org/10.1158/0008-5472.CAN-09-1089>
- Hu, J., Seeger, C., 2015a. Hepadnavirus Genome Replication and Persistence. *Cold Spring Harb. Perspect. Med.* 5, a021386. <https://doi.org/10.1101/cshperspect.a021386>
- Hu, J., Seeger, C., 2015b. Hepadnavirus Genome Replication and Persistence. *Cold Spring Harb. Perspect. Med.* 5, a021386. <https://doi.org/10.1101/cshperspect.a021386>
- Hu, Z., Zhu, D., Wang, W., Li, W., Jia, W., Zeng, X., Ding, W., Yu, L., Wang, X., Wang, L., Shen, H., Zhang, C., Liu, H., Liu, X., Zhao, Y., Fang, X., Li, Shuaicheng, Chen, W., Tang, T., Fu, A., Wang, Z., Chen, G., Gao, Q., Li, Shuang, Xi, L., Wang, C., Liao, S., Ma, X., Wu, P., Li, K., Wang, S., Zhou, J., Wang, J., Xu, X., Wang, H., Ma, D., 2015. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat. Genet.* 47, 158–163. <https://doi.org/10.1038/ng.3178>
- Jemal, A., Ward, E.M., Johnson, C.J., Cronin, K.A., Ma, J., Ryerson, A.B., Mariotto, A., Lake, A.J., Wilson, R., Sherman, R.L., Anderson, R.N., Henley, S.J., Kohler, B.A., Penberthy, L., Feuer, E.J., Weir, H.K., 2017. Annual Report to the Nation on the Status of Cancer, 1975–2014, Featuring Survival. *JNCI J. Natl. Cancer Inst.* 109. <https://doi.org/10.1093/jnci/djx030>
- Jhunjunwala, S., Jiang, Z., Stawiski, E.W., Gnad, F., Liu, J., Mayba, O., Du, P., Diao, J., Johnson, S., Wong, K.-F., Gao, Z., Li, Y., Wu, T.D., Kapadia, S.B., Modrusan, Z., French, D.M., Luk, J.M., Seshagiri, S., Zhang, Z., 2014. Diverse modes of genomic alteration in hepatocellular carcinoma. *Genome Biol.* 15, 436. <https://doi.org/10.1186/s13059-014-0436-9>
- Jiang, S., Yang, Z., Li, W., Li, X., Wang, Y., Zhang, J., Xu, C., Chen, P.-J., Hou, J., McCrae, M.A., Chen, X., Zhuang, H., Lu, F., 2012. Re-evaluation of the Carcinogenic Significance of Hepatitis B Virus Integration in Hepatocarcinogenesis. *PLoS ONE* 7, e40363. <https://doi.org/10.1371/journal.pone.0040363>
- Jiang, Z., Jhunjunwala, S., Liu, J., Haverty, P.M., Kennemer, M.I., Guan, Y., Lee, W., Carnevali, P., Stinson, J., Johnson, S., Diao, J., Yeung, S., Jubb, A., Ye, W., Wu, T.D., Kapadia, S.B., de Sauvage, F.J., Gentleman, R.C., Stern, H.M., Seshagiri, S., Pant, K.P., Modrusan, Z., Ballinger, D.G., Zhang, Z., 2012. The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. *Genome Res.* 22, 593–601. <https://doi.org/10.1101/gr.133926.111>
- Kao, J.-H., Chen, P.-J., Chen, D.-S., 2010. Recent Advances in the Research of Hepatitis B Virus-Related Hepatocellular Carcinoma, in: *Advances in Cancer Research*. Elsevier, pp. 21–72. <https://doi.org/10.1016/B978-0-12-380888-2.00002-9>
- Kawai-Kitahata, F., Asahina, Y., Tanaka, S., Kakinuma, S., Murakawa, M., Nitta, S., Watanabe, T., Otani, S., Taniguchi, M., Goto, F., Nagata, H., Kaneko, S., Tasaka-Fujita, M., Nishimura-Sakurai, Y., Azuma, S., Itsui, Y., Nakagawa, M., Tanabe, M., Takano, S., Fukasawa, M., Sakamoto, M., Maekawa, S., Enomoto, N., Watanabe, M., 2016. Comprehensive analyses of mutations and hepatitis B virus integration in hepatocellular carcinoma with clinicopathological features. *J. Gastroenterol.* 51, 473–486. <https://doi.org/10.1007/s00535-015-1126-4>
- Kay, A., Zoulim, F., 2007. Hepatitis B virus genetic variability and evolution. *Virus Res.* 127, 164–176. <https://doi.org/10.1016/j.virusres.2007.02.021>

BIBLIOGRAPHIE

- Kew, M., 2014. Hepatocellular carcinoma: epidemiology and risk factors. *J. Hepatocell. Carcinoma* 115. <https://doi.org/10.2147/JHC.S44381>
- Kew, M.C., 2011. Hepatitis B virus x protein in the pathogenesis of hepatitis B virus-induced hepatocellular carcinoma: HBx and HCC. *J. Gastroenterol. Hepatol.* 26, 144–152. <https://doi.org/10.1111/j.1440-1746.2010.06546.x>
- Kew, M.C., 2003. Synergistic interaction between aflatoxin B1 and hepatitis B virus in hepatocarcinogenesis. *Liver Int.* 23, 405–409. <https://doi.org/10.1111/j.1478-3231.2003.00869.x>
- Kidd-Ljunggren, K., Holmberg, A., Bläckberg, J., Lindqvist, B., 2006. High levels of hepatitis B virus DNA in body fluids from chronic carriers. *J. Hosp. Infect.* 64, 352–357. <https://doi.org/10.1016/j.jhin.2006.06.029>
- Kim, G.-A., Lee, H.C., Kim, M.-J., Ha, Y., Park, E.J., An, J., Lee, D., Shim, J.H., Kim, K.M., Lim, Y.-S., 2015. Incidence of hepatocellular carcinoma after HBsAg seroclearance in chronic hepatitis B patients: A need for surveillance. *J. Hepatol.* 62, 1092–1099. <https://doi.org/10.1016/j.jhep.2014.11.031>
- Kim, J.H., Kim, Y.D., Lee, M., Jun, B.G., Kim, T.S., Suk, K.T., Kang, S.H., Kim, M.Y., Cheon, G.J., Kim, D.J., Baik, S.K., Choi, D.H., 2018. Modified PAGE-B score predicts the risk of hepatocellular carcinoma in Asians with chronic hepatitis B on antiviral therapy. *J. Hepatol.* 69, 1066–1073. <https://doi.org/10.1016/j.jhep.2018.07.018>
- Kimbi, G.C., 2005. Integration of hepatitis B virus DNA into chromosomal DNA during acute hepatitis B. *World J. Gastroenterol.* 11, 6416. <https://doi.org/10.3748/wjg.v11.i41.6416>
- Kitajima, S., Takahashi, C., 2017. Intersection of retinoblastoma tumor suppressor function, stem cells, metabolism, and inflammation. *Cancer Sci.* 108, 1726–1731. <https://doi.org/10.1111/cas.13312>
- Koh, C., Da, B.L., Glenn, J.S., 2019. HBV/HDV Coinfection. *Clin. Liver Dis.* 23, 557–572. <https://doi.org/10.1016/j.cld.2019.04.005>
- Kondo, Y., Ninomiya, M., Kakazu, E., Kimura, O., Shimosegawa, T., 2013. Hepatitis B Surface Antigen Could Contribute to the Immunopathogenesis of Hepatitis B Virus Infection. *ISRN Gastroenterol.* 2013, 1–8. <https://doi.org/10.1155/2013/935295>
- Konstantinou, D., Deutsch, M., 2015. The spectrum of HBV/HCV coinfection: epidemiology, clinical characteristics, viral interactions and management. *Ann. Gastroenterol.* 8.
- Kotterman, M.A., Schaffer, D.V., 2014. Engineering adeno-associated viruses for clinical gene therapy. *Nat. Rev. Genet.* 15, 445–451. <https://doi.org/10.1038/nrg3742>
- Kramvis, A., 2014. Genotypes and Genetic Variability of Hepatitis B Virus. *Intervirology* 57, 141–150. <https://doi.org/10.1159/000360947>
- Kudo, M., Finn, R.S., Qin, S., Han, K.-H., Ikeda, K., Piscaglia, F., Baron, A., Park, J.-W., Han, G., Jassem, J., Blanc, J.F., Vogel, A., Komov, D., Evans, T.R.J., Lopez, C., Dutcus, C., Guo, M., Saito, K., Kraljevic, S., Tamai, T., Ren, M., Cheng, A.-L., 2018. Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial. *The Lancet* 391, 1163–1173. [https://doi.org/10.1016/S0140-6736\(18\)30207-1](https://doi.org/10.1016/S0140-6736(18)30207-1)
- Kulik, L., El-Serag, H.B., 2019. Epidemiology and Management of Hepatocellular Carcinoma. *Gastroenterology* 156, 477–491.e1. <https://doi.org/10.1053/j.gastro.2018.08.065>
- La Bella, T., Imbeaud, S., Peneau, C., Mami, I., Datta, S., Bayard, Q., Caruso, S., Hirsch, T.Z., Calderaro, J., Morcrette, G., Guettier, C., Paradis, V., Amaddeo, G., Laurent, A., Possenti, L.,

- Chiche, L., Bioulac-Sage, P., Blanc, J.-F., Letouze, E., Nault, J.-C., Zucman-Rossi, J., 2020. Adeno-associated virus in the liver: natural history and consequences in tumour development. *Gut* 69, 737–747. <https://doi.org/10.1136/gutjnl-2019-318281>
- Lamontagne, R.J., Bagga, S., Bouchard, M.J., 2016. Hepatitis B virus molecular biology and pathogenesis. *Hepatoma Res.* 2, 163. <https://doi.org/10.20517/2394-5079.2016.05>
- Lampertico, P., Agarwal, K., Berg, T., Buti, M., Janssen, H.L.A., Papatheodoridis, G., Zoulim, F., Tacke, F., 2017. EASL 2017 Clinical Practice Guidelines on the management of hepatitis B virus infection. *J. Hepatol.* 67, 370–398. <https://doi.org/10.1016/j.jhep.2017.03.021>
- Lau, C.-C., Sun, T., Ching, A.K.K., He, M., Li, J.-W., Wong, A.M., Co, N.N., Chan, A.W.H., Li, P.-S., Lung, R.W.M., Tong, J.H.M., Lai, P.B.S., Chan, H.L.Y., To, K.-F., Chan, T.-F., Wong, N., 2014. Viral-Human Chimeric Transcript Predisposes Risk to Liver Cancer Development and Progression. *Cancer Cell* 25, 335–349. <https://doi.org/10.1016/j.ccr.2014.01.030>
- Lavoie, H., Therrien, M., 2015. Regulation of RAF protein kinases in ERK signalling. *Nat. Rev. Mol. Cell Biol.* 16, 281–298. <https://doi.org/10.1038/nrm3979>
- Lazarevic, I., 2014. Clinical implications of hepatitis B virus mutations: Recent advances. *World J. Gastroenterol.* 20, 7653. <https://doi.org/10.3748/wjg.v20.i24.7653>
- Lee, M.W., Raman, S.S., Asvadi, N.H., Siripongsakun, S., Hicks, R.M., Chen, J., Worakitsitatorn, A., McWilliams, J., Tong, M.J., Finn, R.S., Agopian, V.G., Busuttil, R.W., Lu, D.S.K., 2017. Radiofrequency ablation of hepatocellular carcinoma as bridge therapy to liver transplantation: A 10-year intention-to-treat analysis. *Hepatology* 65, 1979–1990. <https://doi.org/10.1002/hep.29098>
- Lemoine, M., Thursz, M.R., 2017. Battlefield against hepatitis B infection and HCC in Africa. *J. Hepatol.* 66, 645–654. <https://doi.org/10.1016/j.jhep.2016.10.013>
- Letouzé, E., Shinde, J., Renault, V., Couchy, G., Blanc, J.-F., Tubacher, E., Bayard, Q., Bacq, D., Meyer, V., Semhoun, J., Bioulac-Sage, P., Prévôt, S., Azoulay, D., Paradis, V., Imbeaud, S., Deleuze, J.-F., Zucman-Rossi, J., 2017. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. *Nat. Commun.* 8, 1315. <https://doi.org/10.1038/s41467-017-01358-x>
- Levrero, M., Zucman-Rossi, J., 2016. Mechanisms of HBV-induced hepatocellular carcinoma. *J. Hepatol.* 64, S84–S101. <https://doi.org/10.1016/j.jhep.2016.02.021>
- Li, C., Ho, M., Lin, Y., Tzeng, S., Chen, Y., Pai, H., Wang, Y., Chen, C., Lee, Y., Chen, D., Yeh, S., Chen, P., 2020. Cell-free virus-host chimera DNA from Hepatitis B virus integration sites as a circulating biomarker of hepatocellular cancer. *Hepatology* 61, 31230. <https://doi.org/10.1002/hep.31230>
- Li, Chiao-Ling, Li, Chen-Yu, Lin, Y., Ho, M., Chen, D., Chen, P., Yeh, S., 2019. Androgen Receptor Enhances Hepatic Telomerase Reverse Transcriptase Gene Transcription After Hepatitis B Virus Integration or Point Mutation in Promoter Region. *Hepatology* 69, 498–512. <https://doi.org/10.1002/hep.30201>
- Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, H., Durbin, R., 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26, 589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>

BIBLIOGRAPHIE

- Li, X., Zhang, J., Yang, Z., Kang, J., Jiang, S., Zhang, T., Chen, T., Li, M., Lv, Q., Chen, X., McCrae, M.A., Zhuang, H., Lu, F., 2014. The function of targeted host genes determines the oncogenicity of HBV integration in hepatocellular carcinoma. *J. Hepatol.* 60, 975–984. <https://doi.org/10.1016/j.jhep.2013.12.014>
- Lin, S., Nascimento, E.M., Gajera, C.R., Chen, L., Neuhöfer, P., Garbuzov, A., Wang, S., Artandi, S.E., 2018. Distributed hepatocytes expressing telomerase repopulate the liver in homeostasis and injury. *Nature* 556, 244–248. <https://doi.org/10.1038/s41586-018-0004-7>
- Liu, X., Dakic, A., Zhang, Y., Dai, Y., Chen, R., Schlegel, R., 2009. HPV E6 protein interacts physically and functionally with the cellular telomerase complex. *Proc. Natl. Acad. Sci.* 106, 18780–18785. <https://doi.org/10.1073/pnas.0906357106>
- Llovet, J.M., Hilgard, P., de Oliveira, A.C., Forner, A., Zeuzem, S., Galle, P.R., Häussinger, D., Moscovici, M., 2008. Sorafenib in Advanced Hepatocellular Carcinoma. *N Engl J Med* 13.
- Llovet, J.M., Zucman-Rossi, J., Pikarsky, E., Sangro, B., Schwartz, M., Sherman, M., Gores, G., 2016. Hepatocellular carcinoma. *Nat. Rev. Dis. Primer* 2, 16018. <https://doi.org/10.1038/nrdp.2016.18>
- Lucifora, J., Arzberger, S., Durantel, D., Belloni, L., Strubin, M., Levrero, M., Zoulim, F., Hantz, O., Protzer, U., 2011. Hepatitis B virus X protein is essential to initiate and maintain virus replication after infection. *J. Hepatol.* 55, 996–1003. <https://doi.org/10.1016/j.jhep.2011.02.015>
- Ludgate, L., Liu, K., Luckenbaugh, L., Streck, N., Eng, S., Voitenleitner, C., Delaney, W.E., Hu, J., 2016. Cell-Free Hepatitis B Virus Capsid Assembly Dependent on the Core Protein C-Terminal Domain and Regulated by Phosphorylation. *J. Virol.* 90, 5830–5844. <https://doi.org/10.1128/JVI.00394-16>
- Ma, N.-F., Lau, S.H., Hu, L., Xie, D., Wu, J., Yang, J., Wang, Y., Wu, M.-C., Fung, J., Bai, X., Tzang, C.-H., Fu, L., Yang, M., Su, Y.A., Guan, X.-Y., 2008. COOH-Terminal Truncated HBV X Protein Plays Key Role in Hepatocarcinogenesis. *Clin. Cancer Res.* 14, 5061–5068. <https://doi.org/10.1158/1078-0432.CCR-07-5082>
- Ma, Z., Zhang, E., Gao, S., Xiong, Y., Lu, M., 2019. Toward a Functional Cure for Hepatitis B: The Rationale and Challenges for Therapeutic Targeting of the B Cell Immune Response. *Front. Immunol.* 10, 2308. <https://doi.org/10.3389/fimmu.2019.02308>
- Maciejowski, J., de Lange, T., 2017. Telomeres in cancer: tumour suppression and genome instability. *Nat. Rev. Mol. Cell Biol.* 18, 175–186. <https://doi.org/10.1038/nrm.2016.171>
- Margini, C., Dufour, J.F., 2016. The story of HCC in NAFLD: from epidemiology, across pathogenesis, to prevention and treatment. *Liver Int.* 36, 317–324. <https://doi.org/10.1111/liv.13031>
- Martinez, M.G., Testoni, B., Zoulim, F., 2019. Biological basis for functional cure of chronic hepatitis B. *J. Viral Hepat.* 26, 786–794. <https://doi.org/10.1111/jvh.13090>
- Martinez, M.G., Villeret, F., Testoni, B., Zoulim, F., 2020. Can we cure hepatitis B virus with novel direct-acting antivirals? *Liver Int.* 40, 27–34. <https://doi.org/10.1111/liv.14364>
- Mason, W.S., Gill, U.S., Litwin, S., Zhou, Y., Peri, S., Pop, O., Hong, M.L.W., Naik, S., Quaglia, A., Bertoletti, A., Kennedy, P.T.F., 2016. HBV DNA Integration and Clonal Hepatocyte Expansion in Chronic Hepatitis B Patients Considered Immune Tolerant. *Gastroenterology* 151, 986–998.e4. <https://doi.org/10.1053/j.gastro.2016.07.012>
- Mason, W.S., Jilbert, A.R., Summers, J., 2005. Clonal expansion of hepatocytes during chronic woodchuck hepatitis virus infection. *Proc. Natl. Acad. Sci.* 102, 1139–1144. <https://doi.org/10.1073/pnas.0409332102>

- Mason, W.S., Liu, C., Aldrich, C.E., Litwin, S., Yeh, M.M., 2010. Clonal Expansion of Normal-Appearing Human Hepatocytes during Chronic Hepatitis B Virus Infection. *J. Virol.* 84, 8308–8315. <https://doi.org/10.1128/JVI.00833-10>
- Mavilia, M.G., Wu, G.Y., 2018. HBV-HCV Coinfection: Viral Interactions, Management, and Viral Reactivation. *J. Clin. Transl. Hepatol.* 6, 1–10. <https://doi.org/10.14218/JCTH.2018.00016>
- McMahon, B.J., 2009. The natural history of chronic hepatitis B virus infection. *Hepatology* 49, S45–S55. <https://doi.org/10.1002/hep.22898>
- McNaughton, A.L., Revill, P.A., Littlejohn, M., Matthews, P.C., Ansari, M.A., 2020. Analysis of genomic-I-length HBV sequences to determine genotype and subgenotype reference sequences. *J. Gen. Virol.* 13.
- Meng, X., Franklin, D.A., Dong, J., Zhang, Y., 2014. MDM2–p53 Pathway in Hepatocellular Carcinoma. *Cancer Res.* 74, 7161–7167. <https://doi.org/10.1158/0008-5472.CAN-14-1446>
- Mentha, N., Clément, S., Negro, F., Alfaiate, D., 2019. A review on hepatitis D: From virology to new therapies. *J. Adv. Res.* 17, 3–15. <https://doi.org/10.1016/j.jare.2019.03.009>
- Mesri, E.A., Feitelson, M.A., Munger, K., 2014. Human Viral Oncogenesis: A Cancer Hallmarks Analysis. *Cell Host Microbe* 15, 266–282. <https://doi.org/10.1016/j.chom.2014.02.011>
- Moody, C.A., Laimins, L.A., 2010. Human papillomavirus oncoproteins: pathways to transformation. *Nat. Rev. Cancer* 10, 550–560. <https://doi.org/10.1038/nrc2886>
- Moolla, N., Kew, M., Arbuthnot, P., 2002. Regulatory elements of hepatitis B virus transcription. *J. Viral Hepat.* 9, 323–331. <https://doi.org/10.1046/j.1365-2893.2002.00381.x>
- Muir, A.J., 2015. Understanding the Complexities of Cirrhosis. *Clin. Ther.* 37, 1822–1836. <https://doi.org/10.1016/j.clinthera.2015.05.507>
- Nahon, P., Zucman-Rossi, J., 2012. Single nucleotide polymorphisms and risk of hepatocellular carcinoma in cirrhosis. *J. Hepatol.* 57, 663–674. <https://doi.org/10.1016/j.jhep.2012.02.035>
- Nassal, M., 2015. HBV cccDNA: viral persistence reservoir and key obstacle for a cure of chronic hepatitis B. *Gut* 64, 1972–1984. <https://doi.org/10.1136/gutjnl-2015-309809>
- Nassal, M., 2008. Hepatitis B viruses: Reverse transcription a different way. *Virus Res.* 134, 235–249. <https://doi.org/10.1016/j.virusres.2007.12.024>
- Nault, J., Martin, Y., Caruso, S., Hirsch, T.Z., Bayard, Q., Calderaro, J., Charpy, C., Copie-Bergman, C., Zioli, M., Bioulac-Sage, P., Couchy, G., Blanc, J., Nahon, P., Amaddeo, G., Ganne-Carrie, N., Morcrette, G., Chiche, L., Duvoux, C., Faivre, S., Laurent, A., Imbeaud, S., Rebouissou, S., Llovet, J.M., Seror, O., Letouzé, E., Zucman-Rossi, J., 2020. Clinical Impact of Genomic Diversity From Early to Advanced Hepatocellular Carcinoma. *Hepatology* 71, 164–182. <https://doi.org/10.1002/hep.30811>
- Nault, J.C., Calderaro, J., Di Tommaso, L., Balabaud, C., Zafrani, E.S., Bioulac-Sage, P., Roncalli, M., Zucman-Rossi, J., 2014. Telomerase reverse transcriptase promoter mutation is an early somatic genetic alteration in the transformation of premalignant nodules in hepatocellular carcinoma on cirrhosis. *Hepatology* 60, 1983–1992. <https://doi.org/10.1002/hep.27372>
- Nault, J.-C., Datta, S., Imbeaud, S., Franconi, A., Mallet, M., Couchy, G., Letouzé, E., Pilati, C., Verret, B., Blanc, J.-F., Balabaud, C., Calderaro, J., Laurent, A., Letexier, M., Bioulac-Sage, P., Calvo, F., Zucman-Rossi, J., 2015. Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat. Genet.* 47, 1187–1193. <https://doi.org/10.1038/ng.3389>
- Nault, J.-C., Letouzé, E., 2019. Mutational Processes in Hepatocellular Carcinoma: The Story of Aristolochic Acid. *Semin. Liver Dis.* 39, 334–340. <https://doi.org/10.1055/s-0039-1685516>

BIBLIOGRAPHIE

- Nault, J.C., Mallet, M., Pilati, C., Calderaro, J., Bioulac-Sage, P., Laurent, C., Laurent, A., Cherqui, D., Balabaud, C., Zucman-Rossi, J., 2013. High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. *Nat. Commun.* 4, 2218. <https://doi.org/10.1038/ncomms3218>
- Nault, J.-C., Ningarhari, M., Rebouissou, S., Zucman-Rossi, J., 2019. The role of telomeres and telomerase in cirrhosis and liver cancer. *Nat. Rev. Gastroenterol. Hepatol.* 16, 544–558. <https://doi.org/10.1038/s41575-019-0165-3>
- Ng, K.-Y., Chai, S., Tong, M., Guan, X.-Y., Lin, C.-H., Ching, Y.-P., Xie, D., Cheng, A.S.-L., Ma, S., 2016. C-terminal truncated hepatitis B virus X protein promotes hepatocellular carcinogenesis through induction of cancer and stem cell-like properties. *Oncotarget* 7. <https://doi.org/10.18632/oncotarget.8209>
- Nik-Zainal, S., Davies, H., Staaf, J., Ramakrishna, M., Glodzik, D., Zou, X., Martincorena, I., Alexandrov, L.B., Martin, S., Wedge, D.C., Van Loo, P., Ju, Y.S., Smid, M., Brinkman, A.B., Morganella, S., Aure, M.R., Lingjærde, O.C., Langerød, A., Ringnér, M., Ahn, S.-M., Boyault, S., Brock, J.E., Broeks, A., Butler, A., Desmedt, C., Dirix, L., Dronov, S., Fatima, A., Foekens, J.A., Gerstung, M., Hooijer, G.K.J., Jang, S.J., Jones, D.R., Kim, H.-Y., King, T.A., Krishnamurthy, S., Lee, H.J., Lee, J.-Y., Li, Y., McLaren, S., Menzies, A., Mustonen, V., O'Meara, S., Pauporté, I., Pivot, X., Purdie, C.A., Raine, K., Ramakrishnan, K., Rodríguez-González, F.G., Romieu, G., Sieuwerts, A.M., Simpson, P.T., Shepherd, R., Stebbings, L., Stefansson, O.A., Teague, J., Tommasi, S., Treilleux, I., Van den Eynden, G.G., Vermeulen, P., Vincent-Salomon, A., Yates, L., Caldas, C., Veer, L. van't, Tutt, A., Knappskog, S., Tan, B.K.T., Jonkers, J., Borg, Å., Ueno, N.T., Sotiriou, C., Viari, A., Futreal, P.A., Campbell, P.J., Span, P.N., Van Laere, S., Lakhani, S.R., Eyfjord, J.E., Thompson, A.M., Birney, E., Stunnenberg, H.G., van de Vijver, M.J., Martens, J.W.M., Børresen-Dale, A.-L., Richardson, A.L., Kong, G., Thomas, G., Stratton, M.R., 2016. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534, 47–54. <https://doi.org/10.1038/nature17676>
- Nik-Zainal, S., Kucab, J.E., Morganella, S., Glodzik, D., Alexandrov, L.B., Arlt, V.M., Wenginger, A., Hollstein, M., Stratton, M.R., Phillips, D.H., 2015. The genome as a record of environmental exposure. *Mutagenesis* gev073. <https://doi.org/10.1093/mutage/gev073>
- Ningarhari, M., Caruso, S., Hirsch, T.Z., Bayard, Q., Franconi, A., Védie, A.-L., Noblet, B., Blanc, J.-F., Amaddeo, G., Ganne, N., Ziol, M., Paradis, V., Guettier, C., Calderaro, J., Morcrette, G., Kim, Y., MacLeod, A.R., Nault, J.C., Rebouissou, S., Zucman-Rossi, J., Non publié. Telomere length is a major determinant of hepatocellular carcinoma diversity and telomerase addiction an actionable therapeutic target.
- Niu, Z.-S., Niu, X.-J., Wang, W.-H., 2016. Genetic alterations in hepatocellular carcinoma: An update. *World J. Gastroenterol.* 22, 9069. <https://doi.org/10.3748/wjg.v22.i41.9069>
- Papatheodoridis, G., Dalekos, G., Sypsa, V., Yurdaydin, C., Buti, M., Goulis, J., Calleja, J.L., Chi, H., Manolakopoulos, S., Mangia, G., Gatselis, N., Keskin, O., Savvidou, S., de la Revilla, J., Hansen, B.E., Vlachogiannakos, I., Galanis, K., Idilman, R., Colombo, M., Esteban, R., Janssen, H.L.A., Lampertico, P., 2016. PAGE-B predicts the risk of developing hepatocellular carcinoma in Caucasians with chronic hepatitis B on 5-year antiviral therapy. *J. Hepatol.* 64, 800–806. <https://doi.org/10.1016/j.jhep.2015.11.035>
- Papatheodoridis, G.V., Chan, H.L.-Y., Hansen, B.E., Janssen, H.L.A., Lampertico, P., 2015. Risk of hepatocellular carcinoma in chronic hepatitis B: Assessment and modification with current antiviral therapy. *J. Hepatol.* 62, 956–967. <https://doi.org/10.1016/j.jhep.2015.01.002>
- Paradis, V., Zalinski, S., Chelbi, E., Guedj, N., Degos, F., Vilgrain, V., Bedossa, P., Belghiti, J., 2009. Hepatocellular carcinomas in patients with metabolic syndrome often develop without significant liver fibrosis: A pathological analysis. *Hepatology* 49, 851–859. <https://doi.org/10.1002/hep.22734>

- Park, I.Y., Sohn, B.H., Yu, E., Suh, D.J., Chung, Y., Lee, J., Surzycki, S.J., Lee, Y.I., 2007. Aberrant Epigenetic Modifications in Hepatocarcinogenesis Induced by Hepatitis B Virus X Protein. *Gastroenterology* 132, 1476–1494. <https://doi.org/10.1053/j.gastro.2007.01.034>
- Patient, R., Hourieux, C., Roingeard, P., 2009. Morphogenesis of hepatitis B virus and its subviral envelope particles. *Cell. Microbiol.* 11, 1561–1570. <https://doi.org/10.1111/j.1462-5822.2009.01363.x>
- Pellicoro, A., Ramachandran, P., Iredale, J.P., Fallowfield, J.A., 2014. Liver fibrosis and repair: immune regulation of wound healing in a solid organ. *Nat. Rev. Immunol.* 14, 181–194. <https://doi.org/10.1038/nri3623>
- Perugorria, M.J., Olaizola, P., Labiano, I., Esparza-Baquer, A., Marzioni, M., Marin, J.J.G., Bujanda, L., Banales, J.M., 2019. Wnt- β -catenin signalling in liver development, health and disease. *Nat. Rev. Gastroenterol. Hepatol.* 16, 121–136. <https://doi.org/10.1038/s41575-018-0075-9>
- Petruzzello, A., 2018. Epidemiology of Hepatitis B Virus (HBV) and Hepatitis C Virus (HCV) Related Hepatocellular Carcinoma. *Open Virol. J.* 12, 26–32. <https://doi.org/10.2174/1874357901812010026>
- Phung, B.-C., 2014. Hepatitis B and human immunodeficiency virus co-infection. *World J. Gastroenterol.* 20, 17360. <https://doi.org/10.3748/wjg.v20.i46.17360>
- Pollack, J.R., Ganem, D., 1994. Site-specific RNA binding by a hepatitis B virus reverse transcriptase initiates two distinct reactions: RNA packaging and DNA synthesis. *J. Virol.* 68, 5579–5587. <https://doi.org/10.1128/JVI.68.9.5579-5587.1994>
- Pollicino, T., Saitta, C., 2014. Occult hepatitis B virus and hepatocellular carcinoma 20, 12.
- Puigvehí, M., Moctezuma-Velázquez, C., Villanueva, A., Llovet, J.M., 2019. The oncogenic role of hepatitis delta virus in hepatocellular carcinoma. *JHEP Rep.* 1, 120–130. <https://doi.org/10.1016/j.jhepr.2019.05.001>
- Quasdorff, M., Protzer, U., 2010. Control of hepatitis B virus at the level of transcription: Control of hepatitis B virus. *J. Viral Hepat.* 17, 527–536. <https://doi.org/10.1111/j.1365-2893.2010.01315.x>
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Raimondo, G., Allain, J.-P., Brunetto, M.R., Buendia, M.-A., Chen, D.-S., Colombo, M., Craxì, A., Donato, F., Ferrari, C., Gaeta, G.B., Gerlich, W.H., Levrero, M., Locarnini, S., Michalak, T., Mondelli, M.U., Pawlotsky, J.-M., Pollicino, T., Prati, D., Puoti, M., Samuel, D., Shouval, D., Smedile, A., Squadrito, G., Trépo, C., Villa, E., Will, H., Zanetti, A.R., Zoulim, F., 2008. Statements from the Taormina expert meeting on occult hepatitis B virus infection. *J. Hepatol.* 49, 652–657. <https://doi.org/10.1016/j.jhep.2008.07.014>
- Raimondo, G., Locarnini, S., Pollicino, T., Levrero, M., Zoulim, F., Lok, A.S., Allain, J.-P., Berg, T., Bertolotti, A., Brunetto, M.R., Bruno, R., Chen, D.-S., Coppola, N., Cornberg, M., Craxì, A., Dandri, M., Di Marco, V., Ferrari, C., Gaeta, G.B., Glebe, D., Guidotti, L.G., Kramvis, A., Lampertico, P., Li, C., Liang, J., Marzano, A., Michalak, T.I., Pawlotsky, J.-M., Prati, D., Puoti, M., Samuel, D., Soriano, V., Squadrito, G., Sureau, C., Trepo, C., Yuen, M.-F., 2019. Update of the statements on biology and clinical impact of occult hepatitis B virus infection. *J. Hepatol.* 71, 397–408. <https://doi.org/10.1016/j.jhep.2019.03.034>
- Read, S.A., Douglas, M.W., 2014. Virus induced inflammation and cancer development. *Cancer Lett.* 345, 174–181. <https://doi.org/10.1016/j.canlet.2013.07.030>
- Rebouissou, S., Franconi, A., Calderaro, J., Letouzé, E., Imbeaud, S., Pilati, C., Nault, J.-C., Couchy, G., Laurent, A., Balabaud, C., Bioulac-Sage, P., Zucman-Rossi, J., 2016. Genotype-

- phenotype correlation of *CTNNB1* mutations reveals different β -catenin activity associated with liver tumor progression: *Hepatology*, Vol. XX, No. X, 2016 Rebouissou, Franconi, et al. *Hepatology* 64, 2047–2061. <https://doi.org/10.1002/hep.28638>
- Rebouissou, S., Nault, J.-C., 2020. Advances in molecular classification and precision oncology in hepatocellular carcinoma. *J. Hepatol.* 72, 215–229. <https://doi.org/10.1016/j.jhep.2019.08.017>
- Revill, P.A., Chisari, F.V., Block, J.M., Dandri, M., Gehring, A.J., Guo, H., Hu, J., Kramvis, A., Lampertico, P., Janssen, H.L.A., Levrero, M., Li, W., Liang, T.J., Lim, S.-G., Lu, F., Penicaud, M.C., Tavis, J.E., Thimme, R., Zoulim, F., Arbuthnot, P., Boonstra, A., Chang, K.-M., Chen, P.-J., Glebe, D., Guidotti, L.G., Fellay, J., Ferrari, C., Jansen, L., Lau, D.T.Y., Lok, A.S., Maini, M.K., Mason, W., Matthews, G., Paraskevis, D., Petersen, J., Rehmann, B., Shin, E.-C., Thompson, A., van Bömmel, F., Wang, F.-S., Watashi, K., Yang, H.-C., Yuan, Z., Yuen, M.-F., Block, T., Miller, V., Protzer, U., Bréchet, C., Locarnini, S., Peters, M.G., Schinazi, R.F., 2019. A global scientific strategy to cure hepatitis B. *Lancet Gastroenterol. Hepatol.* 4, 545–558. [https://doi.org/10.1016/S2468-1253\(19\)30119-0](https://doi.org/10.1016/S2468-1253(19)30119-0)
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47–e47. <https://doi.org/10.1093/nar/gkv007>
- Rivière, L., Quioc-Salomon, B., Fallot, G., Halgand, B., Féray, C., Buendia, M.-A., Neuveut, C., 2019. Hepatitis B virus replicating in hepatocellular carcinoma encodes HBx variants with preserved ability to antagonize restriction by Smc5/6. *Antiviral Res.* 172, 104618. <https://doi.org/10.1016/j.antiviral.2019.104618>
- Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., Wu, Y.-C., Pfenning, A.R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R.A., Shores, N., Epstein, C.B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R.D., Lister, R., Hong, C., Gascard, P., Mungall, A.J., Moore, R., Chuah, E., Tam, A., Canfield, T.K., Hansen, R.S., Kaul, R., Sabo, P.J., Bansal, M.S., Carles, A., Dixon, J.R., Farh, K.-H., Feizi, S., Karlic, R., Kim, A.-R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T.R., Neph, S.J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R.C., Siebenthal, K.T., Sinnott-Armstrong, N.A., Stevens, M., Thurman, R.E., Wu, J., Zhang, B., Zhou, X., Beaudet, A.E., Boyer, L.A., De Jager, P.L., Farnham, P.J., Fisher, S.J., Haussler, D., Jones, S.J.M., Li, W., Marra, M.A., McManus, M.T., Sunyaev, S., Thomson, J.A., Tlsty, T.D., Tsai, L.-H., Wang, W., Waterland, R.A., Zhang, M.Q., Chadwick, L.H., Bernstein, B.E., Costello, J.F., Ecker, J.R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J.A., Wang, T., Kellis, M., 2015. Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330. <https://doi.org/10.1038/nature14248>
- Roberts, S.A., Lawrence, M.S., Klimczak, L.J., Grimm, S.A., Fargo, D., Stojanov, P., Kiezun, A., Kryukov, G.V., Carter, S.L., Saksena, G., Harris, S., Shah, R.R., Resnick, M.A., Getz, G., Gordenin, D.A., 2013. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat. Genet.* 45, 970–976. <https://doi.org/10.1038/ng.2702>
- Sagnelli, E., Macera, M., Russo, A., Coppola, N., Sagnelli, C., 2020. Epidemiological and etiological variations in hepatocellular carcinoma. *Infection* 48, 7–17. <https://doi.org/10.1007/s15010-019-01345-y>
- Saigo, K., Yoshida, K., Ikeda, R., Sakamoto, Y., Murakami, Y., Urashima, T., Asano, T., Kenmochi, T., Inoue, I., 2008. Integration of hepatitis B virus DNA into the myeloid/lymphoid or mixed-lineage leukemia (MLL4) gene and rearrangements of MLL4 in human hepatocellular carcinoma. *Hum. Mutat.* 29, 703–708. <https://doi.org/10.1002/humu.20701>
- Saitta, C., Tripodi, G., Barbera, A., Bertuccio, A., Smedile, A., Ciancio, A., Raffa, G., Sangiovanni, A., Navarra, G., Raimondo, G., Pollicino, T., 2015. Hepatitis B virus (HBV) DNA integration in

- patients with occult HBV infection and hepatocellular carcinoma. *Liver Int.* 35, 2311–2317. <https://doi.org/10.1111/liv.12807>
- Sarmati, L., Malagnino, V., 2019. HBV Infection in HIV-Driven Immune Suppression. *Viruses* 11, 1077. <https://doi.org/10.3390/v11111077>
- Schulze, K., Imbeaud, S., Letouzé, E., Alexandrov, L.B., Calderaro, J., Rebouissou, S., Couchy, G., Meiller, C., Shinde, J., Soysouvanh, F., Calatayud, A.-L., Pinyol, R., Pelletier, L., Balabaud, C., Laurent, A., Blanc, J.-F., Mazzaferro, V., Calvo, F., Villanueva, A., Nault, J.-C., Bioulac-Sage, P., Stratton, M.R., Llovet, J.M., Zucman-Rossi, J., 2015. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat. Genet.* 47, 505–511. <https://doi.org/10.1038/ng.3252>
- Schulze, K., Nault, J.-C., Villanueva, A., 2016. Genetic profiling of hepatocellular carcinoma using next-generation sequencing. *J. Hepatol.* 65, 1031–1042. <https://doi.org/10.1016/j.jhep.2016.05.035>
- Schuppan, D., Afdhal, N.H., 2008. Liver cirrhosis. *Lancet.*
- Seeger, C., Mason, W.S., 2015. Molecular biology of hepatitis B virus infection. *Virology* 479–480, 672–686. <https://doi.org/10.1016/j.virol.2015.02.031>
- Sherr, C.J., McCormick, F., 2002. The RB and p53 pathways in cancer. *Cancer Cell* 2, 103–112. [https://doi.org/10.1016/S1535-6108\(02\)00102-2](https://doi.org/10.1016/S1535-6108(02)00102-2)
- Shiraishi, Y., Fujimoto, A., Furuta, M., Tanaka, H., Chiba, K., Boroevich, K.A., Abe, T., Kawakami, Y., Ueno, M., Gotoh, K., Ariizumi, S., Shibuya, T., Nakano, K., Sasaki, A., Maejima, K., Kitada, R., Hayami, S., Shigekawa, Y., Marubashi, S., Yamada, T., Kubo, M., Ishikawa, O., Aikata, H., Arihiro, K., Ohdan, H., Yamamoto, M., Yamaue, H., Chayama, K., Tsunoda, T., Miyano, S., Nakagawa, H., 2014. Integrated Analysis of Whole Genome and Transcriptome Sequencing Reveals Diverse Transcriptomic Aberrations Driven by Somatic Genomic Changes in Liver Cancers. *PLoS ONE* 9, e114263. <https://doi.org/10.1371/journal.pone.0114263>
- Shirvani-Dastgerdi, E., Schwartz, R.E., Ploss, A., 2016. Hepatocarcinogenesis associated with hepatitis B, delta and C viruses. *Curr. Opin. Virol.* 20, 1–10. <https://doi.org/10.1016/j.coviro.2016.07.009>
- Shouval, D., Roggendorf, H., Roggendorf, M., 2015. Enhanced immune response to hepatitis B vaccination through immunization with a Pre-S1/Pre-S2/S Vaccine. *Med. Microbiol. Immunol. (Berl.)* 204, 57–68. <https://doi.org/10.1007/s00430-014-0374-x>
- Singal, A.G., Lampertico, P., Nahon, P., 2020. Epidemiology and surveillance for hepatocellular carcinoma: New trends. *J. Hepatol.* 72, 250–261. <https://doi.org/10.1016/j.jhep.2019.08.025>
- Slagle, B.L., Andrisani, O.M., Bouchard, M.J., Lee, C.G.L., Ou, J.-H.J., Siddiqui, A., 2015. Technical standards for hepatitis B virus X protein (HBx) research. *Hepatology* 61, 1416–1424. <https://doi.org/10.1002/hep.27360>
- Slagle, B.L., Bouchard, M.J., 2018. Role of HBx in hepatitis B virus persistence and its therapeutic implications. *Curr. Opin. Virol.* 30, 32–38. <https://doi.org/10.1016/j.coviro.2018.01.007>
- Soussan, P., Tuveri, R., Nalpas, B., Garreau, F., Zavala, F., Masson, A., Pol, S., Brechot, C., Kremsdorf, D., 2003. 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.* 38, 343–348. [https://doi.org/10.1016/S0168-8278\(02\)00422-1](https://doi.org/10.1016/S0168-8278(02)00422-1)
- Sporn, M.B., Liby, K.T., 2012. NRF2 and cancer: the good, the bad and the importance of context. *Nat. Rev. Cancer* 12, 564–571. <https://doi.org/10.1038/nrc3278>

BIBLIOGRAPHIE

- Stanaway, J.D., Flaxman, A.D., Naghavi, M., Fitzmaurice, C., Vos, T., Abubakar, I., Abu-Raddad, L.J., Assadi, R., Bhala, N., Cowie, B., Forouzanfour, M.H., Groeger, J., Hanafiah, K.M., Jacobsen, K.H., James, S.L., MacLachlan, J., Malekzadeh, R., Martin, N.K., Mokdad, A.A., Mokdad, A.H., Murray, C.J.L., Plass, D., Rana, S., Rein, D.B., Richardus, J.H., Sanabria, J., Saylan, M., Shahraz, S., So, S., Vlassov, V.V., Weiderpass, E., Wiersma, S.T., Younis, M., Yu, C., El Sayed Zaki, M., Cooke, G.S., 2016. The global burden of viral hepatitis from 1990 to 2013: findings from the Global Burden of Disease Study 2013. *The Lancet* 388, 1081–1088. [https://doi.org/10.1016/S0140-6736\(16\)30579-7](https://doi.org/10.1016/S0140-6736(16)30579-7)
- Strick-Marchand, H., Dusséaux, M., Darche, S., Huntington, N.D., Legrand, N., Masse-Ranson, G., Corcuff, E., Ahodantin, J., Weijer, K., Spits, H., Kremsdorf, D., Di Santo, J.P., 2015. A Novel Mouse Model for Stable Engraftment of a Human Immune System and Human Hepatocytes. *PLOS ONE* 10, e0119820. <https://doi.org/10.1371/journal.pone.0119820>
- Su, I.-J., Wang, L.H.-C., Hsieh, W.-C., Wu, H.-C., Teng, C.-F., Tsai, H.-W., Huang, W., 2014. The emerging role of hepatitis B virus Pre-S2 deletion mutant proteins in HBV tumorigenesis. *J. Biomed. Sci.* 21, 98. <https://doi.org/10.1186/s12929-014-0098-7>
- Su, Q., Schröder, C.H., Hofmann, W.J., Otto, G., Pichlmayr, R., Bannasch, P., 1998. Expression of hepatitis B virus X protein in HBV-infected human livers and hepatocellular carcinomas: Expression of Hepatitis B Virus X Protein in HBV-Infected Human Livers and Hepatocellular Carcinomas. *Hepatology* 27, 1109–1120. <https://doi.org/10.1002/hep.510270428>
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550. <https://doi.org/10.1073/pnas.0506580102>
- Summers, J., Jilbert, A.R., Yang, W., Aldrich, C.E., Saputelli, J., Litwin, S., Toll, E., Mason, W.S., 2003. Hepatocyte turnover during resolution of a transient hepadnaviral infection. *Proc. Natl. Acad. Sci.* 100, 11652–11659. <https://doi.org/10.1073/pnas.1635109100>
- Sung, W.-K., Zheng, H., Li, S., Chen, R., Liu, X., Li, Y., Lee, N.P., Lee, W.H., Ariyaratne, P.N., Tennakoon, C., Mulawadi, F.H., Wong, K.F., Liu, A.M., Poon, R.T., Fan, S.T., Chan, K.L., Gong, Z., Hu, Y., Lin, Z., Wang, G., Zhang, Q., Barber, T.D., Chou, W.-C., Aggarwal, A., Hao, K., Zhou, W., Zhang, C., Hardwick, J., Buser, C., Xu, J., Kan, Z., Dai, H., Mao, M., Reinhard, C., Wang, J., Luk, J.M., 2012. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat. Genet.* 44, 765–769. <https://doi.org/10.1038/ng.2295>
- Sze, K.M., Ho, D.W., Chiu, Y., Tsui, Y., Chan, L., Lee, J.M., Chok, K.S., Chan, A.C., Tang, C., Tang, V.W., Lo, I.L., Yau, D.T., Cheung, T., Ng, I.O., 2020. HBV- *TERT* Promoter Integration Harnesses Host ELF4 Resulting in *TERT* Gene Transcription in Hepatocellular Carcinoma. *Hepatology* hep.31231. <https://doi.org/10.1002/hep.31231>
- Sze, K.M.F., Chu, G.K.Y., Lee, J.M.F., Ng, I.O.L., 2013. C-terminal truncated hepatitis B virus x protein is associated with metastasis and enhances invasiveness by c-jun/matrix metalloproteinase protein 10 activation in hepatocellular carcinoma. *Hepatology* 57, 131–139. <https://doi.org/10.1002/hep.25979>
- Tahmasebi-Birgani, M., Ansari, H., Carloni, V., 2019. Defective mitosis-linked DNA damage response and chromosomal instability in liver cancer. *Biochim. Biophys. Acta BBA - Rev. Cancer* 1872, 60–65. <https://doi.org/10.1016/j.bbcan.2019.05.008>
- Tan, A.T., Yang, N., Lee Krishnamoorthy, T., Oei, V., Chua, A., Zhao, X., Tan, H.S., Chia, A., Le Bert, N., Low, D., Tan, H.K., Kumar, R., Irani, F.G., Ho, Z.Z., Zhang, Q., Guccione, E., Wai, L.-E., Koh, S., Hwang, W., Chow, W.C., Bertolotti, A., 2019. Use of Expression Profiles of HBV-DNA Integrated Into Genomes of Hepatocellular Carcinoma Cells to Select T Cells for Immunotherapy. *Gastroenterology* 156, 1862-1876.e9. <https://doi.org/10.1053/j.gastro.2019.01.251>

- The Cancer Genome Atlas Research Network, 2017. Integrated genomic and molecular characterization of cervical cancer. *Nature* 543, 378–384. <https://doi.org/10.1038/nature21386>
- The ENCODE Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. <https://doi.org/10.1038/nature11247>
- Thio, C.L., 2009. Hepatitis B and human immunodeficiency virus coinfection. *Hepatology* 49, S138–S145. <https://doi.org/10.1002/hep.22883>
- Thomas, D.L., 2019. Global Elimination of Chronic Hepatitis. *N. Engl. J. Med.* 380, 2041–2050. <https://doi.org/10.1056/NEJMra1810477>
- Tian, Y., Ou, J.J., 2015. Genetic and epigenetic alterations in hepatitis B virus-associated hepatocellular carcinoma. *Viol. Sin.* 30, 85–91. <https://doi.org/10.1007/s12250-015-3582-7>
- Toh, S.T., Jin, Y., Liu, L., Wang, J., Babrzadeh, F., Gharizadeh, B., Ronaghi, M., Toh, H.C., Chow, P.K.-H., Chung, A.Y.-F., Ooi, L.L.-P.-J., Lee, C.G.-L., 2013. Deep sequencing of the hepatitis B virus in hepatocellular carcinoma patients reveals enriched integration events, structural alterations and sequence variations. *Carcinogenesis* 34, 787–798. <https://doi.org/10.1093/carcin/bgs406>
- Tong, S., Revill, P., 2016. Overview of hepatitis B viral replication and genetic variability. *J. Hepatol.* 64, S4–S16. <https://doi.org/10.1016/j.jhep.2016.01.027>
- Torrecilla, S., Sia, D., Harrington, A.N., Zhang, Z., Cabellos, L., Cornella, H., Moeini, A., Camprecios, G., Leow, W.-Q., Fiel, M.I., Hao, K., Bassaganyas, L., Mahajan, M., Thung, S.N., Villanueva, A., Florman, S., Schwartz, M.E., Llovet, J.M., 2017. Trunk mutational events present minimal intra- and inter-tumoral heterogeneity in hepatocellular carcinoma. *J. Hepatol.* 67, 1222–1231. <https://doi.org/10.1016/j.jhep.2017.08.013>
- Totoki, Y., Tatsuno, K., Covington, K.R., Ueda, H., Creighton, C.J., Kato, M., Tsuji, S., Donehower, L.A., Slagle, B.L., Nakamura, H., Yamamoto, S., Shinbrot, E., Hama, N., Lehmkuhl, M., Hosoda, F., Arai, Y., Walker, K., Dahdouli, M., Gotoh, K., Nagae, G., Gingras, M.-C., Muzny, D.M., Ojima, H., Shimada, K., Midorikawa, Y., Goss, J.A., Cotton, R., Hayashi, A., Shibahara, J., Ishikawa, S., Guiteau, J., Tanaka, M., Urushidate, T., Ohashi, S., Okada, N., Doddapaneni, H., Wang, M., Zhu, Y., Dinh, H., Okusaka, T., Kokudo, N., Kosuge, T., Takayama, T., Fukayama, M., Gibbs, R.A., Wheeler, D.A., Aburatani, H., Shibata, T., 2014. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat. Genet.* 46, 1267–1273. <https://doi.org/10.1038/ng.3126>
- Trépo, C., Chan, H.L.Y., Lok, A., 2014. Hepatitis B virus infection. *The Lancet* 384, 2053–2063. [https://doi.org/10.1016/S0140-6736\(14\)60220-8](https://doi.org/10.1016/S0140-6736(14)60220-8)
- Tsai, K.-N., Kuo, C.-F., Ou, J.-H.J., 2018. Mechanisms of Hepatitis B Virus Persistence. *Trends Microbiol.* 26, 33–42. <https://doi.org/10.1016/j.tim.2017.07.006>
- Tu, T., Budzinska, M., Shackel, N., Urban, S., 2017. HBV DNA Integration: Molecular Mechanisms and Clinical Implications. *Viruses* 9, 75. <https://doi.org/10.3390/v9040075>
- Tu, T., Thomas, Budzinska, M.A., Shackel, N.A., Jilbert, A.R., 2015. Conceptual models for the initiation of hepatitis B virus-associated hepatocellular carcinoma. *Liver Int.* 35, 1786–1800. <https://doi.org/10.1111/liv.12773>
- Tu, T., Budzinska, M.A., Vondran, F.W.R., Shackel, N.A., Urban, S., 2018. Hepatitis B Virus DNA Integration Occurs Early in the Viral Life Cycle in an *In Vitro* Infection Model via Sodium Taurocholate Cotransporting Polypeptide-Dependent Uptake of Enveloped Virus Particles. *J. Virol.* 92, e02007-17. <https://doi.org/10.1128/JVI.02007-17>
- Tu, T., Mason, W.S., Clouston, A.D., Shackel, N.A., McCaughan, G.W., Yeh, M.M., Schiff, E.R., Ruszkiewicz, A.R., Chen, J.W., Harley, H.A.J., Stroehrer, U.H., Jilbert, A.R., 2015. Clonal

BIBLIOGRAPHIE

- expansion of hepatocytes with a selective advantage occurs during all stages of chronic hepatitis B virus infection. *J. Viral Hepat.* 22, 737–753. <https://doi.org/10.1111/jvh.12380>
- Valaydon, Z.S., Locarnini, S.A., 2017. The virological aspects of hepatitis B. *Best Pract. Res. Clin. Gastroenterol.* 31, 257–264. <https://doi.org/10.1016/j.bpg.2017.04.013>
- Villanueva, A., 2019. Hepatocellular Carcinoma. *N. Engl. J. Med.* 380, 1450–1462. <https://doi.org/10.1056/NEJMra1713263>
- von Felden, J., 2020. New systemic agents for hepatocellular carcinoma: an update 2020. *Curr. Opin. Gastroenterol.* 1. <https://doi.org/10.1097/MOG.0000000000000626>
- Voulgaris, T., Papatheodoridi, M., Lampertico, P., Papatheodoridis, G.V., 2020. Clinical utility of hepatocellular carcinoma risk scores in chronic hepatitis B. *Liver Int.* 40, 484–495. <https://doi.org/10.1111/liv.14334>
- Walters, K.-A., Joyce, M.A., Addison, W.R., Fischer, K.P., Tyrrell, D.L.J., 2004. Superinfection Exclusion in Duck Hepatitis B Virus Infection Is Mediated by the Large Surface Antigen. *J. Virol.* 78, 7925–7937. <https://doi.org/10.1128/JVI.78.15.7925-7937.2004>
- Wang, H., Liao, P., Zeng, S.X., Lu, H., 2019. It takes a team: a gain-of-function story of p53-R249S. *J. Mol. Cell Biol.* 11, 277–283. <https://doi.org/10.1093/jmcb/mjy086>
- Wang, J., Chenivesse, X., Henglein, B., Brechot, C., 1990. Hepatitis B virus integration in a cyclin A gene in a hepatocellular carcinoma 343, 3.
- Wei, Y., Neuveut, C., Tiollais, P., Buendia, M.-A., 2010. Molecular biology of the hepatitis B virus and role of the X gene. *Pathol. Biol.* 58, 267–272. <https://doi.org/10.1016/j.patbio.2010.03.005>
- Wong, G.L.-H., Chan, H.L.-Y., Mak, C.W.-H., Lee, S.K.-Y., Ip, Z.M.-Y., Lam, A.T.-H., Lu, H.W.-H., Leung, J.M.-S., Lai, J.W.-Y., Lo, A.O.-S., Chan, H.-Y., Wong, V.W.-S., 2013. Entecavir treatment reduces hepatic events and deaths in chronic hepatitis B patients With liver cirrhosis: *Hepatology*. *Hepatology* 58, 1537–1547. <https://doi.org/10.1002/hep.26301>
- Woo, H.G., Wang, X.W., Budhu, A., Kim, Y.H., Kwon, S.M., Tang, Z., Sun, Z., Harris, C.C., Thorgeirsson, S.S., 2011. Association of TP53 Mutations With Stem Cell-Like Gene Expression and Survival of Patients With Hepatocellular Carcinoma. *Gastroenterology* 140, 1063-1070.e8. <https://doi.org/10.1053/j.gastro.2010.11.034>
- Wooddell, C.I., Yuen, M.-F., Chan, H.L.-Y., Gish, R.G., Locarnini, S.A., Chavez, D., Ferrari, C., Given, B.D., Hamilton, J., Kanner, S.B., Lai, C.-L., Lau, J.Y.N., Schluep, T., Xu, Z., Lanford, R.E., Lewis, D.L., 2017. RNAi-based treatment of chronically infected patients and chimpanzees reveals that integrated hepatitis B virus DNA is a source of HBsAg. *Sci. Transl. Med.* 9, eaan0241. <https://doi.org/10.1126/scitranslmed.aan0241>
- World Health Organization, 2017. Global hepatitis report, 2017.
- World Health Organization, 2016a. Global health sector strategy on viral hepatitis 2016-2021.
- World Health Organization, 2016b. Global Health Estimates 2015: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2015.
- Wu, C.-Y., Lin, J.-T., Ho, H.J., Su, C.-W., Lee, T.-Y., Wang, S.-Y., Wu, C., Wu, J.-C., 2014. Association of Nucleos(t)ide Analogue Therapy With Reduced Risk of Hepatocellular Carcinoma in Patients With Chronic Hepatitis B—A Nationwide Cohort Study. *Gastroenterology* 147, 143-151.e5. <https://doi.org/10.1053/j.gastro.2014.03.048>
- Wu, J.-C., Huang, Y.-H., Chau, G.-Y., Su, C.-W., Lai, C.-R., Lee, P.-C., Huo, T.-I., Sheen, I.-J., Lee, S.-D., Lui, W.-Y., 2009. Risk factors for early and late recurrence in hepatitis B-related hepatocellular carcinoma. *J. Hepatol.* 51, 890–897. <https://doi.org/10.1016/j.jhep.2009.07.009>

- Xu, W., Yu, J., Wong, V.W.-S., 2017. Mechanism and prediction of HCC development in HBV infection. *Best Pract. Res. Clin. Gastroenterol.* 31, 291–298. <https://doi.org/10.1016/j.bpg.2017.04.011>
- Yan, H., Yang, Y., Zhang, L., Tang, G., Wang, Y., Xue, G., Zhou, W., Sun, S., 2015. Characterization of the genotype and integration patterns of hepatitis B virus in early- and late-onset hepatocellular carcinoma: *HEPATOLOGY*, Vol. XX, No. X, 2015. *Hepatology* 61, 1821–1831. <https://doi.org/10.1002/hep.27722>
- Yang, J., Trépo, E., Nahon, P., Cao, Q., Moreno, C., Letouzé, E., Imbeaud, S., Bayard, Q., Gustot, T., Deviere, J., Bioulac-Sage, P., Calderaro, J., Ganne-Carrié, N., Laurent, A., Blanc, J.F., Guyot, E., Sutton, A., Ziol, M., Zucman-Rossi, J., Nault, J., 2019a. A 17-Beta-Hydroxysteroid Dehydrogenase 13 Variant Protects From Hepatocellular Carcinoma Development in Alcoholic Liver Disease. *Hepatology* hep.30623. <https://doi.org/10.1002/hep.30623>
- Yang, J., Trépo, E., Nahon, P., Cao, Q., Moreno, C., Letouzé, E., Imbeaud, S., Gustot, T., Deviere, J., Debette, S., Amouyel, P., Bioulac-Sage, P., Calderaro, J., Ganne-Carrié, N., Laurent, A., Blanc, J.F., Guyot, E., Sutton, A., Ziol, M., Zucman-Rossi, J., Nault, J., 2019b. PNPLA3 and TM6SF2 variants as risk factors of hepatocellular carcinoma across various etiologies and severity of underlying liver diseases. *Int. J. Cancer* 144, 533–544. <https://doi.org/10.1002/ijc.31910>
- Yang, J.D., Hainaut, P., Gores, G.J., Amadou, A., Plymoth, A., Roberts, L.R., 2019. A global view of hepatocellular carcinoma: trends, risk, prevention and management. *Nat. Rev. Gastroenterol. Hepatol.* 16, 589–604. <https://doi.org/10.1038/s41575-019-0186-y>
- Yang, L., Ye, S., Zhao, X., Ji, L., Zhang, Y., Zhou, P., Sun, J., Guan, Y., Han, Y., Ni, C., Hu, X., Liu, W., Wang, H., Zhou, B., Huang, J., 2018. Molecular Characterization of HBV DNA Integration in Patients with Hepatitis and Hepatocellular Carcinoma. *J. Cancer* 9, 3225–3235. <https://doi.org/10.7150/jca.26052>
- Yang, W., Mason, W.S., Summers, J., 1996. Covalently closed circular viral DNA formed from two types of linear DNA in woodchuck hepatitis virus-infected liver. *J. Virol.* 70, 4567–4575. <https://doi.org/10.1128/JVI.70.7.4567-4575.1996>
- Yang, W., Summers, J., 1999. Integration of Hepadnavirus DNA in Infected Liver: Evidence for a Linear Precursor. *J. Virol.* 73, 9710–9717. <https://doi.org/10.1128/JVI.73.12.9710-9717.1999>
- Yang, W., Summers, J., 1998. Infection of Ducklings with Virus Particles Containing Linear Double-Stranded Duck Hepatitis B Virus DNA: Illegitimate Replication and Reversion. *J. Virol.* 72, 8710–8717. <https://doi.org/10.1128/JVI.72.11.8710-8717.1998>
- Yang, X., Wu, L., Lin, J., Wang, A., Wan, X., Wu, Y., Robson, S.C., Sang, X., Zhao, H., 2017. Distinct hepatitis B virus integration patterns in hepatocellular carcinoma and adjacent normal liver tissue: HBV integration hotspot lncRNAs. *Int. J. Cancer* 140, 1324–1330. <https://doi.org/10.1002/ijc.30547>
- Yi, K., Ju, Y.S., 2018. Patterns and mechanisms of structural variations in human cancer. *Exp. Mol. Med.* 50, 98. <https://doi.org/10.1038/s12276-018-0112-3>
- Yim, H.J., Lok, A.S.-F., 2006. Natural history of chronic hepatitis B virus infection: What we knew in 1981 and what we know in 2005. *Hepatology* 43, S173–S181. <https://doi.org/10.1002/hep.20956>
- Yip, T.C.-F., Chan, H.L.-Y., Wong, V.W.-S., Tse, Y.-K., Lam, K.L.-Y., Wong, G.L.-H., 2017. Impact of age and gender on risk of hepatocellular carcinoma after hepatitis B surface antigen seroclearance. *J. Hepatol.* 67, 902–908. <https://doi.org/10.1016/j.jhep.2017.06.019>
- Yip, T.C.-F., Wong, G.L.-H., Chan, H.L.-Y., Tse, Y.-K., Lam, K.L.-Y., Lui, G.C.-Y., Wong, V.W.-S., 2019. HBsAg seroclearance further reduces hepatocellular carcinoma risk after complete viral

- suppression with nucleos(t)ide analogues. *J. Hepatol.* 70, 361–370. <https://doi.org/10.1016/j.jhep.2018.10.014>
- Yizhak, K., Aguet, F., Kim, J., Hess, J.M., Kübler, K., Grimsby, J., Frazer, R., Zhang, H., Haradhvala, N.J., Rosebrock, D., Livitz, D., Li, X., Arich-Landkof, E., Shores, N., Stewart, C., Segrè, A.V., Branton, P.A., Polak, P., Ardlie, K.G., Getz, G., 2019. RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science* 364, eaaw0726. <https://doi.org/10.1126/science.aaw0726>
- Yoo, S., Wang, W., Wang, Q., Fiel, M.I., Lee, E., Hiotis, S.P., Zhu, J., 2017. A pilot systematic genomic comparison of recurrence risks of hepatitis B virus-associated hepatocellular carcinoma with low- and high-degree liver fibrosis. *BMC Med.* 15, 214. <https://doi.org/10.1186/s12916-017-0973-7>
- Zapatka, M., Borozan, I., Brewer, D.S., Iskar, M., Grundhoff, A., Alawi, M., Desai, N., Sülthmann, H., Moch, H., Cooper, C.S., Eils, R., Ferretti, V., Lichter, P., 2020. The landscape of viral associations in human cancers. *Nat. Genet.* 52, 320–330. <https://doi.org/10.1038/s41588-019-0558-9>
- Zhang, W., He, H., Zang, M., Wu, Q., Zhao, H., Lu, L., Ma, P., Zheng, H., Wang, N., Zhang, Y., He, S., Chen, X., Wu, Z., Wang, X., Cai, J., Liu, Z., Sun, Z., Zeng, Y.-X., Qu, C., Jiao, Y., 2017. Genetic Features of Aflatoxin-Associated Hepatocellular Carcinoma. *Gastroenterology* 153, 249-262.e2. <https://doi.org/10.1053/j.gastro.2017.03.024>
- Zhao, L.-H., Liu, X., Yan, H.-X., Li, W.-Y., Zeng, X., Yang, Yuan, Zhao, J., Liu, S.-P., Zhuang, X.-H., Lin, C., Qin, C.-J., Zhao, Y., Pan, Z.-Y., Huang, G., Liu, H., Zhang, J., Wang, R.-Y., Yang, Yun, Wen, W., Lv, G.-S., Zhang, H.-L., Wu, H., Huang, S., Wang, M.-D., Tang, L., Cao, H.-Z., Wang, L., Lee, T.-L., Jiang, H., Tan, Y.-X., Yuan, S.-X., Hou, G.-J., Tao, Q.-F., Xu, Q.-G., Zhang, X.-Q., Wu, M.-C., Xu, X., Wang, J., Yang, H.-M., Zhou, W.-P., Wang, H.-Y., 2016. Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nat. Commun.* 7, 12992. <https://doi.org/10.1038/ncomms12992>
- Zhu, A.X., Finn, R.S., Edeline, J., Cattani, S., Ogasawara, S., Palmer, D., Verslype, C., Zagonel, V., Fartoux, L., Vogel, A., Sarker, D., Verset, G., Chan, S.L., Knox, J., Daniele, B., Webber, A.L., Ebbinghaus, S.W., Ma, J., Siegel, A.B., Cheng, A.-L., Kudo, M., Alistar, A., Asselah, J., Blanc, J.-F., Borbath, I., Cannon, T., Chung, K., Cohn, A., Cosgrove, D.P., Damjanov, N., Gupta, M., Karino, Y., Karwal, M., Kaubisch, A., Kelley, R., Van Laethem, J.-L., Larson, T., Lee, J., Li, D., Manhas, A., Manji, G.A., Numata, K., Parsons, B., Paulson, A.S., Pinto, C., Ramirez, R., Ratnam, S., Rizell, M., Rosmorduc, O., Sada, Y., Sasaki, Y., Stal, P.I., Strasser, S., Trojan, J., Vaccaro, G., Van Vlierberghe, H., Weiss, A., Weiss, K.-H., Yamashita, T., 2018. Pembrolizumab in patients with advanced hepatocellular carcinoma previously treated with sorafenib (KEYNOTE-224): a non-randomised, open-label phase 2 trial. *Lancet Oncol.* 19, 940–952. [https://doi.org/10.1016/S1470-2045\(18\)30351-6](https://doi.org/10.1016/S1470-2045(18)30351-6)
- Zhu, A.X., Kang, Y.-K., Yen, C.-J., Finn, R.S., Galle, P.R., Llovet, J.M., Assenat, E., Brandi, G., Pracht, M., Lim, H.Y., Rau, K.-M., Motomura, K., Ohno, I., Merle, P., Daniele, B., Shin, D.B., Gerken, G., Borg, C., Hiriart, J.-B., Okusaka, T., Morimoto, M., Hsu, Y., Abada, P.B., Kudo, M., 2019. Ramucicromab after sorafenib in patients with advanced hepatocellular carcinoma and increased α -fetoprotein concentrations (REACH-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* 20, 282–296. [https://doi.org/10.1016/S1470-2045\(18\)30937-9](https://doi.org/10.1016/S1470-2045(18)30937-9)
- Zhu, M., Lu, T., Jia, Y., Luo, X., Gopal, P., Li, L., Odewole, M., Renteria, V., Singal, A.G., Jang, Y., Ge, K., Wang, S.C., Sorouri, M., Parekh, J.R., MacConmara, M.P., Yopp, A.C., Wang, T., Zhu, H., 2019. Somatic Mutations Increase Hepatic Clonal Fitness and Regeneration in Chronic Liver Disease. *Cell* 177, 608-621.e12. <https://doi.org/10.1016/j.cell.2019.03.026>
- Zhu, Y.-Z., Zhu, R., Fan, J., Pan, Q., Li, H., Chen, Q., Zhu, H.-G., 2010. Hepatitis B virus X protein induces hypermethylation of p16^{INK4A} promoter via DNA methyltransferases in the early stage

- of HBV-associated hepatocarcinogenesis. *J. Viral Hepat.* 17, 98–107.
<https://doi.org/10.1111/j.1365-2893.2009.01156.x>
- Zoulim, F., Lebossé, F., Levrero, M., 2016. Current treatments for chronic hepatitis B virus infections. *Curr. Opin. Virol.* 18, 109–116. <https://doi.org/10.1016/j.coviro.2016.06.004>
- Zucman-Rossi, J., Villanueva, A., Nault, J.-C., Llovet, J.M., 2015. Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. *Gastroenterology* 149, 1226-1239.e4.
<https://doi.org/10.1053/j.gastro.2015.05.061>
- Zulfiqar, H.F., Javed, A., Sumbal, Afroze, B., Ali, Q., Akbar, K., Nadeem, T., Rana, M.A., Nazar, Z.A., Nasir, I.A., Husnain, T., 2017. HIV Diagnosis and Treatment through Advanced Technologies. *Front. Public Health* 5, 32. <https://doi.org/10.3389/fpubh.2017.00032>