Role of the DEAD-box Helicases DDX5 and DDX17 in Hepatitis B Virus RNA processing

Fleur Chapus

N°d'ordre NNT : 2020LYSE098

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale N° 340
Biologie Moléculaire Intégrative et Cellulaire (BMIC)

Spécialité de doctorat : Virologie et Biologie Moléculaire
Discipline : Infectiologie

Soutenue publiquement le 28 Mai 2020 par
Fleur Chapus

Role of the DEAD-box Helicases DDX5 and DDX17 in Hepatitis B Virus RNA processing

---

Directeurs de Thèse : **Professeur Fabien Zoulim** et **Docteur Barbara Testoni**

Devant le jury composé de :

Mme. Docteur **Christine Neuveut**, DR INSERM, IGH, Montpellier	Rapporteur
Mr. Professeur **Michael Nassal**, PU, Université de Fribourg en Brisgau	Rapporteur
Mme. Docteur **Sofia Pérez-del-Pulgar**, Hospital Clinic de Barcelona	Examinateur
Mr. Professeur **Massimo Levrero**, PU-PH, CRCL/HCL, Lyon	Examinateur
Mme. Docteur **Barbara Testoni**, CR INSERM, CRCL, Lyon	Co-directrice de Thèse
Mr. Professeur **Fabien Zoulim**, PU-PH, CRCL/HCL, Lyon	Directeur de Thèse

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université	**M. le Professeur Frédéric FLEURY**
Vice-président du Conseil Académique	M. le Professeur Hamda BEN HADID
Vice-président du Conseil d'Administration	M. le Professeur Didier REVEL
Vice-président du Conseil Formation et Vie Universitaire	M. le Professeur Philippe CHEVALIER
Vice-président de la Commission Recherche	M. Jean-François MORNEX
Directeur Général des Services	M. Damien VERHAEGHE

COMPOSANTES SANTÉ

Faculté de Médecine Lyon Est – Claude Bernard	Doyen : M. Gilles RODES
Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux	Doyenne : Mme. Carole BURILLON
Faculté d'Odontologie	Doyenne : Mme Dominique SEUX
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : Mme. Christine VINCIGUERRA
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. Xavier PERROT
Département de formation et Centre de Recherche en Biologie Humaine	Directeur : Mme. Anne-Marie SCHOTT

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

UFR Biosciences	Directrice : Mme Kathrin GIESELER
Département Génie Electrique et des Procédés	Directrice : Mme. Rosaria FERRIGNO
Département Informatique	Directeur : M. Behzad SHARIAT
Département Mécanique	Directeur : M. Marc BUFFAT
UFR – Faculté des Sciences	Administrateur provisoire : M. Bruno ANDRIOLETTI
UFR STAPS	Directeur : M. Y. VANPOULLE
Observatoire de Lyon	Directrice : Mme Isabelle DANIEL
Ecole Polytechnique Universitaire Lyon 1	Directeur : M. Emmanuel PERRIN
Ecole Supérieure de Chimie, Physique, Electronique (CPE Lyon)	Directeur : M. Gérard PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur : M. Christophe VITON
Institut de Science Financière et d'Assurances	Directeur : M. Nicolas LEBOISNE
Ecole Supérieure du Professorat et de l'Education (ESPE)	Administrateur provisoire : M. Pierre CHAREYRON

Résumé

Rôle des hélicases DDX5 et DDX17 dans la régulation transcriptionnelle et la maturation des ARN du Virus de l'hépatite B.

La chronicité du virus de l'hépatite B (VHB) repose sur la persistance de l'ADN circulaire et clos de manière covalente (ADNccc) dans le noyau des hépatocytes infectés. Le génome viral présente une structure chromatinisée sujette à des régulations épigénétiques impactant son activité biologique à différents niveaux. Une meilleure connaissance des facteurs cellulaires orchestrant la régulation transcriptionnelle et post-transcriptionnelle de l'ADNccc est fondamentale dans la compréhension des mécanismes à l'origine de la persistance du VHB.

Afin d'identifier les acteurs cellulaires impliqués dans la biologie de l'ADNccc, un ambitieux projet de protéomique de l'ADNccc (ChROP) a été initié par le Dr. Barbara Testoni. Parmi les candidats identifiés, les hélicases à ARN DDX5 et DDX17 ont particulièrement attiré notre attention. DDX5 et DDX17 jouent un rôle crucial dans la régulation de la transcription et le métabolisme des ARN. Nous avons donc évalué leur rôle dans la régulation transcriptionnelle de l'ADNccc et le métabolisme des ARN viraux. Afin d'étudier de manière précise les différents transcrits du VHB, une technique de 5' RACE a été mise au point au laboratoire par le Dr. Bernd Stadelmayer, et a fait l'objet d'une publication. Cette technique a permis l'étude des ARN du VHB dans un contexte de déplétion des hélicases DDX5 et DDX17.

Par ailleurs, DDX5/17 appartiennent au complexe insulateur de CCCTC-binding protein (CTCF). Nous avons donc en parallèle étudié le rôle de CTCF dans la biologie de l'ADNccc et le métabolisme des ARN viraux.

Dans des cellules HepG2-NTCP et des hépatocytes primaires humains infectés par le VHB, la répression de DDX5/17 entraine un raccourcissement de tous les ARN viraux. Des séquençages de dernière et troisième génération ont permis l'identification de variants d'épissage alternatif ainsi qu'une utilisation différentielle du site de polyadénylation lors de la transcription des transcrits viraux. Des expériences d'immunoprécipitation de l'ARN ont montré que DDX5 et DDX17 s'associent directement aux ARN viraux et recrutent CPSF6 et NUDT21, deux facteurs impliqués dans le choix du site de polyadénylation. Par ailleurs, nous avons identifié des sites de liaison de CTCF sur le génome du VHB et par mutagénèse dirigé,

nous avons mis en évidence que la mutation de ces sites impacte le recrutement de CTCF et de DDX5/17 sur l'ADNccc et affecte métabolisme des ARN viraux.

L'ensemble de ces données met donc en lumière un rôle essentiel de DDX5 et DDX17 dans la maturation des ARN viraux, en complexe avec la protéine insulatrice CTCF et des facteurs de temrination, à l'interface entre l'ADNccc et les transcripts du VHB.

Mots clés : Virus de l'hépatite B, ADNccc, métabolisme des ARN, Hélicases à ARN à motif DEAD, CTCF, 5' RACE, Immunoprecipitation de la Chromatine et de l'ARN

Abstract

Role of the DEAD-box helicases DDX5 and DDX17 in HBV transcriptional regulation and RNA processing

Chronicity of hepatitis B virus (HBV) infection hinges on the persistence of covalently-closed-circular DNA (cccDNA) in the nucleus of infected hepatocytes. The viral genome associates with histones and non-histone proteins to build a chromatin structure that is subjected to epigenetic regulation translating into different levels of biological activity. A better understanding of the host factors orchestrating HBV minichromosome transcriptional regulation and RNA processing is fundamental for deciphering the mechanisms at the basis of HBV persistence and reactivation.

In order to identify the cellular factors regulating cccDNA biology, an ambitious project of cccDNA proteomics (ChroP) has been initiated by Dr. Barbara Testoni. Among the identified cccDNA-associated proteins, the DEAD-box RNA helicases DDX5 and DDX17 particularly interested us for their driving role in mammalian transcriptional regulation and RNA metabolism. Thus, we investigated their role in cccDNA transcriptional activity regulation and HBV RNA processing. Precise characterization of HBV transcripts was performed with a 5' RACE approach set up and published in our lab by Dr. Bernd Stadelmayer. This technique was applied to study viral transcript in a context of DDX5/17 depletion. Furthermore, DDX5/17 belong to the insulator complex CCCTC-binding protein (CTCF). We therefore investigated the role of CTCF in cccDNA biology and viral RNA metabolism.

In HBV infectedHepG2-NTCP and Primary Human Hepatocytes, siRNA knockdown of DDX5/17 led to a shortening of all the viral transcripts, together with an increase in viral transcript levels and viral particles accumulation in the cytoplasm, without affecting the global level of cccDNA. Next and third generation sequencing allowed the identification of alternative splicing of pgRNA-derived spliced variants and differential usage of polyadenylation site during HBV RNA transcription. Moreover, RNA immunoprecipitation of DDX5 and DDX17 revealed that both of these proteins are directly associated to the viral transcripts and recruit two factors, CPSF6 and NUDT21, involved in alternative polyadenylation site choice. Moreover, we identified CTCF binding sites on HBV genome and by site directed mutagenesis we showed that

mutations in CTCF binding sites affect CTCF and DDX5/17 recruitment to cccDNA and subsequently impact HBV RNA processing.

Altogether, our data highlight an essential role of DDX5 and DDX17 in the fine tuning of HBV RNA processing, in complex with the insulator protein CTCF and termination factors at the interface between cccDNA and HBV transcripts.

Key words: Hepatitis B virus, cccDNA, RNA processing, DEAD-Box RNA helicases, CTCF, 5' RACE, Chromatin and RNA immunoprecipitation, RNA sequencing

Résumé Substantiel

Rôle des hélicases à ARN DDX5 et DDX17 dans la régulation transcriptionnelle du minichromosome du virus de l'hépatite B et dans la maturation des ARN viraux.

Travaux de thèse réalisés par Fleur Chapus, sous la direction du Pr. Fabien Zoulim et du Dr. Barbara Testoni, au sein du Centre de Recherche en Cancérologie de Lyon (CRCL), UMR INSERM U1052, CNRS 5286, Centre Léon Bérard, dirigé par Dr. Patrick Mehlen

Introduction

Le Virus de l'Hépatite B (VHB) infecte chroniquement plus de 250 millions d'individus dans le monde. Les porteurs chroniques du VHB ont un risque accru de développer de sévères pathologies hépatiques, notamment une cirrhose voire un carcinome hépatocellulaire, qui est le deuxième cancer le plus meurtrier dans le monde. Malgré l'existence d'un vaccin prophylactique efficace et de traitements antiviraux, l'hépatite B demeure un problème majeur de santé publique qui cause plus de 800 000 décès chaque année (Trépo et al., 2014). La molécule en cause de la persistance du virus malgré les traitements est un épisome viral maintenu dans le noyau des hépatocytes infectés appelé ADNccc pour ADN circulaire et clos de manière covalente. Les stratégies thérapeutiques actuelles cherchent donc à cibler l'ADNccc, soit par destruction, soit par répression transcriptionnelle, dans le but de développer un traitement curatif de l'hépatite B chronique (Nassal, 2015a).

Le VHB est le plus petit virus à ADN identifié jusqu'à présent. C'est un virus enveloppé de la famille des *Hepadnaviridae* qui comporte un génome partiellement double brin et circulaire relaxé, appelé ADNrc, qui comporte 3.2 kb. Après son entrée dans le noyau, l'ADNrc est complété et chromatinisé afin de former un minichromosome viral: l'ADNccc (C. Thomas Bock et al., 2001). Ce dernier est l'unique intermédiaire réplicatif transcriptionnellement actif qui code les 6 ARN viraux, parmi lesquels l'ARN prégénomique (ARNpg) est l'ARN intermédiaire de réplication. En effet, après avoir été exporté dans le cytoplasme, l'ARNpg est rétro-transcrit en ADNrc et simultanément encapsidé afin de former de nouvelles nucléocapsides matures (Beck and Nassal, 2007).

La transcription virale permet la synthèse de six transcrits: un transcrit de 0.7 kb appelé X, deux transcrits de 2.1 kb nommés PreS2 et S, un transcrit de 2.4 kb nommé PreS1 et deux

transcrits dont la longueur est supérieure à celle du génome, appelés ARN précore et ARN prégénomique mesurant 3.5 kb (Seeger and Mason, 2000a). L'ADNccc contient également tous les éléments régulateurs nécessaires à une fine régulation de la transcription. En effet, la transcription est modulée par quatre promoteurs et deux enhancers (Moolla et al., 2002). De plus, bien que chaque ARN viral soit initié en 5' au niveau d'un site d'initiation de la transcription qui lui est propre, tous les ARN viraux sont polyadénylés en 3' au niveau d'un site commun de polyadénylation. En outre, l'ARNpg et l'ARN précore sont initiés environ 100 paires de bases en amont du site de polyadénylation, puis sont transcrits tout le long du génome, et se terminent à leur deuxième rencontre avec le site de polyadénylation. Par ce mécanisme encore très peu défini, les deux ARN de 3.5 kb comportent tous deux une séquence redondante correspond à la région transcrite deux fois (Beck and Nassal, 2007). D'autre part, de nombreux variants d'épissages ont été identifiés *in vivo* dans le sérum de patients chroniquement infectés par le VHB, ainsi qu'*in vitro* en culture cellulaire. 16 variants dérivent de l'épissage de l'ARNpg et les sites donneurs et accepteurs d'épissage varient en nature et en longueur en fonction des génotypes viraux considérés (Candotti and Allain, 2016a). La complexité de l'organisation du génome viral, ainsi que sa similarité avec la chromatine cellulaire et avec les processus de maturation des ARN, suggèrent une coopération de nombreux facteurs cellulaires et viraux afin de précisément orchestrer une transcription virale optimale.

Afin d'identifier les acteurs cellulaires qui régulent l'activité transcriptionnelle de l'ADNccc, ainsi que les évènement co-transcriptionnels associés, un projet visant à identifier les partenaires protéiques de l'ADNccc a été réalisé par le Dr. Barbara Testoni. Une majorité de protéines liant l'ADN et l'ARN et impliquées dans la biologie de la chromatine et la maturation des ARN ont été listées. Parmi ces candidats, les hélicases à ARN constituent la plus grande famille d'enzymes impliquées dans le métabolisme des ARN cellulaires (Bourgeois et al., 2016a). Elles se déclinent en six sous-familles, parmi lesquelles la famille des hélicases à ARN à motif DEAD (Asparagine-Glutamine-Alanine-Asparagine) comporte le plus de membres. Les deux paralogues DDX5 et DDX17 sont des membres prototypiques des hélicases à motif DEAD. Ces deux protéines sont connues pour participer à de nombreux processus de la régulation de l'expression des gènes de la cellule (Giraud et al., 2018a). Notamment, DDX5 et DDX17 sont importantes dans la maturation de l'ARN pré-messager, au niveau transcriptionnel puis lors

de l'épissage alternatif (Dardenne et al., 2014a; Giraud et al., 2018a). De plus, ces deux hélicases sont impliquées dans le métabolisme des ARN de nombreux virus (Wenyu Cheng et al., 2018a). Notamment, une récente étude identifie DDX5 comme répresseur transcriptionnel de l'ADNccc, en particulier à travers le complexe PRC2 et le dépôt de marques épigénétiques répressives H3K27me3 (Zhang et al., 2016a).

Le rôle de DDX5 et de DDX17 dans la maturation des ARN viraux reste donc à être déterminé.

Par ailleurs, DDX5 et DDX17 sont nécessaires aux fonctions insulatrices de la protéine CTCF (H. Yao et al., 2010a). De manière intéressante, CTCF est également requise pour l'épissage alternatif (Shukla et al., 2011a) et est associé à l'organisation tri-dimensionnelle d'épisomes viraux (Lang et al., 2017; Mehta et al., 2015; Pentland et al., 2018a; Washington et al., 2018). Nous nous sommes donc également intéressés au rôle de CTCF dans la biologie de l'ADNccc et dans la maturation des ARN du VHB.

Objectifs

Les objectifs de mes travaux de thèse ont donc porté sur : i) le rôle de DDX5 et DDX17 dans le métabolisme des ARN viraux ; iii) l'impact de la déplétion de ces hélicases dans la réplication virale ; iii) le rôle de CTCF dans la régulation transcriptionnelle des ARN viraux.

Matériel et Méthode

Des cellules HepG2-NTCP, exprimant stablement le récepteur NTCP du VHB, et des Hépatocytes Primaires Humains (PHH) provenant de résection de patients, ont été infectés par le VHB à une multiplicité d'infection de 250 équivalents génomes par cellule. Après quatre jours d'infection, lorsque celle-ci est considérée établie, les cellules ont été transfectées avec des petits ARN interférents ciblant de manière concomitante DDX5 et DDX17 afin de prévenir d'une potentielle redondance fonctionnelle. Après validation de l'efficacité de la déplétion de DDX5 et DDX17 au niveau ARNm par RT-qPCR et protéique par Western-blot, les paramètres viraux ont été évalués. L'ARN total et les ARN 3.5 kb, ainsi que l'ADN total et l'ADNccc ont été quantifiés par (RT)-qPCR. Afin de discriminer les différents ARN viraux, ceux-ci ont été soumis à des Northern-blot. Par ailleurs, l'activité transcriptionnelle de l'ADNccc a été évaluée par capture des ARN naissants et RT-qPCR spécifique des ARN viraux. Afin de caractériser précisément les ARN viraux, un séquençage d'ARN a été fait. L'étude qualitative précise des ARN viraux a été complétée par des expériences de 5' RACE et 3' RACE, deux techniques

permettant d'amplifier spécifiquement les ARN depuis leur extrémités 5' et 3'. Les produits de 3'RACE ont par la suite été séquencés par la technologie Oxford Nanopore. Finalement, la morphologie des cellules infectées et déplétées ou non pour DDX5 et DDX17 a été comparée par microscopie électronique à transmission (MET). Un immunomarquage aux billes d'or des protéines virales HBc et HBs, ainsi qu'un double marquage a été réalisé dans le but de valider la présence de particules virales cytoplasmiques en MET. Ces particules ont ensuite été caractérisées biochimiquement par isolation des capsides cytoplasmiques suivie d'une migration en gel d'agarose en condition native ou par extraction d'ADN et ARN encapsidés suivies de (RT)-qPCR spécifiques ou de séquençage Nanopore. Enfin, la présence de DDX5 et DDX17 sur l'ADNccc et les ARN viraux a été étudiée par Immunoprécipitation de la chromatine et par immunoprécipitation des ARN viraux après fixation aux UV.

Par ailleurs, une mutagénèse dirigée spécifiquement sur les sites de liaison à CTCF préalablement identifiés grâce à un outils de prédiction *in sillico*, a permis de générer des molécules d'ADNccc-like par la technique Minicircle. L'étude de l'association de CTCF à l'ADNccc-like sauvage ou mutant a été réalisée par ChIP. Les paramètres viraux ont été étudiés de la même manière qu'explicité précédemment.

Résultats

Après une déplétion efficace de DDX5 et DDX17, nous avons observé en MET une accumulation de vésicules cytoplasmiques contenant des particules virales. Nous avons confirmé biochimiquement cette augmentation et évalué que deux fois plus de particules contenant HBc et HBs étaient présentes dans le cytoplasme des cellules déplétées pour DDX5 et DDX17. Deux fois plus d'ARN et ADN encapsidés ont été quantifiés dans ces particules cytoplasmiques, et sont qualitativement identiques (caractérisation par séquençage Nanopore). Par ailleurs, l'inhibition de l'expression de DDX5 et DDX17 entraine une augmentation globale des niveaux d'ARN viraux, et de leur transcription, sans modification du nombre de copies d'ADNccc par cellule. Qualitativement, nous avons observé que les ARN du VHB étaient plus courts en condition siDDX5/17. Un séquençage des ARN totaux par la technique Illumina nous a permis d'identifier cinq exons plus fréquemment exclus lorsque les hélicases sont absentes. Cependant, les transcrits épissés ne représentant qu'une faible proportion des transcrits viraux, l'exclusion plus fréquente observée n'explique que partiellement la diminution de taille des ARN du VHB en condition siDDX5/17. De manière

intéressante, l'ensemble des données obtenues par séquençage des ARN ainsi qu'une étude des variants d'épissages décrits dans la littérature a permis la génération du premier fichier d'annotation du génome complet du VHB.

Nous avons par la suite caractérisé plus précisément les ARN viraux en condition siDDX5/17. Bien qu'aucune différence ne soit observée en 5', l'extrémité 3' révèle une modification du phénotype lorsque les hélicases sont absentes. Afin d'étudier qualitativement le raccourcissement en 3' des ARN viraux en absence de DDX5/17, un séquençage en molécule unique a été réalisé par la technologie Oxford Nanopore. Les résultats suggèrent qu'en condition sauvage, les transcrits viraux ne s'arrêtent pas systématiquement au site de polyadénylation comme historiquement proposé, mais continuent sur plusieurs dizaines voire centaines de nucléotides, résultant en une région 3'UTR plus longue. L'absence de DDX5 et DDX17 permet en revanche une utilisation quasi systématique du site de polyadénylation du VHB. Par ailleurs, le séquençage au Nanopore nous a permis d'identifier précisément le site de polyadénylation des transcrits viraux 13 +/-1 nucléotides en aval du site de polyadénylation.

De plus, nous avons identifié que DDX5 et DDX17 interagissaient directement avec les ARN viraux, et notamment les ARN de 3.5 kb, ainsi qu'avec l'ADNccc. Ces dernières semblent recruter deux facteurs cellulaires impliqués dans le choix de sites de polyadénylation distaux : CPSF6 et NUDT21 (Hardy and Norbury, 2016a). En effet, la déplétion de DDX5 et DDX17 résulte en une diminution du recrutement de de CPSF6 et NUDT21 de plus de 50% sur les ARN viraux.

Parallèlement, nous avons démontré que la mutation des sites de liaison à CTCF affecte le recrutement de celui sur l'ADNccc, ainsi que le recrutement de DDX5 et DDX17. Par ailleurs, nous avons montré que la diminution du recrutement de CTCF, DDX5 et DDX17 sur l'ADNccc était corrélée à une modification de la taille des transcrits viraux.

Conclusion et perspectives

L'ensemble des résultats de cette étude suggère que DDX5 et DDX17 sont nécessaires au métabolisme des ARN viraux. De plus, les données générées suggèrent que CTCF constitue un point d'ancrage pour DDX5 et DDX17 sur l'ADNccc et que le complexe CTCF/DDX5/DDX17 semble nécessaire à la correcte maturation des ARN viraux. Cependant, le rôle direct de CTCF et de DDX5/17 reste à déterminer. Un recrutement de ces protéines sur des régions

spécifiques de l'ADNccc par la technologie CRISPR/dCas9 permettrait de mettre en évidence l'effet direct de ces protéines dans la biologie des ARN viraux.

Nous avons démontré également que DDX5 et DDX17 régulent l'épissage alternatif des ARN viraux en favorisant l'inclusion de cinq exons. Par ailleurs, DDX5 et DDX17 sont impliqués dans la terminaison de la transcription virale, notamment en recrutant CPSF6 et NUDT21, deux facteurs nécessaires à la reconnaissance de sites de polyadénylation alternatifs. Bien qu'une démonstration directe reste à établir, le raccourcissement des ARN viraux semble être corrélé à une augmentation des particules virales cytoplasmiques contenant de l'ARN et de l'ADN viraux. Le raccourcissement des ARN viraux en condition siDDX5/17 entraine un raccourcissement du 3'UTR, région essentielle au devenir et à la régulation du transcrit.

La validation de telles observations sur l'utilisation de sites alternatifs de polyadénylation chez les patients chroniquement infectés remettrait en question les biomarqueurs utilisés pour la détection des ARN viraux qui s'appuient sur une reconnaissance de ceux-ci par l'utilisation d'amorces chevauchant le site de polyadénylation et le début de la queue polyA. En effet, si 50% des transcrits ne terminent pas au site identifié de polyA, il est envisageable que les biomarqueurs actuels sous-estiment la quantité réelle des transcrits viraux chez les patients. De manière générale, l'identification de facteurs cellulaires dans la régulation de l'activité transcriptionnelle de l'ADNccc est nécessaire pour une meilleure compréhension du rôle des facteurs cellulaires dans la réplication virale, pour une avancée vers le développement de traitements curatifs efficaces contre le VHB.

Table of contents

Figures

Abbreviations

(G)TF: (Genera) Transcription Factor

3' / 5' UTR: 3' / 5' Untranslated Region

3C: Chromosome Conformation Capture

4C: Circular Chromosome Conformation Capture

5' / 3' RACE: 5' / 3' Rapide Amplification cDNA End

AAV: adeno associated virus

ASPP1/2: Apoptosis-Stimulating of P53 Protein ½

ATP: Adenosine Tri-Phosphate

BCP: Basic Core Promoter

bp: base pair

BRE: TFIIB Recognition Element

BS: Binding Site

CAGE: Cap Analysis Gene Expression

cccDNA: Covalently Closed Circular DNA

CCL2: C-C motif Chemokine Ligand 2

CDK: Cycline Dependent Kinase

CFI/II: Cleavage Factor I/II

CPA: Cleavage and Polyadenylation

CPSF: Cleavage and Polyadenylation Specificity Factor

CPSF6: Cleavage and Polyadenylation Specificity Factor 6

CRISPR/Cas9: Clustered Regularly Interspaced Short Palindromic repeats / Caspase 9

CstF: Cleavage Stimulatory Factor

CTCF: CCCTC-Binding Factor

CTD: Carboxy Trerminal Domain

Cul4: Cullin 4

DCE: Downstream Core Element

DDB1: DNA Damage-Binding protein 1

DDX17: DEAD-Box Helicase 17

DDX5: DEAD-Box Helicase 5

DEAD: Asp-Glu-Ala-Asp

DHBV: Duck Hepatitis B Virus

DMSO: Dimethyl Sulfoxide

DNA: Deoxyribonucleic Acide

DNMT: DNA methyltransferase

DPE: Downstream core promoter element

DR1/2: Direct Repeat 1/2

DSIF: DRB Sensitivity Inducing Factor

dsl-DNA: Double-Stranded Linear DNA

EBV: Epstein-Barr Virus

EnhI/II: Enhancer I/II

ESCRT: Export and Sorting Complex Required for Transport

ESE: Exon Splicing Enhancer

ESS: Exon Splicing Suppressos

FEN1: Flap Endonuclease 1

FRG: FAH-/- RAG2-/- FCIIγR-/-

GTase: Guanylyltransferase

H1/2/3/4/5: Histone 1/2/3/4/5

HBeAg: Hepatitis B e Antigen

HBsAg: Hepatitis B s Antigen

HBSP: Hepatitis B Spliced Protein

HBV: Hepatitis B Virus

HCC: Hepatocellular Carcinoma

HCV: Hepatitis C Virus

HDAC: Histone Deacetylase

HIV-1: Human Immunodeficiency Virus 1

HNF: Hepatocyte Nuclear Factor

hnRNP: heterogeneous nuclear Ribonucleoprotein

HPV: Human Papilloma Virus

HSP: Heat Shock Protein

HSPG: Heparant Sulfate Proteoglycan

HSV: Human Simplex Virus

HTLV-1 : Human T-Lymphotrophic Virus-1

IFN: Interferon

Inr: Initiator

ISE: Intron Splicing Enhancer

ISS: Intron Splicing Suppressor

JEV: Japanese Encephalitis Virus

kb: kilobase

kDa: kilodalton

lncRNA: Long Non-Coding RNA

m6A: N6 methylAdenosine

m7G: N7 mehtylguanosine

METTL3/14: Methyltransferase-Like 3/14

miRISC : microRNA Induced Silencing Complex

miRNA: MicroRNA

MNase: Micrococal Nuclease

mRNA: Messenger RNA

mRNP: messenger Ribonucleoprotein

MTase: Methyltransferase

MTE: Motif Ten Element

MVB: Multivesicular Body

MYXV: Myxoma Virus

ncRNA: Non Coding RNA

NELF: Negative Elongation Factor

NF-κB: Nuclear Factor-κ B

NGD: Non-Go Decay

NHEJ: Non Homologus End Joning

NLS: Nuclear Localization Signal

NMD: Nonsense Mediated Decay

NPC: Nuclear Pore Complex

NPC: Nuclear Pore Complex

NSD: Non-stop Decay

Nsp : Non-structural protein

nt: nucleotide

NTCP: Na+-Taurocholate Cotransporting Polypeptide

NUC: Nucleos(t)ide Analog

NUDT21: Nudix Hydrolase 21

ORF: Open Riding Frame

ORI: Origin of replication

PABP1: PolyA Binding Protein 1

PAMP: Pathogen-Associated Molecular Pattern

PAPD5/7: PolyA-RNA Polymerase Associated Domain containing protein 5/7

PAS: Polyadenylation Signal

PCR: Polymerase Chain Reaction

pgRNA: Pregenomic RNA

PHH: Primary Human Hepatocytes

PIC: Preinitiation Complex

PRC2: Polycomb Repressor Complex 2

PRR: Pattern Recognition Receptor

PRRSV: Porcine Reproductive and Respiratory Syndrome Virus

P-TEFb: Positive Transcription Elongation Factor b

PTM: Post-Translational Modification

PUF60: Poly(U) Binding Splicing Factor 60

Rab5/7: Ras Associated Protein 5/7

RAD21: double-strand-break repair Protein RAD21

rcDNA: Relaxed Circular DNA

RER: Rough Endoplasmic Reticulum

RNA: Ribonucleic Acid

RT: Reverse-Transcriptase

SAMHD1: SAM and HD Domain containing deoxynucleoside triphosphate tryphosphohydrolase 1

SARS-CoV: Severe Acute Respiratory Syndrom-coronavirus

SETDB1: SET Domain Bifurcated Histone Lysine Methyltransferase1

Sirt1: Sirtuin 1

Smc5/6: Structural Maintenance of Chromosomes 5/6

snRNP: small nuclear Ribonucleoprotein

SP1: Singly Spliced Product 1

SRA: Steroide Receptor Activator

SRSF: Serine-Arginine Splicing Factor

STAT1/2: Signal Transducer and Activator of Transcription1/2

SUV39H1: Suppressor of Variegation 3-9 Homolog 1

SUZ12: Subunit of PRC2 complex

SV40: Simian Virus 40

SVP: Subviral Particle

TAD : Topologically Associated Domain

TARDBP: TAR DNA Binding Protein

TDP2: Tyrosyl-DNA Phosphodiesterase 2

TEM: Transmission Electron Microscopy

THO: TREX complex subunit

TIP60: Histone Acetyltransferase KAT5/TIP60

TNFα: Tumor Necrosis Factor α

TP: Terminal Protein

TPase: Triphosphatase

TPH: Tupaia Primary Hepatocyte

TREX: Transcription and Export Factor

TSS: Transcription Start Site

U2AF: U2 Associated Factor

uPA: urokinase-type Plasminogen Activator

WHV: Woodchuck Hepatitis Virus

WT: Wild Type

WT: Wild-Type

XCPE1: X Core Promoter Element 1

Xrn1/2: Exoribonuclease 1/2

YY1: Yin-Yang 1

Introduction

Overview

Hepatitis B Virus

Transcriptional Events

Co-transcriptional Events

Roles of DDX5 and DDX17

Introduction

I. Overview

Hepatitis B Virus (HBV) is the causative agent of a chronic hepatitis that affects more than 250 million people in the world. Chronic carriers have increased risk of developing severe liver diseases, such as liver cirrhosis and hepatocellular carcinoma (HCC) (Trépo et al., 2014). Despite effective prophylactic vaccine and antiviral treatments, chronic hepatitis B remains a major health burden, causing more than 800 000 deaths every year (WHO, 2019). The key viral molecule for HBV persistence is its episomal replicative intermediate, also called covalently closed circular DNA (cccDNA), which stably resides in the nucleus of infected hepatocytes. cccDNA bears the complete viral genetic information and is the only template for viral replication. Hence, directly targeting cccDNA is decisive for a complete HBV cure (Nassal, 2015a).

Current treatments, mainly nucleo(s)tide analogs (NUCs), maintain under control the viral load to undetectable levels by inhibiting the viral polymerase and the subsequent new infectious particle release. However, these treatments only marginally affect the intrahepatic pool of cccDNA (Lok et al., 2017), since they act downstream of cccDNA formation and viral RNA synthesis from cccDNA. Moreover, recent data suggest an incomplete blocking of intrahepatic HBV replication by NUCs, thus probably allowing for cccDNA pool replenishment either by de novo infection of non-infected hepatocytes or by intracellular nucleocapsid recycling within already infected cells (Boyd et al., 2016). For these reasons, new strategies toward a curative therapy for HBV (Lok et al., 2017) involving direct targeting of cccDNA, either by degradation, or by transcriptional repression are highly envisaged.

At the same time, the development of non-invasive biomarkers for intrahepatic cccDNA amount and activity will have to support the evaluation of emerging therapeutic regimens. In this respect, the quantification of circulating HBV RNAs has raised much interest recently and opened new questions about the nature of these RNAs (full length, spliced, truncated…) and their biological and clinical predictive role (Charre et al., 2019).

cccDNA transcriptional activity and RNA processing is controlled by viral and cellular factors. If the essential role of the HBV X protein for full viral transcription is well characterized (L. Belloni et al., 2009; Lucifora et al., 2011), the host factors involved are still poorly identified. Therefore, investigating the implication of host proteins in the regulation of cccDNA transcription and RNA processing would help paving the way for the development of new therapeutic strategies toward a definitive cure of HBV.

II. Hepatitis B Virus

1. Background

Hepatitis B virus (HBV) is an enveloped, hepatotropic, non-cytopathic virus, belonging to the *Hepadnaviridae* family. The chronicity of HBV infection leads to major health issues (Locarnini et al., 2015). Indeed, infected patients can develop acute hepatitis, which can be resolved in 1-2 weeks, or can develop chronic hepatitis. The onset of a chronic infection strongly depends on the age at the moment of infection and is defined by the persistence of serum HBsAg for more than 6 months (Locarnini et al., 2015). Untreated chronic infections can lead to the development of liver fibrosis, cirrhosis, and, after decades, hepatocellular carcinoma (HCC), which represents the 3rd cancer-related death worldwide, causing more than 800 000 deaths each year in the world (Villanueva, 2019). Endemicity of HBV, measured by HBsAg seropositivity, is categorized as low (<2%), low-intermediate (2-4.9%), high-intermediate (5-7.9%), and high (>8%) **(Figure 1)**. The highest HBV endemicity is found in sub-Saharan Africa regions, with 8.63%, and in East Asia with 5.26% (Chang and Nguyen, 2017) .

A. Genotypes

HBV can be classified in different genotypes and sub-genotypes according to genome sequences. The classification takes into account the entire genome sequence and defines a variation between two strains when the difference is higher than 8% for the entire nucleotide sequence, and at least 4.1% in the surface genes (PreS1/PreS2/S). To date, 10 genotypes of HBV genome have been identified: A to J. Genotypes A-D and F can be further classified into sub-genotypes, with 4% to 8% nucleotide sequence differences between two sub-genotypes. Genotypes and sub-genotypes show a specific geographical distribution and are correlated to disease and clinical progressions, treatment response and prognosis (Rajoriya et al., 2017; Sunbul, 2014). Notably, genotype C infected patients are five times more prone to develop HCC than patients infected with other HBV genotypes (Yu et al., 2005). Genotype C is also associated to an increased resistance to IFN therapies (Lin and Kao, 2015).

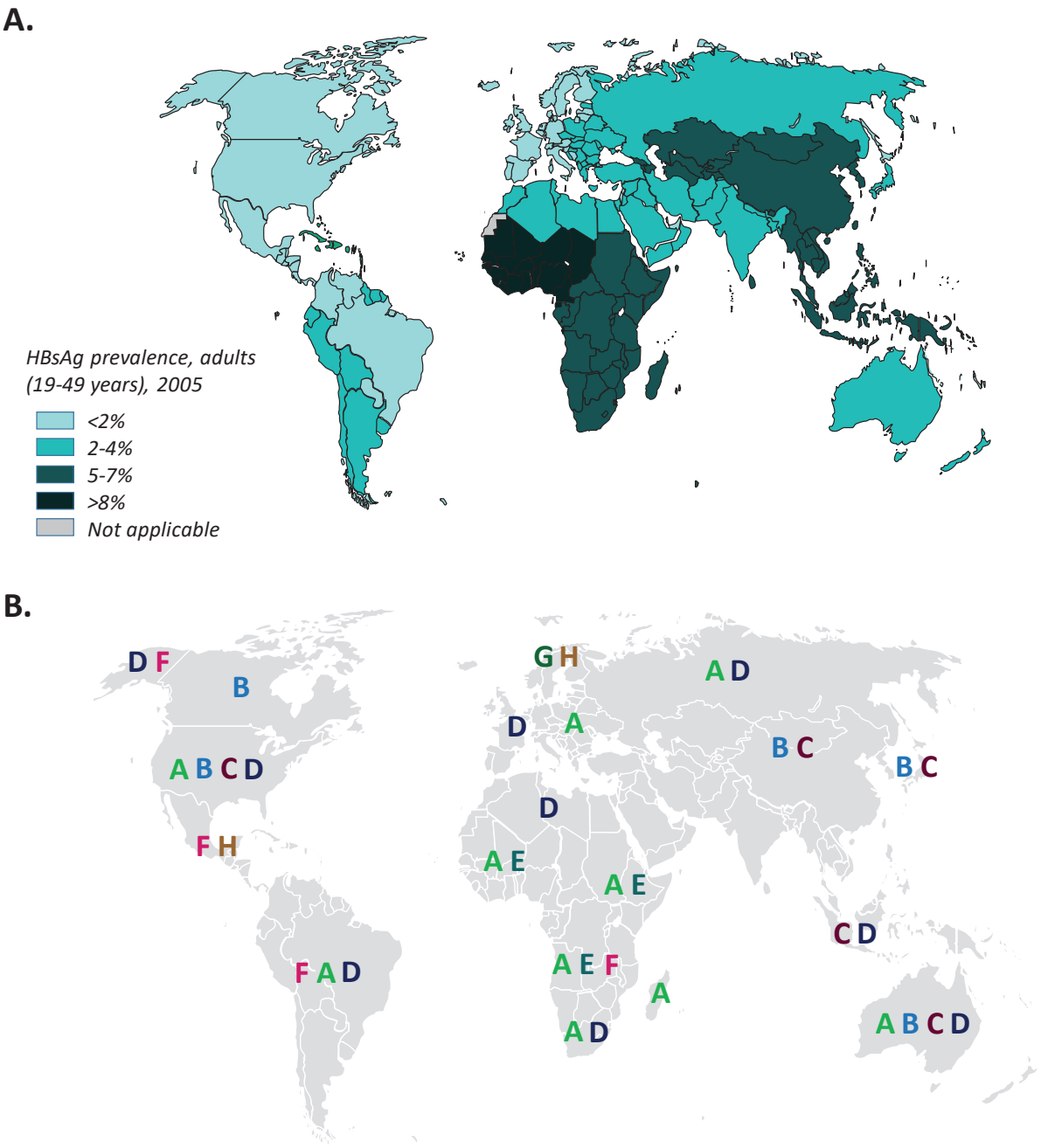


Figure 1: Seroprevalence and genotypes of Hepatitis B Virus worldwide

A. Prevalence of HBV is defined as the percentage of chronically infected people. It can be classified in 4 groups: low (<2%), low-intermediate (2-4.9%), high-intermediate (5-7.9%), and high (>8%).
-Adapted from CDC, 2016

B. Hepatitis B virus can be classified in 10 genotypes: A to J. Genotypes are defined by a strain difference higher than 8% in the entire nucleotide sequence.
-Adapted from A.Valsamakis

B. Routes of transmission

HBV can be transmitted either vertically or horizontally **(Figure 2)**.

Vertical transmission, from mother to child, is the predominant route of transmission in highly endemic countries. It is defined as HBsAg seropositivity at 6-12 months of life for an infant born from an infected mother. The risk of transmission is higher in HBsAg+ and HBeAg+ mothers (70%-90%) compared to low HBsAg+ and HBeAg- mothers (10%-40%). Mother to infant transmission can occur through 3 different ways: i) intrauterine transmission, ii) transmission during delivery, due to newborn contact with infected mother blood or fluids, and iii) postpartum transmission, caused by close contact between infected mother and infant (Borgia and Gentile, 2014).

Horizontal transmission mainly occurs through parenteral exposure to infected blood and blood products such as transfusion, contaminated needles or razors. It could also occur via unprotected sexual intercourses with exchange of body fluids. These two modes of horizontal transmission are the predominant routes of HBV infection in low endemicity countries. It is important to note that several body fluids, such as saliva, urine and faeces were evaluated positive for HBsAg presence without being a certain and direct cause of HBV transmission.

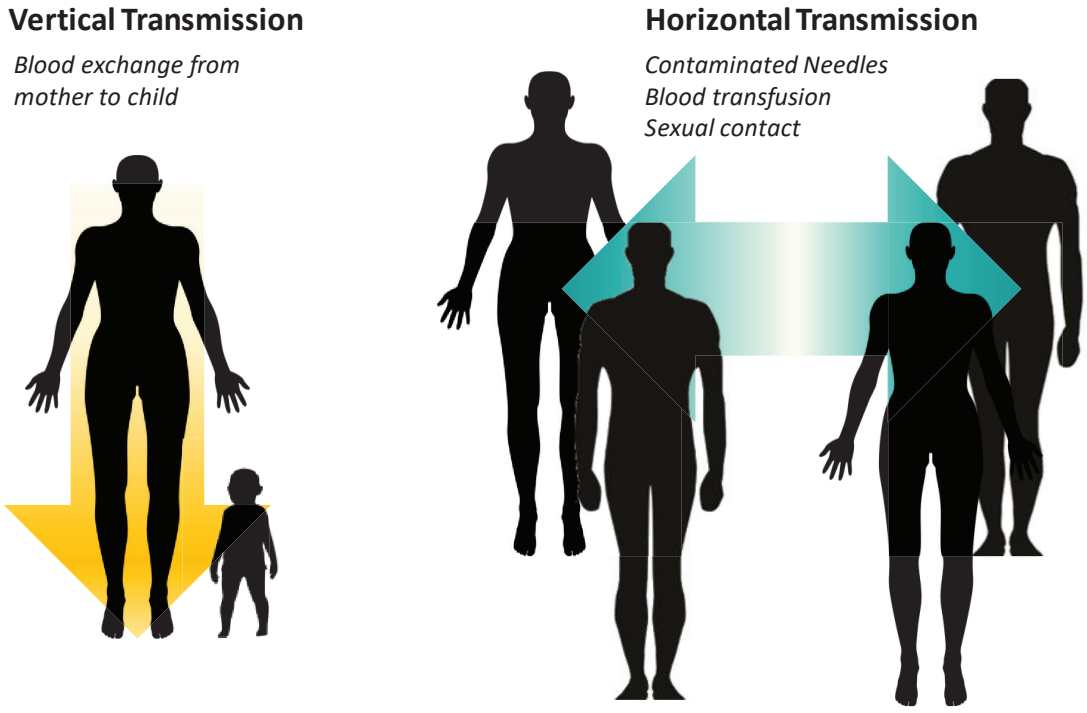


Figure 2 : Routes of transmission

HBV can be transmitted either vertically, which is the main route of transmission in endemic areas, or horizontally, mostly represented in low endemic countries.

2. HBV biology

A. Viral classification

HBV is a partially double stranded DNA virus replicating *via* an RNA intermediate that is retro-transcribed in DNA by the viral polymerase/retro-transcriptase. HBV hepatotropism, genetic organization, morphology and replication mechanisms make it belonging to the Baltimore class VII of pararetroviruses, which is composed of Hepadnaviridae and Caulimaviridae. This viral family can further be divided into two genera: i) the orthohepadnaviruses that infect mammals, such as HBV, Woodchuck Hepatitis Virus (WHV), ground squirrel hepatitis virus and ii) the avihepadnaviruses infecting birds such as Duck Hepatitis B Virus (DHBV). Each of these viruses is indigenous to its host (Schaefer, 2007).

B. Viral structure

a. *Virions and subviral particles*

HBV virus is secreted out of hepatocytes in the form of infectious particles, or sub-viral particles that are found circulating in infected patients **(Figure 3)**.

The infectious particle, also called Dane particle, is a 42 nm diameter particle, whose envelope surrounds an icosahedral nucleocapsid that contains the 3.2 kb partially double stranded relaxed circular DNA (rcDNA) genome, covalently linked to the viral polymerase. The assembly of 90-120 HBV Core protein (HBc) dimers forms the nucleocapsid. Capsids containing 90 HBc homodimers are symmetrically icosahedral with T=3 organization and measure 30nm diameter. Besides, capsids containing 120 HBc homodimers are symmetrically icosahedral with T=4 organization and are 34 nm diameter. The T=4 organization is preferentially used for envelopment. The viral envelop consists in a cellular lipid bi-layer containing the three viral envelop proteins S-, M-, L-HBsAg, in a ratio 4:1:1. In infected patients, the number of Dane particles (presumably infectious particles) can go up to 10^{10} genome-copies/mL (Patient et al., 2009).

HBV envelope proteins can also assemble without any nucleocapsid and be further secreted out of hepatocytes. These empty particles are called sub-viral particles (SVP) and are not infectious. SVP can be released as octahedral spheres of 20 nm in diameter containing only S-

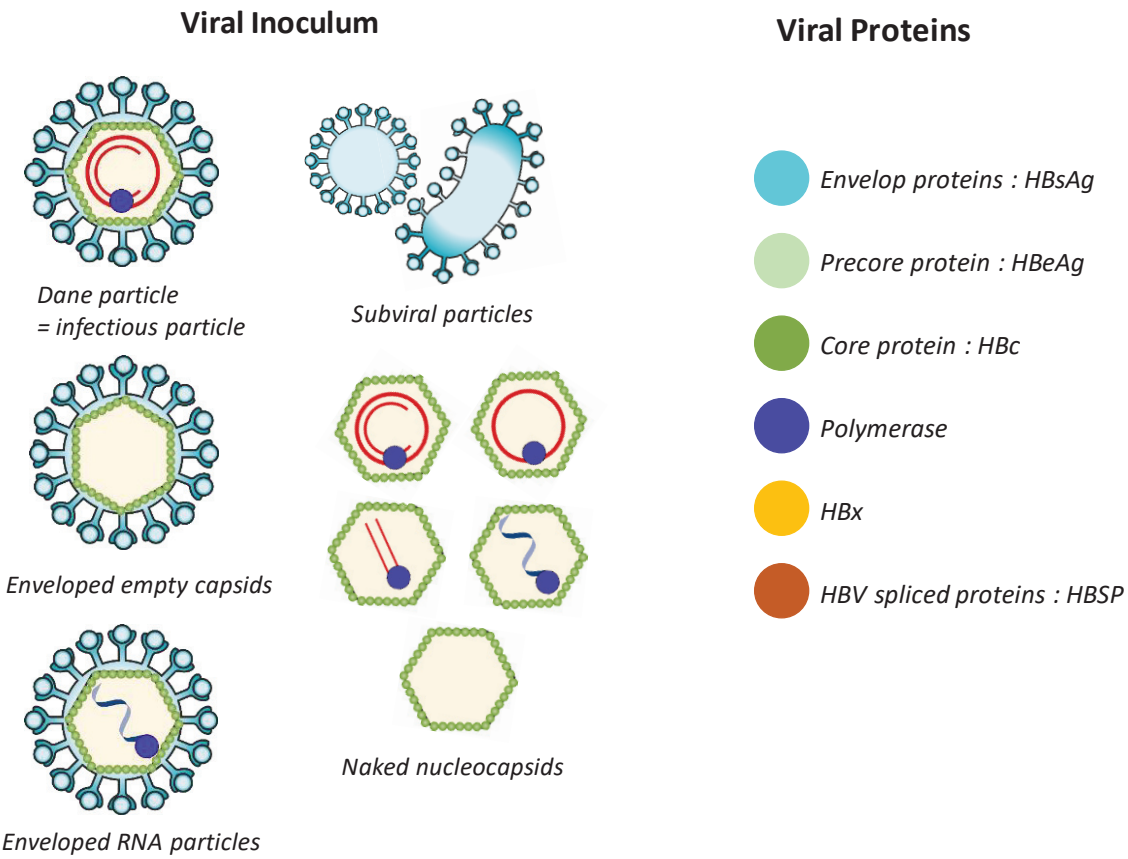


Figure 3: Viral components secreted by HBV infected hepatocytes

The viral inoculum can be composed of Dane particles, the infectious form of HBV, up to 10^9 particles/mL in patient serum. Sub-viral particles are secreted in a higher extent, and found 10^4 to 10^5-fold more frequently than Dane particles in patient serum.
Legend: the seven proteins encoded by HBV full-length mRNA, notice that HBsAg represents S-, M- and L-HBs proteins. The sixth protein is HBV spliced protein HBSP encoded by the major pgRNA-deriving spliced variant SP1.

HBsAg oligomerized in 48 homodimers, or as filaments having the same diameter and the same ratio of HBsAg proteins as Dane particles, but that can vary in length. SVP are found in 10 000 to 100 000 fold excess compared to Dane particles in blood circulation (Ganem, 2004) and are assumed to act as decoys by trapping the host's adaptive immune system (Patient et al., 2009).

In addition to HBsAg containing particles, two other viral particles have been identified in patient's sera. Immature nucleocapsids have been shown to be incompetent for envelopment and secretion (Gerelsaikhan et al., 1996). However, non-enveloped nucleocapsids have been identified *in vitro* in a hepatoma cell line transfected with the WT HBV genome (Watanabe et al., 2007). These nucleocapsids may contain the full viral genome, but also all the replicative intermediates between pgRNA encapsidation and rcDNA synthesis including pgRNA/DNA hybrids (Hu and Liu, 2017a).

Empty virions, or genome-free enveloped capsids, have also been identified in the blood of infected patients. These particles are described as "light" virions as they appear empty under electron microscopy. It is hypothesized that these empty virions are coming from damaged virions from which the rcDNA may have leaked out. Nevertheless, it still remains unclear if the secretion of such particles is part of the normal virion secretion process and what the functions of such empty virions are (Hu and Liu, 2017a).

Finally, HBV RNA species have recently been described to circulate in serum of chronically infected patients and are under evaluation for the development of non-invasive biomarkers (Charre et al., 2019). The predominant form of circulating HBV RNA seems to be encapsidated pgRNA that can be present in the serum of HBV-chronic carriers within enveloped nucleocapsids (Ning et al., 2011; Wang et al., 2016) or within naked nucleocapsids (Bai et al., 2018).

b. Viral DNA

rcDNA

HBV is the smallest DNA virus described, harboring a 3.2 kb-long genome. HBV genome is a relaxed circular DNA, also called rcDNA, and appears to be a partially double stranded DNA

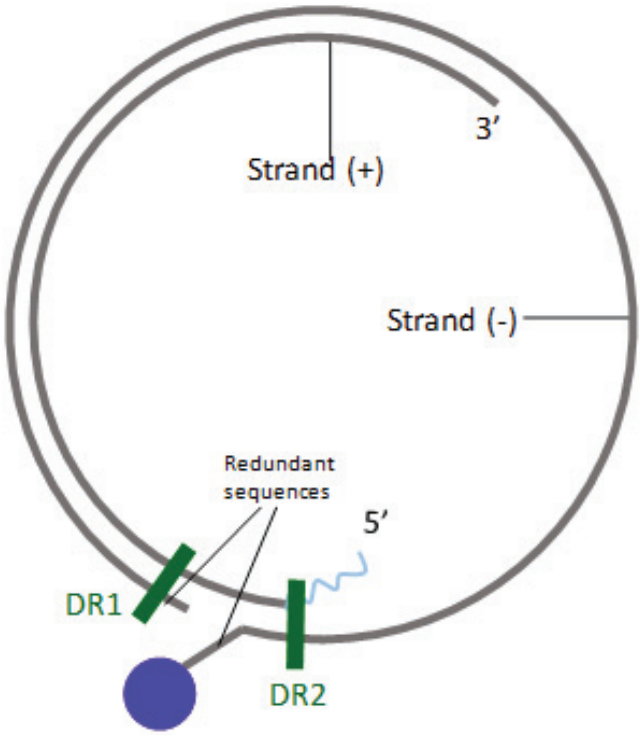


Figure 4 HBV rcDNA

rcDNA is a partially double-stranded relaxed circular DNA of 3.2 kb length. The negative strand is covalently attached to the viral polymerase via the terminal redundancy sequence. The incomplete positive strand is bound to an RNA primer in its 5'end.
The viral polymerase is represented as a blue circle, linked to the (-)-strand. RNA primer is represented as a light blue wavy line, linked to the (+)-strand.

(Erreur ! Source du renvoi introuvable.**)**. Indeed, the negative strand is complete and the positive strand has a gap that can vary between 600 and 2100 nucleotides. The complete negative strand is the template for viral transcription and is covalently linked in its 5'end to the viral polymerase (Summers et al., 1975). The 5' end of the uncomplete positive strand is associated with a capped RNA oligomer of around 18 nucleotides, deriving from the pregenomic RNA (pgRNA), and serving as a primer for the synthesis of the DNA positive strand. The negative and positive strands are held together thanks to an overlap region of around 200 bp, thus maintaining the circular shape of the rcDNA (Gao and Hu, 2007). This cohesive overlapping region contains the 11 bp direct repeat sequences DR1 and DR2 (5' - TTCACCTCTGC-3'), involved in rcDNA synthesis and found in HBV integrated forms (Bonilla Guerrero and Roberts, 2005). Indeed, the RNA primer associated to the 5' end of the negative strand has to be translocated onto the 3'end of the positive strand, and annealed to the DR2 sequence by sequence complementarity. Together with hairpin formation at DR1 and sequence identity between DR1 and DR2, these phenomena are mandatory for an efficient retrotranscription of pgRNA (Habig and Loeb, 2006) *(See below "pgRNA transcription, encapsidation and retro-transcription")*.

dsL-DNA

During rcDNA synthesis, RNA primer translocation can be altered, resulting in an *in situ* priming, from DR1 (Habig and Loeb, 2006). This process occurs within 10% of mature nucleocapsids and leads to the formation of a linear double-stranded DNA (dsL-DNA). dsL-DNA can then i) form defective cccDNA, unable to support rcDNA synthesis due to insertion(s) or deletion(s) appearing during the ligation of dsL-DNA extremities; or ii) be integrated into the host cell genome. HBV genome integration can support HBsAg and HBx synthesis, but no other viral protein production. Indeed, while remaining intact, Pol and HBeAg ORFs are physically separated from their promoters. Hence, integrated genome cannot support pgRNA transcription, then leading to non-replicative viruses. However, dsL-DNA integration was assumed to be a causative process in tumorigenesis and associated to HCC initiation and progression. It is important to note that the mechanisms of HBV-induced HCC carcinogenesis are still poorly understood as they involve many other mechanisms besides viral integration (Tu et al., 2017).

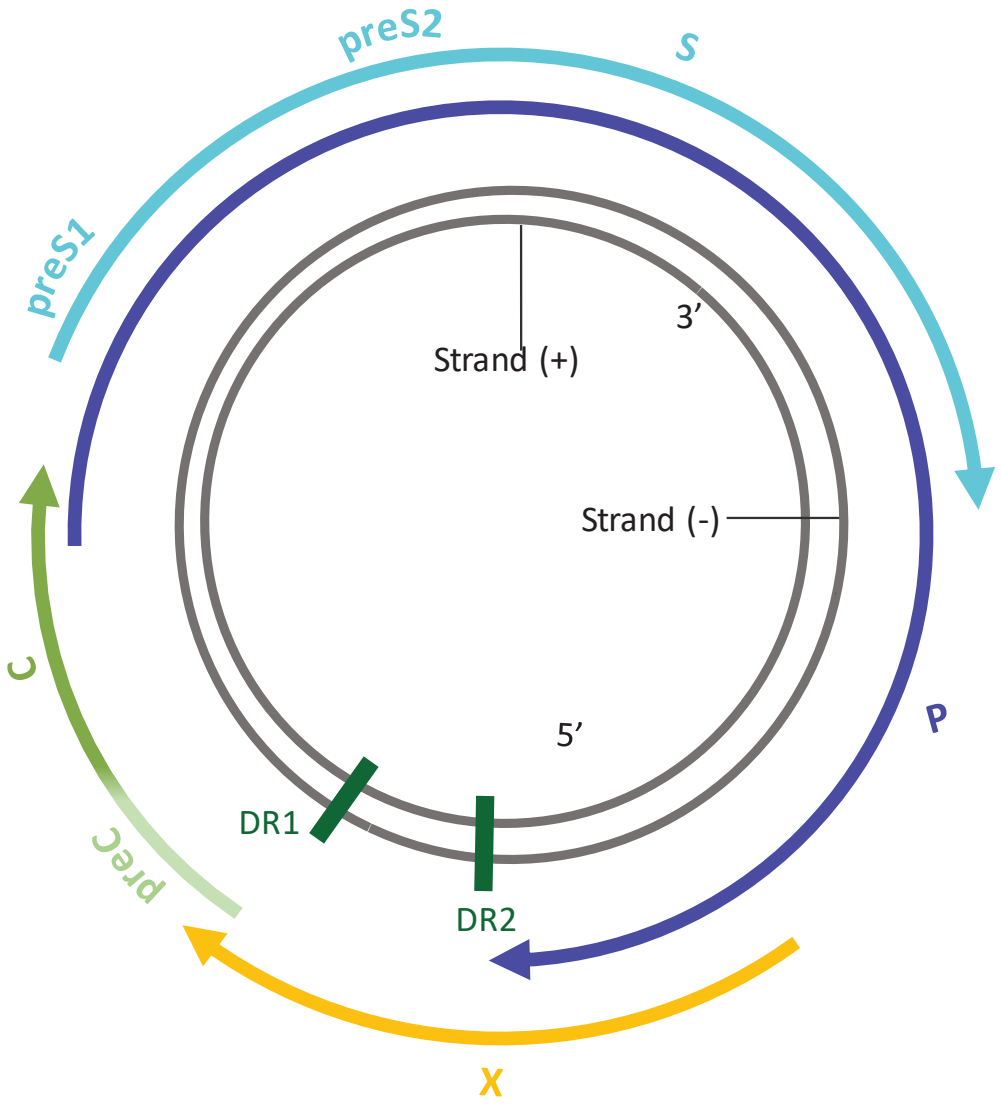


Figure 5 : cccDNA and ORFs

HBV covalently closed circular DNA (cccDNA) is the transcriptionally active DNA intermediate of HBV life cycle, responsible for HBV persistence. It contains 4 overlapping reading frames (ORF), namely preCore/Core ORF, for the synthesis of HBeAg and HBc; S ORF, for L-, M-, and S-HBsAg; Pol ORF, for the viral polymerase/retro-transcriptase, and X ORF, for HBx.

cccDNA

Into the nucleus of infected hepatocytes, rcDNA is converted into a covalently closed circular DNA, also called cccDNA. This cccDNA is the persistent episomal form of HBV genome, and is the unique template for all the viral RNAs, including the pregenomic RNA, or pgRNA, which is the RNA intermediate of the viral replication. The viral transcription is regulated by 4 promoters, 2 enhancers, and a common polyadenylation signal. cccDNA contains four overlapping open reading frames (ORFs) with every base coding for at least one ORF (Nassal, 2015a) **(**Erreur ! Source du renvoi introuvable.**)**. The viral ORF are oriented in the same direction and are the following:

- Pre-core/Core ORF: 2 in-phase start codons, encoding HBc and HBeAg.
- Polymerase ORF: the largest ORF (80% of the viral genome), encoding the viral polymerase.
- PreS1/PreS2/S ORF: 3 in-phase start codons, within the polymerase ORF, encoding the 3 different envelop proteins (S-, M-, L-HBsAg; for small, medium and large HBsAg.)
- HBx ORF: encoding the viral transactivator HBx.

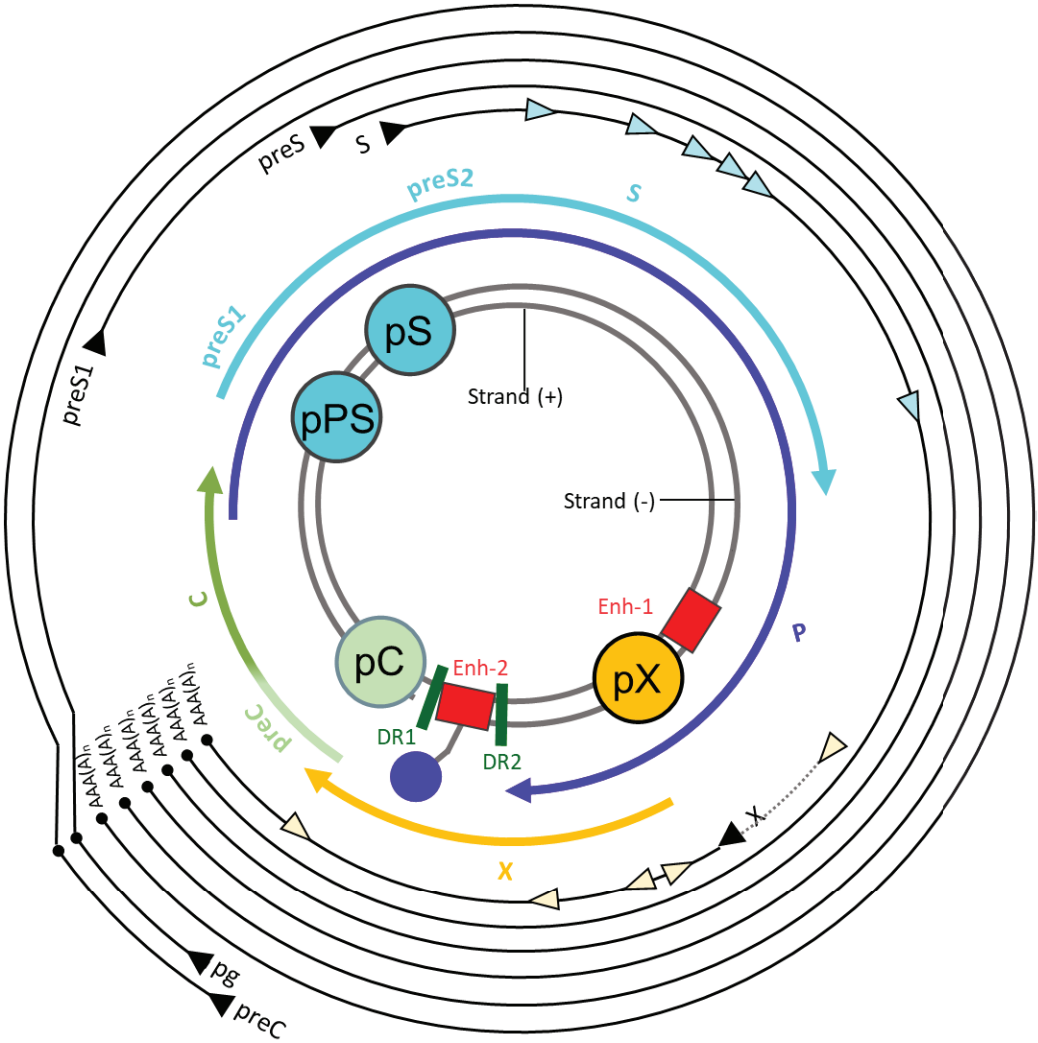


Figure 6 : Viral transcripts and regulatory elements

HBV genome encodes 6 viral mRNAs: the 2 longer-than-genome length 3.5 kb precore and pregenomic RNAs, the 2.4 kb preS1 RNA, the 2.1 kb preS2 RNA, the 2.1 kb S RNA and the 0.7 kb X RNA. Several spliced isoforms have been identified. Major viral transcripts start at distinct TSS (black arrow) and end at a common polyA site (black dot – AAAA(A)n). Weaker TSS identified by Altinel et al. 2016, are represented in light blue triangle for S transcripts and light yellow triangle for X.
Viral transcription is regulated by 4 promoters: promoter C (pC), promoter preS (pPS), promoter S (pS) and promoter X (pX); and by 2 enhancers: Enh-I and Enh-II.

c. RNA transcripts

HBV genome encodes 6 viral RNAs, having their own Transcription Start Sites (TSS), but sharing a common polyadenylation signal (PAS) **(Figure 6)**. Every RNA harbors a cap at its 5' extremity. The cellular RNA Polymerase II is the enzyme catalyzing the viral transcription (Rall et al., 1983).

The 3.5 kb RNAs:

- The pregenomic RNA (pgRNA) is synthesized from the Pre-core/Core promoter and is 3.5 kb long (i.e. 1.1-fold longer than the viral genome). It contains the whole genome information and is the RNA intermediate necessary to achieve the complete viral replication, through reverse-transcription onto rcDNA. The pgRNA is also an mRNA as it encodes the viral polymerase and the capsid protein HBc. The polymerase ORF is less efficiently used than the Pre-core/Core ORF, with one polymerase protein translated for 240 HBc (Sen et al., 2004).
- The Pre-core messenger RNA (mRNA) is also synthesized from the Pre-core/Core promoter and is 3.5 kb long. This RNA encodes the secreted HBeAg.

pgRNA and pre-Core mRNA are only 30 bp different in length, pre-core mRNA being longer, and come from two physically distinct promoters, while generally assumed as the "Pre-core/Core" promoter (Yu and Mertz, 1996).

The 2.4 kb RNA:

The PreS1 mRNA is transcribed from the PreS1 promoter and encodes the Large HBsAg.

The 2.1 kb RNAs:

- The PreS2 mRNA is 2.1 kb long and is synthesized from the PreS2/S promoter. It encodes the Medium HBsAg.
- The S mRNA is also 2.1 kb long, and is also regulated by the PreS2/S promoter. It encodes the Small HBsAg.

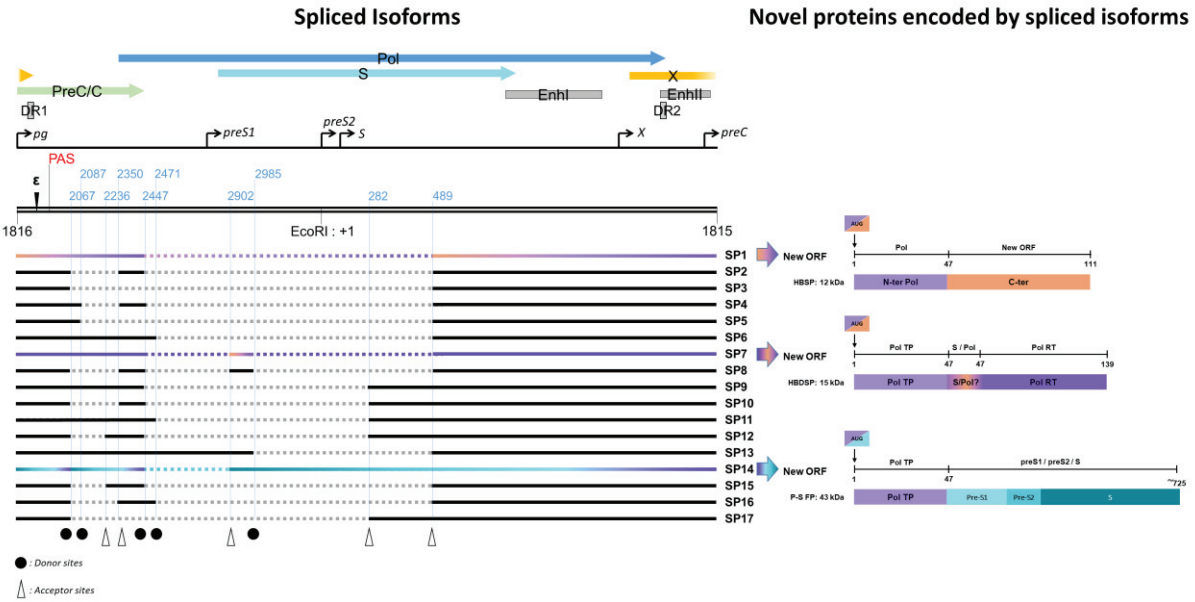


Figure 7 : HBV spliced isoforms and encoded proteins

pgRNA can be spliced in 17 isoforms according to the use of major donor (black dots) and acceptor (white triangles) splicing sites (Chen et al., 2015). Viral ORFs are annotated on top of the scheme (preC/C, Pol, S and X ORFs). The major TSS are annotated with black arrows (pg, preS1, preS2, S, X and preC). TSS are positioned according to Altinel et al, 2016 & Stadelmayer et al, 2020.

pgRNA-deriving spliced variants are named from SP1 to SP17. To date, only three of these spliced variants are known to encode viral proteins.

SP1 encodes the 12kDa Hepatitis B Spliced Protein (HBSP), consisting of the fusion of Pol ORF at N-ter and of a new ORF at C-ter.

SP7 encodes the 15 kDa Hepatits B Doubly Spliced Protein (HBDSP), which ORF is not clearly defined but is assumed to be the fusion of different subunits of Pol and potentially S ORFs.

SP14 encodes the 43 kDa Pol-S Fusion Protein (P-S FP), corresponding to the fusion of the extreme N-ter of Pol ORF and the PreS1 ORF at its C-ter.

PreS2 TSS, starting at position 3182 in D genotype, is only 26 bp upstream S TSS, starting at position 25 in the same genotype, thus giving rise to two independent mRNAs (Altinel et al., 2016).

The 0.7 Kb RNA:

The X mRNA is the smallest HBV RNA with a 0.7 kb length. It encodes the HBx protein (Moolla et al., 2002).

It is important to note that 17 TSS have been recently identified. Among them, 6 are the already well-described major TSS initiating the 6 major HBV transcripts. Eleven weaker TSS have also been identified along HBV genome, notably in the Pre-core/Core promoter and in the X transcript. Alternative TSS in the HBx sequence could give rise to a shorter HBx RNA (Altinel et al., 2016). Noteworthy, technic used by Altinel and colleagues allows the identification of approximately 50 nt in 5' from the cap structure. Therefore, the length of the potential new transcripts remains to be studied. Moreover,.functions of these transcripts have not been elucidated so far and further investigations are required to fully characterize the products of these eleven "weak TSS".

Spliced variants

Besides the abovementioned HBV transcripts, several spliced variants have been identified *in vivo*, in patient samples (sera and biopsies) and *in vitro*, in cell culture. More precisely, sixteen spliced variants have been discovered, deriving from the pgRNA and four spliced variants have been identified deriving from the PreS2/S mRNA (Candotti and Allain, 2016a). A seventeenth variant, called SP17 has been recently identified in sera of CHB patients (Chen et al., 2015) **(Figure 7)**. These spliced variants are resulting from the activation of cryptic splicing sites, or from the repression of constitutive splicing sites, by *cis*-acting cellular and/or viral factors. All these variants share the splicing features generally found in the cellular RNAs, with a donor 5' splicing site (5'ss), an Adenine located within the exon to skip, and an acceptor 3' splicing site (3'ss), upstream controlled by a pyrimidine tract (Lee and Rio, 2015). Splicing of HBV (genotype D) pgRNA varies in donor 5'ss but remains similar for all the variants in acceptor 3'ss. Contrarily, the PreS2/S deriving spliced isoforms share a common donor 5'ss and vary on their acceptor 3'ss. Moreover, spliced isoforms are dependent of HBV genotypes. So far, only single and double spliced isoforms have been identified (Candotti and Allain, 2016a). The most

reported pgRNA deriving spliced isoform is the so called Singly spliced Product 1, or SP1, which is 2.2 kb long, and spliced between nucleotides 2447 and 489, according to *ayw* genome (genotype D) nomenclature (Su et al., 1989). This spliced variant appears to be found in around 30% of pgRNA transcripts in transfected cell culture, and up to 60% in patient samples (Wu, n.d.). pgRNA deriving spliced isoforms can be encapsidated and retro-transcribed onto defective HBV DNA particles. Indeed, such abnormal particles circulate in the sera of infected patients, and are correlated with liver diseases (Soussan et al., 2008a). Moreover, the amount of HBV spliced isorforms is increasing in patient sera prior HCC development (Bayliss et al., 2013a).

SP1, together with two other spliced variants, namely SP7 and SP14, have been demonstrated to encode spliced proteins, called HBSP, HBDSP and P-S FP respectively (*see below "HBSP" and "Discussion"*) **(Figure 7)**.

d. Viral proteins

Surface proteins: L-, M-, S-HBsAg

The viral RNAs PreS1 (2.4 kb), PreS2 and S (2.1kb) encode the three proteins constituting the HBV envelope. PreS1 mRNA encodes the Large surface protein, or L-HBsAg, from the PreS1 ORF, which is 42 kDa, while PreS2 and S mRNAs encode the Medium and Small surface proteins, respectively M-HBsAg of 31 kDa, and S-HBsAg, of 27 kDa, from the same PreS2/S ORF. Two specific translation start codons are used from the same PreS2/S ORF, however, mechanisms of leaky scanning lead to the differential expression of the M- and S-HBsAg in the ratio 1:4 (Sheu and Lo, 1992). L-, M- and S-HBsAg share the same carboxyterminal domain (C-Ter), corresponding to the whole S-HBsAg. L- and M-HBsAg share the PreS2 domain, while L-HBsAg is the only envelope protein containing the PreS1 domain **(Figure 8).**

HBsAg are synthesized through the Rough Endoplasmic Reticulum (RER) and then matured into the Golgi Apparatus. Envelope proteins contain four transmembrane domains. The first transmembrane domain of PreS1 can be either translocated onto the viral membrane (e-preS topology), or can remain cytosolic (i-preS topology) (Prange, 2012). The common S domain contains the antigenic loop (AGL) region, which is N-glycosylated on its residue 146, in a ratio glycosylated/non-glycosylated 1:1. The non-glycosylated N146 is necessary for HBV infectivity, while the glycosylated form of N146 is essential for virion secretion (Yan et al., 2012a). The

AGL also bears the so-called immunodominant a-determinant, the first HBV marker identified and conserved in all HBV strains (Julithe et al., 2014). The a-determinant is a conformational epitope that can attach Heparan Sulfate (HS) at the surface of hepatocytes (Le Duff et al., 2009). This close contact to the hepatocyte allows the myristoylated PreS1 domain of L-HBsAg, in its e-preS conformation, to interact with the sodium taurocholate cotransporting polypeptide (NTCP) receptor (Yan et al., 2012a). Moreover, the PreS1 domain, in its i-preS topology, interacts with the preformed cytosolic nucleocapsid. This interaction is mostly due to 17 C-ter residues stretch in PreS1. Another region was predicted to interact with the nucleocapsid: the first cytosolic loop of the S domain, but this hypothesis is not confirmed so far (Prange, 2012). L-HBsAg and S-HBsAg can also be non- or mono-glycosylated in the viral envelope, while M-HBsAg can be mono- or di-glycosylated. In certain HBV strains, M-HBsAg can also be O-glycosylated (Julithe et al., 2014). It is important to note that M-HBsAg is not essential for HBV assembly and infectivity, but necessary for HBV virion secretion (Bruss and Ganem, 1991; Prange, 2012).

Mutated or truncated HBV surface proteins can also be synthesized from integrated HBV genome. These mutant forms are associated to ER stress response that may increase the risk of HCC. Furthermore, mutated S proteins could confer proliferative advantages to hepatocytes, for example by stimulating their expansion. Finally, in HCC animal models, over-expression of HBs mutants drives precancerous liver damages (Tu et al., 2017).

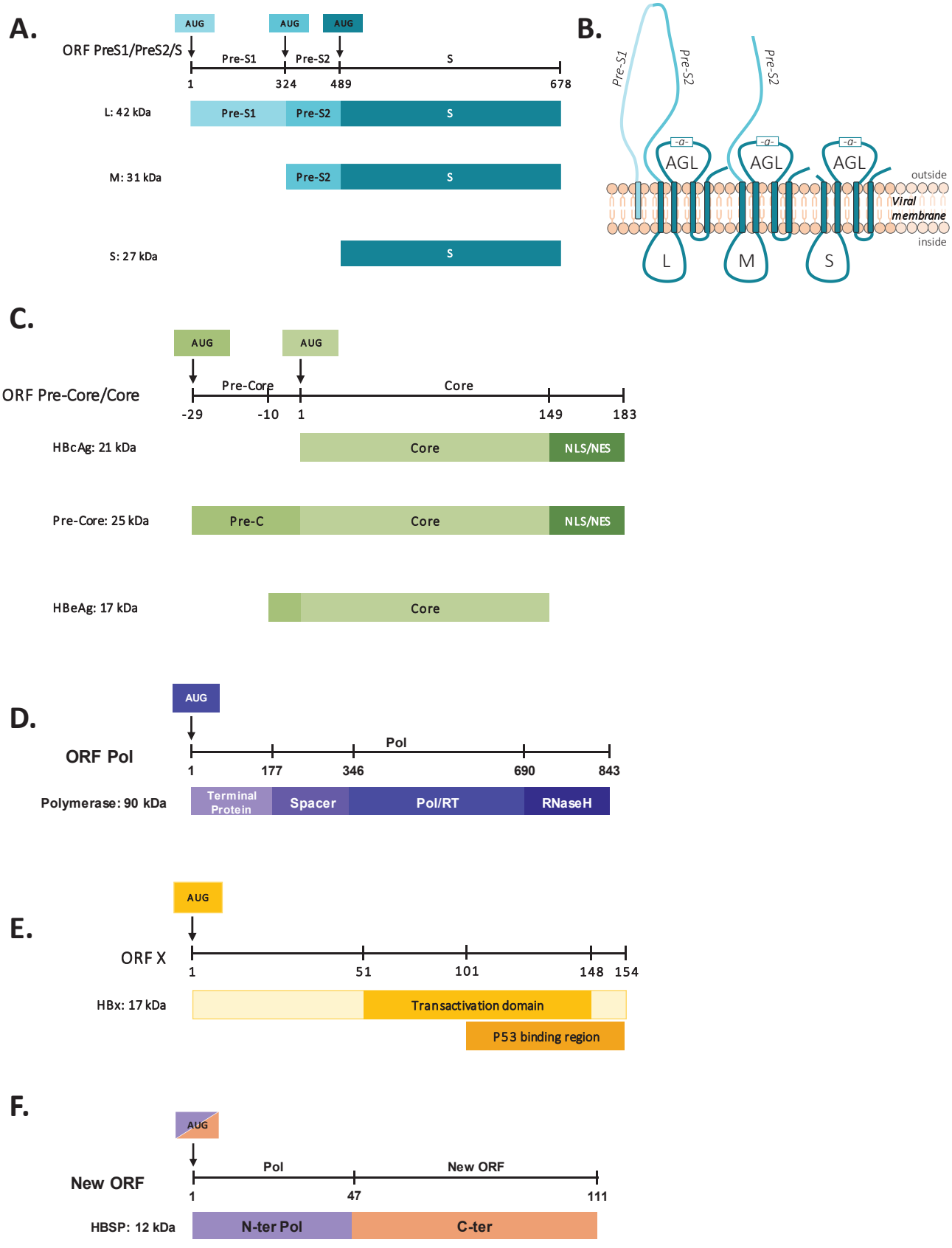

A.
ORF PreS1/PreS2/S
AUG
AUG
AUG
Pre-S1
Pre-S2
S
1
324
489
678
L: 42 kDa
Pre-S1
Pre-S2
S
M: 31 kDa
Pre-S2
S
S: 27 kDa
S
B.
Pre-S1
Pre-S2
Pre-S2
AGL
AGL
AGL
outside
Viral membrane
inside
L
M
S
C.
ORF Pre-Core/Core
AUG
AUG
Pre-Core
Core
-29
-10
1
149
183
HBcAg: 21 kDa
Core
NLS/NES
Pre-Core: 25 kDa
Pre-C
Core
NLS/NES
HBeAg: 17 kDa
Core
D.
ORF Pol
AUG
Pol
1
177
346
690
843
Polymerase: 90 kDa
Terminal Protein
Spacer
Pol/RT
RNaseH
E.
ORF X
AUG
1
51
101
148
154
HBx: 17 kDa
Transactivation domain
P53 binding region
F.
New ORF
AUG
Pol
New ORF
1
47
111
HBSP: 12 kDa
N-ter Pol
C-ter

Figure 8 : Viral proteins

A. Large-, Medium- and Small-S proteins constitute the viral envelope. These 3 proteins share the "S" domain, M- and L- also share the "Pre-S2" domain. Only L- contains the "Pre-S1" domain.
B. Envelope proteins contain 4 transmembrane domains, anchored in the host cell lipid bilayer. The "S" domain bears the so-called antigen-loop (AGL), which contains the "a-determinant".
C. HBcAg, Pre-core and HBeAg share the "core" domain, while only HBcAg and Pre-core proteins contain the NLS sequence. Pre-core is maturated through "Pre-core" and NLS domain cleavage to form the secreted HBeAg.
D. The viral polymerase is composed of a "TP" domain that mediates the covalent interaction with rcDNA, a "Spacer" domain, a "Pol/RT" domain in charge of the (-)- and the (+)-DNA synthesis during pgRNA retro-transcription, and a "RNaseH" domain that degrades pgRNA during the (-)-DNA strand synthesis.
E. HBx is a non-structural viral transactivator. HBx is mandatory for the viral transcription and is implicated in HCC development.
-Adapted from Tan, 2005 and from Elmore (1997)
F. HBSP is translated from the major form of pgRNA-derived spliced isoform SP1. This chimeric protein share its N-ter with the viral polymerase/retro-transcriptase, and has its own C-ter encoded by a newly generated ORF originating from the splicing process.
-Adapted from Assrir et al. 2010

Capsid Protein: HBc

The capsid protein, also called core protein, or HBc, is 183 to 185 aa long, depending on the HBV genotype. This 21 kDa protein can assemble in dimers and in capsid shell, *via* the "core" domain (residues 1 - 140). The ability of HBc of establishing dimers is due to its secondary structure in α-helices. The dimerization consists in the union 4 α-helices forming spikes on the surface of the particle (Prange, 2012) **(Figure 8).** On the other hand, the oligomerization of HBc leads to nucleocapsid formation, composed of 180 or 240 HBc units, with an icosahedral symmetry in T=3, measuring 30 nm in diameter, or T=4 organization, measuring 34 nm in diameter, respectively. Both of these capsids can be present into infected hepatocytes, but the T=4 geometry is preferentially incorporated in viral particles (Patient et al., 2009).

A 10 aa linker separates the "core" domain from the "protamine" domain (residues 150 - 183/185). The latter, also called CTD for C-ter domain, is arginine-rich and drives the interaction with nucleic acids (Nassal, 1992). Hence, while it is dispensable for assembly processes, it is essential for pgRNA packaging and reverse-transcription. The protamine domain contains nuclear localization signal (NLS) and nuclear export signal (NES), required for the shuttling of HBc between the nucleus and the cytoplasm. Capsid entry into the nucleus and subsequent HBV rcDNA release is required for the establishment and the maintenance of infection through the persistence of the cccDNA pool. This process is mediated by nuclear

transporters such as the family of importin-β, in concert with the nucleporin 153 (Schmitz et al., 2010). Moreover, the CTD can be phosphorylated in three major serine residues (S157, S164 and S172) by different cellular kinases, thereby inducing conformational changes by neutralizing the positive charges of the CTD and thus loosening the contact with nucleic acids. These phosphorylations correlate with pgRNA packaging and (+)-ssDNA synthesis. Contrarily, the dephosphorylation is associated with capsid maturation, DNA replication and subsequent nucleocapsid envelopment and virion release (Lan et al., 1999).

Pre-core Protein: HBeAg

HBeAg is a soluble and non-particulate HBV antigen circulating in the serum of infected patients (Nassal et al., 1992). HBeAg share sequence with HBc, but is encoded by the Pre-Core mRNA **(Figure 8)**. With the exception of a 29 residues extension in its N-ter, HBeAg is similar to HBc and contains the same domains. Its higher length is due to the location of its start codon, 87 nucleotides upstream of HBc start codon. Its N-ter hydrophobic sequence of 19 residues directs the 25 kDa HBeAg precursor to the RER where it undergoes two successive cleavages to be fully mature. The first cleavage leads to a 22 kDa protein, which can either be cleaved a second time in the Golgi network in its C-ter to form the secreted form of HBe of 17 kDa (Messageot et al., 2003), or that can traffic to the cytosol, forming the p22 viral protein (Dandri and Locarnini, 2012).

The secreted HBeAg is not required for the viral replication but exhibits immune-modulating roles that contribute to the viral persistence (Dandri and Locarnini, 2012). The presence of circulating HBeAg may be one of the determinants involved in thenimmune tolerance state in infected patients. Indeed, HBeAg-negative mutants have been associated to fulminant hepatitis B and more severe chronic hepatitis B. By mimicking HBc antigen, that can be recognized by the immune system at the surface of hepatocytes and induce a T-cell mediated immune response, HBeAg saturates the immune receptors and allows the virus to escape the immune response (Lee, 1997).

Polymerase

The viral polymerase, designated as Pol, is 845 aa long and is the only enzymatic viral protein encoded by HBV genome. This multifunctional 90 kDa protein is translated from the pregenomic RNA and contains four different domains; from the N-terminal : i) the terminal

protein (TP) domain; ii) the spacer domain; iii) the reverse-transcriptase (RT) domain; and iv) the RNase H domain **(Figure 8)**.

The TP domain, from aa 1 to aa 180, connects the viral polymerase to the 5' end of the DNA minus-strand through its residue Y63, thus leading to a primase activity that allows the initiation of pgRNA reverse transcription by synthesizing a 3 or 4 nucleotide primer (Lanford et al., 1997; Zoulim and Seeger, 1994). Moreover, the TP domain is also required for the pgRNA packaging by interacting with the pgRNA 5' "epsilon" ε-loop.

The spacer domain, from aa 181 to aa 359, tethers TP and RT domains and can tolerate amino acid insertion or deletion. The role of this genetically variable region is still unclear and appears to overlap with the PreS domain of the S protein, thus suggesting a role in environmental adaptation and providing flexibility in conformational changes (Chen et al., 2013).

The RT domain, from aa 360 to aa 693, harbors a polymerase/retro-transcriptase activity, responsible for pgRNA retro-transcription into negative strand DNA with subsequent DNA positive strand synthesis.

The C-terminal domain, from aa 694 to aa 845, carries a RNase H activity driving the pgRNA degradation after retro-transcription of the DNA minus strand, and is also important for generating the short RNA primer, required for the DNA plus strand synthesis (Wei and Peterson, 1996). Blocking the RNase H activity leads to an accumulation of RNA:DNA hybrids, and inhibits the plus strand DNA synthesis, making the RNase H domain of HBV polymerase a potential therapeutic target (Tavis and Lomonosova, 2015).

X Protein: HBx

HBx is a 154 aa-long protein (17 kDa) encoded by the smallest ORF of HBV genome, and designated as X as it does not share any homology with any other known cellular or viral protein. HBx ORF is present in mammalian hepadnaviruses, but absent in avian viruses. This non-structural protein has been studied for long time, and the different roles described for HBx vary according to the model studied, and some demonstrations remain controversial (Benhenda et al., 2009). However, despite still unclear functions, two main roles have been proposed for HBx protein: i) a role in viral replication, ii) a role in hepatocellular carcinogenesis.

First, in 1994, Zoulim et al. demonstrated for the first time that X protein plays a critical role in Woochuck Hepatitis Virus (WHV) life cycle. Indeed, in mutants expressing truncated C-ter of X protein or in mutants preventing the expression of X protein, by X- AUG mutation, amount of virion DNA was decreased by 60% compared to wild-type HBV genome transfection and the establishment of viral infection failed *in vivo* (Zoulim et al., 1994.). The essential role for maintaining the viral infection was further demonstrated for HBV infection, where the transcomplementation with WT HBx protein restored HBV replication levels (Keasler et al., 2007). HBx ability of modulating HBV transcription and replication is dependent on its C-terminal domain, from aa 51 to aa 148, designated as "transactivation domain" (Tang et al., 2005) **(Figure 8)**. Later, it was determined that HBx-negative HBV impaired the transcriptional level of cccDNA (Lucifora et al., 2011a). WT and HBx mutant viruses were both able to form the same levels of cccDNA, while HBx mutant virus showed no viral RNA transcription in an infectious model. These observations highlighted the importance of HBx for cccDNA transcriptional activity, but not for the establishment of the infection (Lucifora et al., 2011a).

Another highlighted role of HBx is its potential implication in carcinogenesis. Moreover, albeit HBV DNA integration occurs with a terminal deletion of around 200 bp into the host genome, enhancer-1 sequence activating HBx transcription remains intact. In the context of integration, HBx RNA can be translated onto a 3-aa truncated protein, which is still functional. The expression of this C-ter truncated HBx has been associated with stem-like properties, cell transformation and apoptosis inhibition, and appears to be important in HCC tumorigenesis (Tu et al., 2017).

Several studies uncovered the role of HBx as transcriptional activator through a direct role on HBV cccDNA. Indeed, HBx was demonstrated to bind cccDNA and to regulate its epigenetic state. HBx mutant virus harbored a decreased cccDNA association of transcriptional activators, such as the acetyltransferase p300, and contrarily, an increased binding of repressing factors, such as histone deacetylases HDAC1 and Sirt1. These differential interactions were associated to a reduction of pgRNA transcription and viral replication (Laura Belloni et al., 2009). HBx protein was also shown to interact with many different cellular pathways, such as cell cycle, apoptosis, or DNA repair and with several cellular partners as members of the proteasome, p53 and the DNA damage-binding protein DDB1. In particular, the direct binding of HBx to DDB1 is important to redirect the ubiquitin ligase activity of the CUL4-DDB1 E3 ligase (Li et al.,

2010, p. 1). HBx hijacking of CUL4-DDB1 E3 ligase was shown, more recently, to drive Smc5/6 complex degradation, implicated in HBV replication restriction by blocking cccDNA transcriptional activity (Decorsière et al., 2016). Important finding from this study highlighted for the first time a mechanism by which HBx confers active transcription of cccDNA.

The role of HBx as a transcriptional modulator of cellular genes was also assessed by recent studies. HBx ChIP-seq analyses revealed a repertoire of new genes and ncRNAs, targeted by HBx, and involved in the positive regulation of cccDNA transcriptional activity and viral replication. Indeed, HBx activates genes and miRNA involved in endocytosis and autophagy, and thus favoring HBV replication. Contrariwise, HBx can also repress certain miRNA, such as miR-224 and miR-138, implicated in HBV replication inhibition and pgRNA targeting (Guerrieri et al., 2017).

By interacting with several cellular proteins, HBx can overcome senescence by inhibiting p53 nucleotide excision repair, through its p53 binding region, from aa 101 to aa 154 (Elmore et al., 1997), or by down-regulating the expression of p53 activators ASPP1 and ASPP2 (Levrero and Zucman-Rossi, 2016). HBx also influences epigenetic modification levels of target genes that could be implicated in HCC development. For example, HBx induces DNA hypermethylation by upregulating DNA methyltransferases (DNMT), thus leading to hypermethylated genes implicated in cell cycle regulation, cell growth, dedifferentiation, invasiveness and other pro-oncogenic processes. Inversely, HBx can promote gene-specific DNA hypomethylation through different cellular partners. HBx can also induce aberrant histone post-transcriptional modifications or alterations in miRNA (Tian et al., 2013).

Hepatitis B Splice-generated Protein: HBSP

All HBV proteins described above are encoded by unspliced HBV RNAs. However, spliced variants were identified in patients' samples (Candotti and Allain, 2016a) and a novel HBV protein, encoded by the singly spliced variant SP1 of 2.2 kb, was detected *in vivo* (Soussan et al., 2000). Hepatitis B splice-generated Protein, or HBSP, is a 12 kDa protein that has been found in HBV-infected livers. This protein corresponds to the fusion of 47 amino acids of the HBV polymerase N-terminal domain to a 64 amino-acid polypeptide synthesized from a new ORF generated after the splicing event **(Figure 8)**. Moreover, circulating anti-HBSP antibodies have been found in the serum of patients with detectable viremia. However, no correlation was established between anti-HBSP antibodies and HBV viral load in chronic carriers. *In vitro*,

the overexpression of HBSP did not show any effect on HBV DNA replication and transcription (Soussan et al., 2000). HBSP expression was first associated to severe liver fibrosis score, with no association with liver necroinflammation or other histological activities. Furthermore, an increased serum level of TNF-α was observed in anti-HBSP positive patients (Soussan et al., 2003a). Further studies investigated the correlation between HBSP expression and TNF-α signaling pathway, and demonstrated that HBSP reduced hepatic immune infiltration in HBSP transgenic mice injured liver with successive LPS injections. This mechanism was related to a down-regulation of immune response, notably through NF-κB activity weakening when HBSP-expressing cells were stimulated with TNF-α (Pol et al., 2015). Finally, the immune-modulatory role of HBSP has been evidenced through the down-regulation of C-C motif chemokine ligand 2, CCL2. Duriez and colleagues demonstrated that spliced variants were upregulated upon liver injury, leading to an increased HBSP synthesis. HBSP expression impairs NF-κB expression by changing the phosphorylation status of its inhibitor IκB, resulting in the down-regulation of the NF-κB-induced CCL2 transcription and in the subsequent reduction of CCL2 secretion. Consequently, inflammatory monocytes/macrophages recruitment during liver injury declines, leading to a protective effect against liver damage, then supporting the hypothesis of a HBV-driven feedback loop (Duriez et al., 2017). However, these studies were conducted in HBSP transgenic mice upon artificial TNF-α stimulation. More physiological studies remain necessary to fully elucidate the role of HBSP in HBV associated pathologies.

3. Viral life cycle

A. Viral Entry

L-, and S-HBsAg, but not M-HBsAg, constituting HBV envelope, are crucial for viral entry into hepatocytes. The preS1 domain, forming the N-ter domain specific to L-HBsAg, harbors 3 specific domains, between aa 2 and 75, that are required for HBV infectivity : i) the myristoyl anchor, ii) a putative receptor binding site (aa 2-48) and iii) a still unknown function domain between aa 49 and 75. Moreover, the a-determinant, shared by the 3 envelope proteins, is a conformational epitope that can attach Heparan Sulfate Proteoglycans (HSPG) at the surface of hepatocytes (Le Duff et al., 2009; Leistner et al., 2007). HSPG is a low affinity HBV receptor and is abundantly expressed in the extracellular matrix of various cell types, including

hepatocytes. Among HSPG family, Glypican 5 was demonstrated to specifically mediate HBV entry into hepatocytes (E. R. Verrier et al., 2016).

Until 2012, the exact process of viral entry was not completely elucidated, as attachment of the viral envelope proteins was demonstrated to be necessary but not sufficient for achieving complete viral infection. Through protein affinity purification and Mass Spectrometry analysis, the Sodium Taurocholate Cotransporting Polypeptide (NTCP) was identified as HBV receptor (Yan et al., 2012a). NTCP expression in non-permissive cells is sufficient to make cells permissive to HBV entry, and the stable expression of NTCP in a hepatoma cell line (HepG2 cells) allows an infection reaching the levels obtained in Primary Human Hepatocytes (PHH), the natural host of the virus (Yan et al., 2012a). Besides, myristoylation of PreS1 domain of L-HBsAg was shown to be crucial for NTCP interaction, and thus for HBV infection (Gripon et al., 1995; Liu et al., 2016).

HBV internalization after NTCP binding are still poorly understood processes. Majority of enveloped viruses enter cells through the endocytic pathway. It was hypothesized that HBV also escapes from endosomes through low pH-dependent membrane disruption sustained by preS1 (9-24) hydrophobic peptide (Liu et al., 2016). Evidences for endosomal trafficking were also highlighted, notably *via* Rab5/Rab7, two proteins regulating newly endocytosed vesicles transport to early endosomes, that were demonstrated to be required for HBV internalization (Macovei et al., 2013). However, it remains controversial whether HBV particles are internalized through clathrin- (Huang et al., 2012) or caveolin- (Macovei et al., 2010) dependent mechanisms.

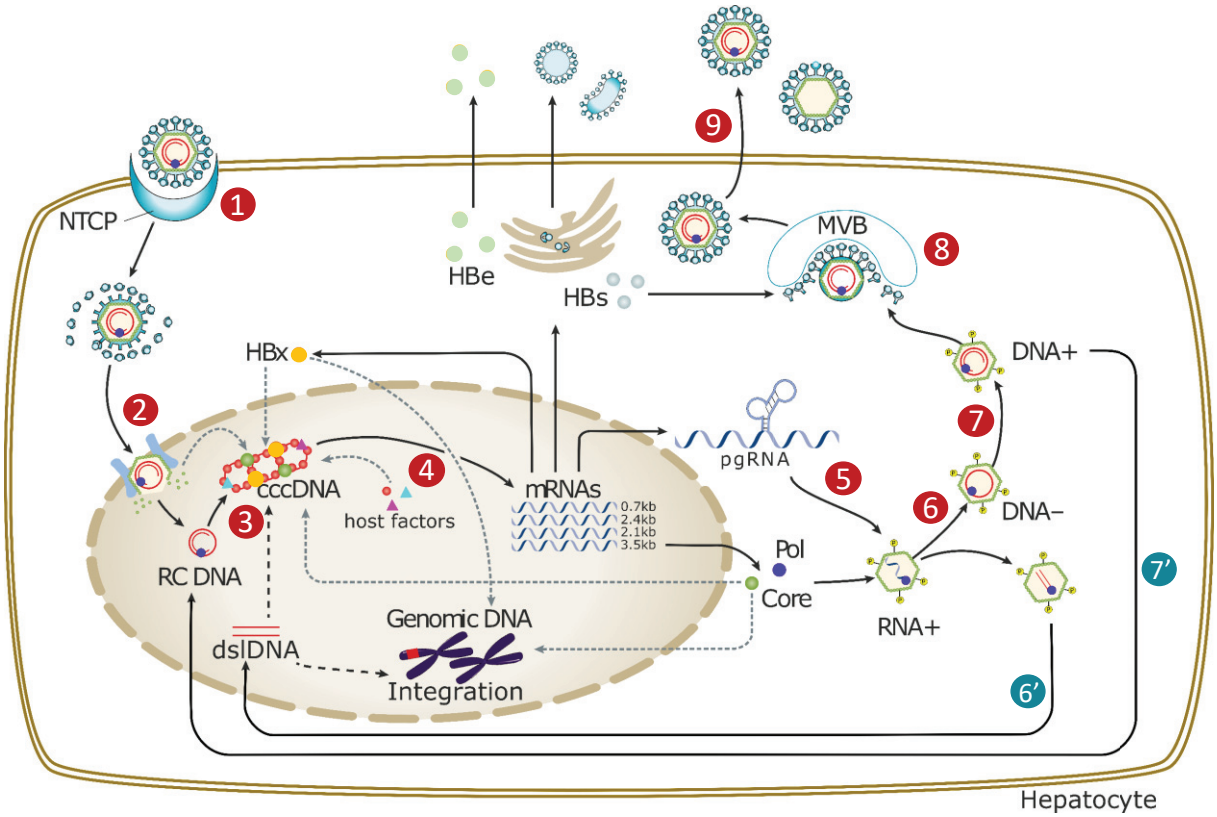


Figure 9 : HBV viral life cycle

HBV enters differentiated hepatocytes *via* its receptor NTCP (step1). At the nuclear pore, the capsid dissociates and rcDNA is released (step2), rcDNA is subsequently converted into cccDNA (step3), which is the unique template for the viral transcription (step 4). The viral transcription generates an RNA intermediate called pregenomic RNA, which is encapsidated (step5) while retro-transcribed in a (-)-DNA strand (step6). The (+)-DNA strand is further synthesized (step 6) to form the rcDNA. In 10% of cases, *in-situ* priming of (+)-DNA strand synthesis occurs and leads to the synthesis of a dsL-DNA (step 6'), that can go to the nucleus to form a defective cccDNA or to be integrated in the host genome. rcDNA-containing nucleocapsids can be enveloped (step 7) and secreted (step 8) to form new infectious particles, or recycled to the nucleus to replenish the cccDNA pool (step 7').

B. Nuclear import

HBV rcDNA is a Pathogen Associated Molecular Pattern (PAMP) that can be sensed by Pattern Recognition Receptors (PRRs) while located into the cytoplasm of infected hepatocytes (Mogensen, 2009). To lower the risk of recognition and remain protected against the intrinsic cellular immune system, the nucleocapsid is assumed to stay intact until genome release into the nucleus. During retrograde transport toward the nucleus, nucleocapsids interact with the microtubular system, helped by the functional binding partner DynLL1, with empty and mature capsid affinity for microtubules being stronger than RNA containing capsids (Osseman et al., 2018).

Once located at the Nuclear Pore Complex (NPC), HBV nucleocapsid can directly interact with nuclear import receptors, the so-called karyopherines (importin α and β). Indeed, HBc C-ter domain bears an importin-β binding domain (IBB) that allows direct interactions with importin-β and a NLS that can bind importin-β via an importin-α intermediate (Gallucci and Kann, 2017). However, only non-assembled core proteins can exhibit IBB in their CTD, meaning that assembled capsids only interact with NPC through exposed NLS in order to translocate nucleocapsids into the nucleus. Interaction between HBc CTD and importins depends on CTD exposure and capsid status. Indeed, genome maturation into nucleocapsids is linked to increased CTD exposure as the synthesis of the negative DNA strand by reverse-transcription leads to the release of the C-ter domain of HBc. Subsequently, exposed CTD becomes phosphorylated by PKC (Kann and Gerlich, 1994), resulting in a strong binding capacity of nucleocapsids to the nuclear envelope at the NPC (Gallucci and Kann, 2017). Immature and mature capsids are both able to interact with NPC, notably by strongly binding Nup153, a key component of NPC (Schmitz et al., 2010, p. 153) and to reach the nuclear basket. However, only mature capsids can release their genome into the karyoplasm, meaning that the outcome of HBV genome is regulated into the nuclear basket (Rabe et al., 2003). It is important to note that other factors may be involved in capsid import to the nucleus as 20% of capsids are still able to reach the karyoplasm in Nup153 silenced cells (Schmitz et al., 2010). Once capsids attain the karyoplasm, HBV rcDNA has to be uncoated and viral polymerase has to be released in order to complete the (+) strand DNA and ligate both extremities to form cccDNA. HBV

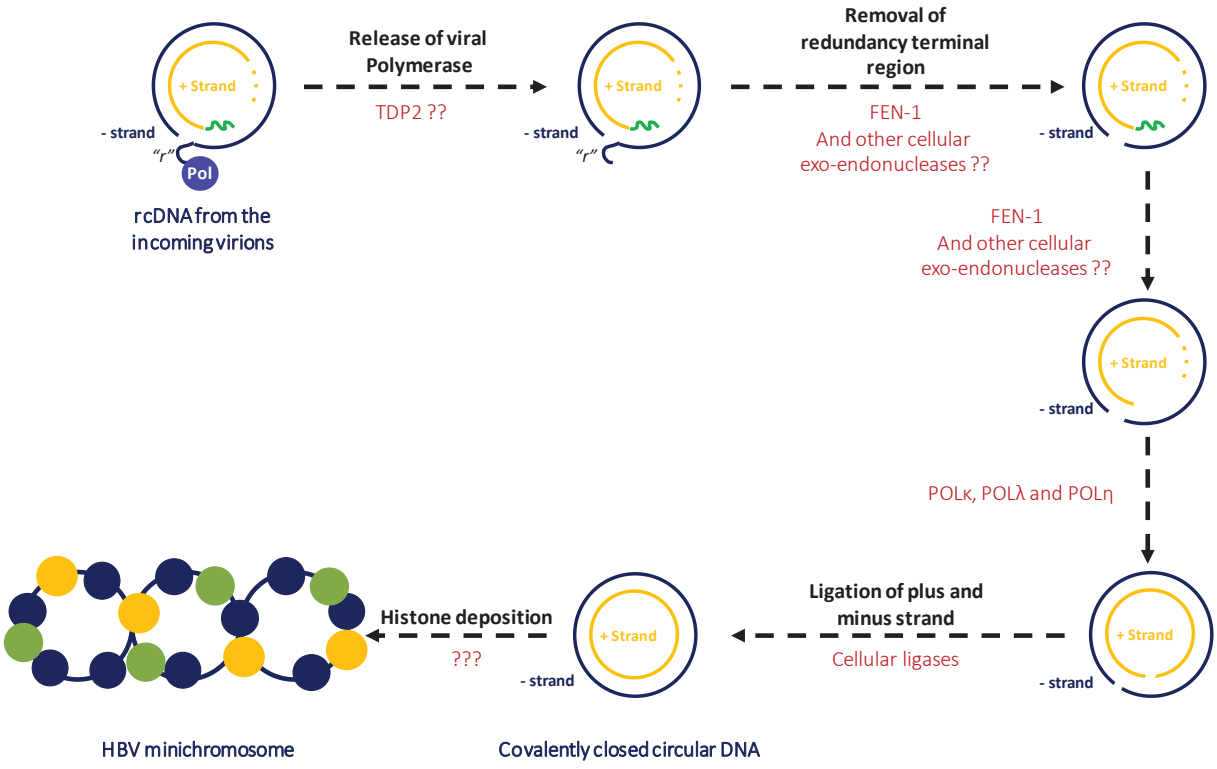


Figure 10 : rcDNA conversion to cccDNA

A multi-stepped process is required to proper rcDNA to cccDNA conversion. The following steps are required: i) the removal of the covalently bound viral polymerase from the 5'end of the negative strand and the removal of the redundancy terminal region "r"; ii) the degradation of the RNA oligomer (used to prime plus-strand DNA synthesis during pgRNA retro-transcription (see Figure 12)) from the 5'end of the positive strand; iii) the completion of the plus strand; iv) the ligation of both DNA extremities. In red are named the host proteins identified to be involved in the different conversion steps. It remains unknown whether these steps are concomitant or kinetically ordered.

nucleocapsid disassembly seems to initiate at specific sites of HBV genome, and partially opened capsid shells have been identified as uncoated intermediates, yet, 40% of rcDNA is found to be capsid-free in the nucleoplasm (Köck et al., 2010). Hypothetical models suggest that after capsid disassembly and genome release at the NPC, HBc proteins would dimerize to subsequently re-assemble to empty capsids (Gallucci and Kann, 2017) and/or stably associate with cccDNA to insure proper nucleosomal distribution (C. Thomas Bock et al., 2001).

C. cccDNA regulation

a. Formation

cccDNA is the persistent form of HBV DNA and is the crucial intermediate of viral replication. With no Origin of Replication (ORI), cccDNA pool is maintained through *de novo* infection or rcDNA-containing nucleocapsid recycling to the nucleus. Bromodeoxyuridine incorporation experiments allowed the identification of an asymetric replication of cccDNA, in agreement with reverse-transcription of an RNA intermediate (Pourcel and Summers, 1986).

After releasing into the nucleoplasm, rcDNA conversion into transcriptionally active DNA is a multi-step process that strongly relies on cellular protein activity. rcDNA conversion to cccDNA requires the following steps: i) the removal of the covalently bound viral polymerase from the 5'end of the negative strand and removal of the redundancy terminal region "r"; ii) the degradation of the RNA oligomer (used to prime plus-strand DNA synthesis during pgRNA retro-transcription (*see below "pgRNA transcription, retro-transcription and encapsidation"*) from the 5'end of the positive strand; iii) the completion of the plus strand; iv) the ligation of both DNA extremities **(Figure 10)**.

For some steps required for rcDNA conversion to cccDNA, cellular proteins have been clearly identified, yet, for other steps, technical limitations impair the uncovering of cellular factors. In addition, it is important to note that the above-described steps can occur either in a time-structured manner, or concomitantly, and that this question still remains open.

It has been described that removal of the covalently attached viral polymerase from the 5'end of the minus strand was completed by the Tyrosyl phosphodiesterase 2 (TDP2) (Königer et al., 2014). Moreover, being the only viral protein with an enzymatic activity, the viral polymerase was assumed responsible for the plus strand completion during cccDNA formation. However,

recent infection experiments demonstrated that inhibition of viral polymerase activity with specific inhibitors did not impair cccDNA formation and brought the hypothesis that cccDNA synthesis mostly depends on cellular factors (Qi et al., 2016). Qi and colleagues also demonstrated with this recent study that the polymerase κ (Polκ), and to a lesser extent Polη and Polλ, play a crucial role in the plus strand DNA synthesis in *de novo* infection.

Cleavage and/or degradation of the RNA oligomer at the 5'end of the plus strand and the removal of the redundant terminal region "r" from the 5'end of minus strand (that is remaining after viral polymerase detachment) prior ligation of DNA strands are both decisive events to achieve cccDNA formation. According to the molecular mechanism theoretically required to achieve these processes, cellular exo- and/or endo-nucleases and ligases seem to be fitting candidates. Besides, these enzymes belong to DNA repair pathways, assumed to be involved in rcDNA conversion to cccDNA (Gómez-Moreno and Garaigorta, 2017). The "r" region, located in the 5'end of the minus strand, is 10 nucleotides long and is presumed to form a 5'-flap DNA structure. A flap structure is a single-stranded DNA (or RNA) that protrudes from a dsDNA molecule (Tsutakawa et al., 2011). Through its ability to cleave 5'-flap DNA structures, flap structure-specific endonuclease 1 (FEN1) constitutes a good candidate in cccDNA formation processes by cleaving "r" sequence from rcDNA. Indeed, by specific FEN1 activity inhibitor, and siRNA, shRNA and CRISPR/Cas9 downregulation, Kitamura and colleagues demonstrated that FEN1 is involved in cccDNA formation. Interestingly, by *in vitro* assay, they also demonstrated that DNA polymerase, ligase and FEN1 activities were sufficient to convert rcDNA to cccDNA, and that DNA polymerase and ligase, alone, were not efficient to achieve cccDNA conversion (Kitamura et al., 2018). Noteworthy, DNA Damage Response (DDR) elements are largely regulated by cell cycle. As HBV replicates in quiescent hepatocytes, it would be of high interest to study DNA repair elements that are enzymatically active during G0 phase.

In addition, through its capacity to bind ssDNA and to act as scaffolding protein involved in homologous recombination, the role of SAMHD1 was recently investigated in cccDNA formation. SAMHD1 was demonstrated to be important for cccDNA levels in early infection, where its silencing induced a decrease of cccDNA amount of around 90% (Wing et al., 2019). However, the question of which step(s) is regulated by SAMHD1 has not been addressed.

Finally, an alternative pathway for cccDNA formation has been highlighted for DHBV cccDNA synthesis. Indeed, during pgRNA retro-transcription, priming failure can occur and leads to the formation of a dsL-DNA instead of rcDNA. DsL-DNA synthesis happens in 10% of the cases (Staprans et al., 1991). By Non-homologous End Joining (NHEJ) mechanism, DHBV dsL-DNA can be converted onto cccDNA, notably with the implication of Ku80, yet leading to defective cccDNA molecules, containing insertions or deletions around the ligated sites (Guo et al., 2012). Noteworthy, dsL-DNA is the predominant form leading to viral DNA integration into host genome (Yang and Summers, 1999).

b. Chromatin Structure

In 1994, cccDNA was, for the first time, demonstrated to harbor a chromatin-like structure, since electron microscopy characterization of HBV nucleoprotein complexes revealed a "beads-on-a-string" nucleosomal structure. H3, H4, H2A and H2B histones were detected in the sedimented fractions containing HBV DNA, and their presence was further demonstrated by Western-Blot. H1 and H5 histones were also identified to be present with these experiments (Bock et al., 1994). The year after, Newbold and colleagues confirmed the presence of nucleosomes on DHBV cccDNA, in infected ducks, describing a peculiar nucleosomal arrangement. Indeed, in eukaryotic cells, the nucleosome is composed of 146 bp DNA wrapped around (H3, H4, H2A, H2B)$_2$ octamer. Nucleosomes are separated by approximatively 50 bp of linker DNA associated to H1 and H5. For DHBV cccDNA, micrococcal nuclease digestion, cleaving the linker DNA between nucleosomes, revealed smaller DNA fragments, of around 150 bp, compared to the 200 bp DNA fragments expected. This higher compaction of DHBV cccDNA compared to eukaryotic chromatin confers a lower sensitivity of DHBV cccDNA to soft micrococcal nuclease activity. Moreover, with two-dimension electrophoresis of DHBV nucleoproteins, cccDNA was shown to cluster in two major structural subgroups. A first population with 18, 19 or 20 nucleosomes was identified with a 150 bp DNA surrounding nucleosomes and a small amount of linker DNA. A second population was highlighted, with half of nucleosomes, on a single tract or organized in various nucleosome groups (Newbold et al., 1995). The decrease in nucleosomal spacing described by Bock *et al.* (1994), and later confirmed by Newbold *et al*. (1995) can be explained by the presence of non-histone proteins associated to the viral minichromosome. HBV nucleoprotein immunoprecipitation experiments revealed the presence of HBc as part of the complex,

suggesting that HBc is bound to HBV DNA (Bock et al., 1994). Later, HBc was confirmed to be present on cccDNA, with a preferential binding affinity for dsDNA compared to partially dsDNA, and to be involved in the reduction of the nucleosomal spacing of around 10% (C. Thomas Bock et al., 2001).

c. Epigenetic Regulation

Chromatin structures undergo several epigenetic modifications that play key roles in gene expression regulation. The DNA wrapped around nucleosomes can be methylated and the methylation profile influences association strength with nucleosomes. Moreover, histone tails protruding from the core nucleosome can also experience several chemical modifications modulating the compaction state of the chromatin and consequently its transcriptional state. As for cellular chromatin, HBV minichromosome is subjected to different epigenetic regulation processes.

DNA Methylation

HBV DNA methylation on specific sequences has been identified by bisulfite-specific PCR enabling the identification of methylated cytosines by conversion of these residues to uracil. Three CpG islands were found on HBV genotype D DNA sequence at the following locations : CpG#1 (67-392), is near the preS2 ATG, CpG#2 (1033-1749) is overlapping X ORF and EnhII and located 1 bp upstream of the basic core promoter, and CpG#3 (2215-2490) is at the polymerase ATG (Jain et al., 2015). Moreover, it has been demonstrated that HBV infection positively regulates the expression of all the cellular DNA methyltransferases (DNMT1, DNMT2, DNMT3a and DNMT3b), in turn leading to HBV DNA methylation and viral protein downregulation **(Figure 11)**. Among them, DNMT3a was demonstrated to be implicated in HBV DNA methylation and reduced HBe and HBs antigens secretion (Vivekanandan et al., 2010). The methylation of HBV DNA sequence is also linked to a global decrease of viral mRNA and protein levels. A cccDNA-like molecule artificially methylated and transfected into hepatocytes also demonstrated a decreased secretion of HBsAg (Vivekanandan et al., 2009). Interestingly, methylation of CpG#3 is enriched on HBV genome extracted from tumoral tissue of HCC patients, compared to CHB and HBV-induced cirrhosis patient samples (Jain et al., 2015).

Histone modifications: methylation and acetylation

HBV chromatin has been reported to be associated with histone proteins. All the canonic histones composing the core nucleosome (H3, H4, H2A, H2B) have been described associated to cccDNA (C.Thomas Bock et al., 2001). N-ter histone tails are known to undergo several post-translational modifications (PTMs): acetylation and methylation are broadly studied, yet, many other have been identified (Zhao and Garcia, 2015). The chromatinized structure of cccDNA suggests a transcriptional regulation mediated by histone PTM **(Figure 11)**. Pollicino et al. (2006) assessed acetylation status of cccDNA-bound H3 and H4 and correlated hyper-acetylation to high viremia in HBV-infected patients. Contrarily, H3/H4 hypo-acetylation was rather correlated to low viral load. Moreover, *in vitro* experiments confirmed these results by artificially maintaining the acetylation status of cccDNA-bound histones, and an increase in viral transcripts, cytoplasmic viral intermediates and secreted particles was observed (Pollicino et al., 2006). *In vitro* specific cccDNA Chromatin immunoprecipiation (cccDNA-ChIP) also revealed the presence of histone acetyltransferases such as CPB, p300 and PCAF/GCN5 on the viral minichromosome. Histone deacetylases (HDAC) also associate to cccDNA, notably HDAC1 and Sirt1 (L. Belloni et al., 2009b; Pollicino et al., 2006). HDAC1 recruitment to cccDNA lead to cccDNA transcriptional repression (Pollicino et al., 2006).

Pegylated-IFN treatment inhibits the viral replication through global cccDNA-bound histone hypo-acetylation and the recruitment of transcriptional co-repressors together with a decreased recruitment of the transcription factors STAT1 and STAT2, in cell culture and uPA-SCID mice (Belloni et al., 2012). Recently, ChIP-seq experiments following MNase digestion of cccDNA confirmed the presence of active marks (H3K4me3, H3K27Ac and H3K122Ac) and an underrepresented presence of negative marks (H3K9me3 and K3K27me3) on HBV cccDNA (Tropberger et al., 2015). The cccDNA transcriptional repression induced by IFN treatment was also confirmed in this study with a reduced level of active histone PTMs. Interestingly, a different pattern of nucleosome distribution was observed in hepatoma cell line (HepG2-NTCP) and primary human hepatocytes. Besides nucleosomal distribution, also histone PTMs seem to vary according to the cell type of tissue samples. Indeed, active marks, such as H3K4me3, H3K27Ac and H3K122Ac are mainly enriched at preC/pgRNA TSS in HepG2-NTCP cells with poor representation at other TSS, while peaks representing association of these PTMs to cccDNA are mostly visible at X and preS1 TSS in PHH and X and preS2 TSS in HBV-

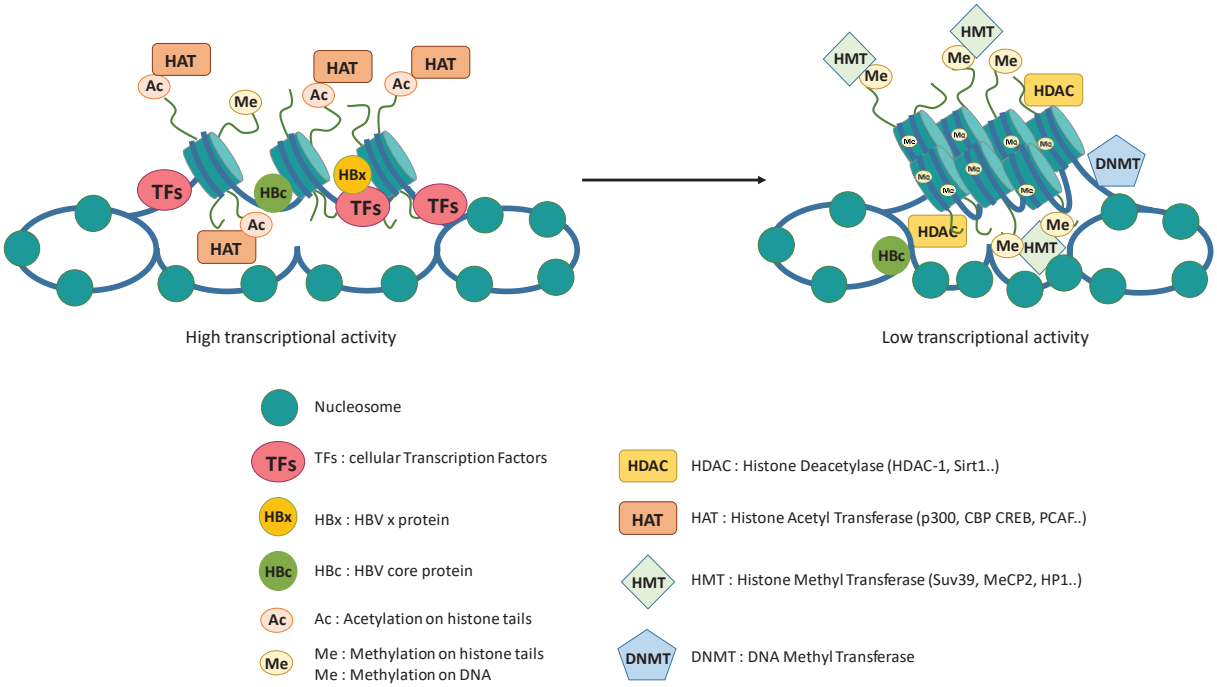


Figure 11 : Epigenetic regulation of cccDNA

cccDNA can be methylated by different DNA methyltransferases (DNMTs). cccDNA is chromatinized and harbors nucleosomes. cccDNA-bound histones car undergo PTMs modulated by histone acetyl transferases (HATs), histone deacetylases (HDACs) and histone methyl transferases (HMTs). Viral proteins, HBc and HBx, are also associated to cccDNA and participate to its epigenetic regulation.

-Adapted from Levrero et al. 2009

positive liver samples. Enrichment for positive histone PTMs downstream of the Enhancer I was however shared for all these three models, reflecting its active state. Overall, distribution of histone PTMs can vary between cell culture models and infected tissues, meaning that the viral chromatin organization probably depends on and adapt to the cell type-specific environment (Tropberger et al., 2015).

Viral proteins associated to cccDNA

Besides being associated to histone proteins, cccDNA was also demonstrated to be associated with non-histone proteins, and particularly with the viral HBx and HBc proteins (C.Thomas Bock et al., 2001; Lucifora et al., 2011). Notably, the viral transactivator HBx is recruited onto cccDNA and is mandatory for full transcriptional activity of cccDNA (Lucifora et al., 2011), notably *via* influencing the epigenetic state of cccDNA **(Figure 12)**. Indeed, while WT virus harbors active histone marks, HBx negative (HBV-X^{neg}) mutant displays H3/H4 global hypo-acetylation, less H3K4me3 and an increased recruitment of transcriptional repressors. Moreover, the repressive H3K9me2/3 marks is more associated to cccDNA in HBV-X^{neg} leading to the recruitment of Heterochromatin Protein 1 (HP)1, a reader of H3K9me3. SETDB1, one of the H3K9 methyl-transferases, was demonstrated to repress the transcriptional activity of cccDNA by increasing the deposition of the methyl group on H3K9 (Rivière et al., 2015). In addition, CREB transcription factor activity on HBV DNA is maintained by proper phosphorylation mediated by cAMP, and can be dephosphorylated and inactivate by the Protein Phosphatase 1 (PP1). Interestingly, HBx inhibits PP1 activity, thus insuring a long-lasting CREB phosphorylation and subsequent HBV DNA permissive transcription state (Cougot et al., 2012). Recently, HBx was demonstrated to promote the interaction between Structural maintenance of chromosomes protein 5 and 6 (Smc5/6) and a subunit of the ubiquitin proteasome complex (the E3-ubiquitin ligase Cul4). This HBx-mediated interaction induces the degradation of Smc5/6 and allows the maintenance of a transcriptionally active state of cccDNA (Decorsière et al., 2016). Interestingly, cccDNA-encoded HBx can undergo mutations in HCC patients. Most of HBx mutants, yet, are still able to support cccDNA transcriptional activity by degrading Smc5/6 complex. On the contrary, some mutations, principally in HBx C-ter domain and in the region between aa 52 and aa 87, exhibit a loss of ability to promote cccDNA transcriptional activity due to their inability of degrading Smc5/6 (Rivière et al., 2019).

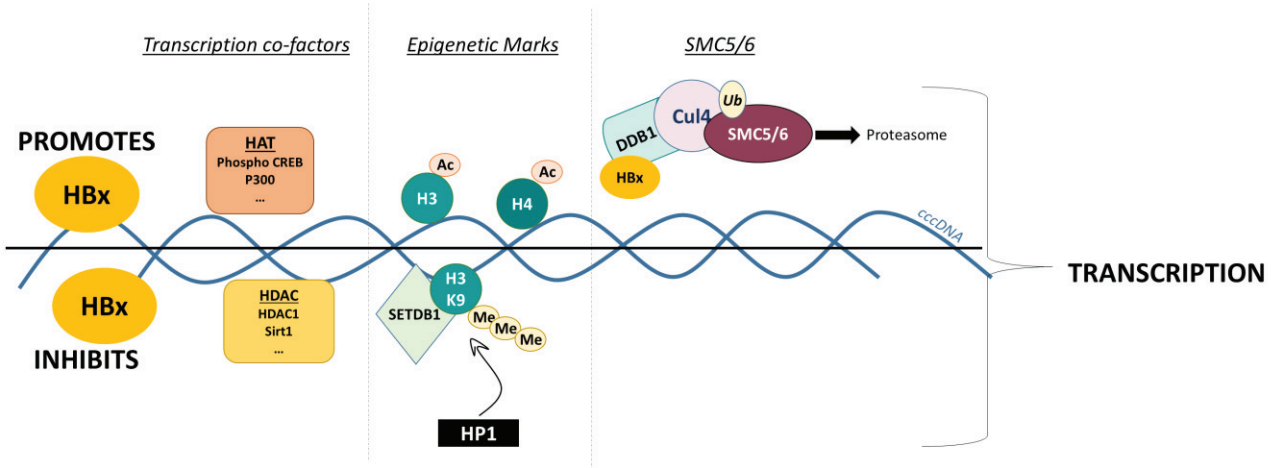


Figure 12 : HBx promotes cccDNA active transcriptional state

HBx acts by different but converging mechanisms to establish a global permissive epigenetic status to cccDNA crucial for viral transcriptional activity.
HBx favors the recruitment onto cccDNA of transcriptional activators, such as Histone Acetyltransferases (PhosphoCREB, p300..), positive epigenetic marks (H3Ac and H4Ac), and promotes the degradation of the restiction factor SMC5/6 through interaction with DDB1 and intermediate of the E3-Ubiquitine ligase Cul4.
HBx also prevents the recruitment onto cccDNA of transcriptional repressors such as Histone Deacetylases (HDAC1, Sirtuin1,…), negative histone marks such as H3K9me3, deposited by the Histone Metyltransferase SETDB1, and further recruitment of Heterochromatin Protein (HP) 1.

This study suggests that HBx C-ter and 52-87 aa region is important for Smc5/6 interaction. However, the precise interaction site needs to be precisely defined.

Moreover, the viral core protein HBc has also been found on the viral minichromosome, and influences the nucleosomal organization on cccDNA (C.Thomas Bock et al., 2001). It has been suggested that HBc preferentially binds the CpG#2 on the viral genome, and HBc binding abundance on CpG#2 positively correlates with the level of serum DNA in patients. Moreover, HBc binding to CpG#2 was correlated to CREB-binding protein (CBP) binding at that same location and DNA hypo-methylation (Guo et al., 2011). These observations were obtained after MNase digested and sonication treated extracts followed by cccDNA-ChIP. Nonetheless, no proof of an effective digestion of the viral minichromosome allows the confirmation of such a preferential recruitment site, and methods to efficiently digest cccDNA still need to be improved.

Host proteins associated to cccDNA

Hepatotropism of HBV suggests a key role of hepatocyte-specific factors in HBV life cycle. Indeed, host factors regulating cccDNA transcriptional activity are mostly enriched in hepatocytes compared to other cell types. Liver-enriched factors are called hepatocyte nuclear factors (HNF). Among them, HNF4α is a key regulator of approximately 40% of hepatocyte genes and its inhibition by shRNA also leads to strong decrease in HBV RNA transcripts and DNA replication in hepatoma cell line and mice (He et al., 2012). Moreover, other HNF have been identified to regulate HBV RNA transcription. Notably, HNF1α and C/EBPα, another liver-specific transcription factor, seem to control cccDNA regulatory elements, such as preS1 promoter and Enhancer II to favor transcriptional activation (Kim et al., 2016). Contrarily, some HNF have been reported to be transcriptional inhibitors. That is the case for HNF3β and HNF6 that both inhibit HBV gene expression in hepatocytes (Kim et al., 2016).

On another hand, other factors which are not liver-enriched nuclear factors have also been described to regulate HBV transcription, positively or negatively. First, TARDBP depletion impairs the levels of HBV replicative intermediates, HBV RNA and proteins. The two RNA recognition motifs of TARDBP have been identified to be essential for TARDBP ability to activate the transcription from the core promoter (Makokha et al., 2019). PUF60 have also been reported to positively regulate HBV core promoter activity as its overexpression

increases levels of 3.5 kb RNA (Sun et al., 2017). Oppositely, TIP60 complex has been described to negatively regulate HBV transcription. The level of HBV RNA, HBe- and HBs-antigens strongly increase in TIP60 depleted cells, without any significant impact on cccDNA level nor 3.5 kb RNA stability. The acetyltransferase TIP60 seems to inhibit viral transcription through the recruitment of the bromodomain-containing cellular repressor (Brd4) on TIP60-acetylated H4 (Nishitsuji et al., 2018). Furthermore, Sirtuin 3 (SIRT3), a cellular deacetylase, is also involved in HBV replication restriction. Notably, SIRT3 ectopic expression in HBV-infected cells decreases HBV RNA transcription and HBV DNA replicative intermediate. SIRT3 associates to cccDNA and modulates its epigenetic state by recruiting onto cccDNA the histone methyltransferase suppressor SUV39H1 that catalyzes H3K9me3 deposition, and contrarily leads to a decreased recruitment of the histone methyltransferase SET1A catalyzing H3K4me3 active mark deposition. Importantly, SIRT3 ectopic expression also decreases RNA Pol II recruitment onto cccDNA, leading to a global repressive state of cccDNA and further transcription inhibition (Ren et al., 2018).

D. pgRNA transcription, encapsidation and retro-transcription

Pregenomic RNA is a key molecule in HBV life cycle as it is the RNA intermediate for the viral replication. It is the template for retro-transcription into a new rcDNA and subsequent formation of new infectious particles. Indeed, being longer-than-genome and measuring 3.5 kb long, pgRNA bears the whole genome sequence. Its transcription encompasses the genome length by 120 redundant nucleotides. pgRNA is capped in its 5'end and polyadenylated in 3'end. pgRNA is a bicistronic mRNA encoding both HBc and Pol. This latter ORF partially overlaps HBc ORF and is less translated than HBc, in a ratio 1:240, resulting from a ribosomal shunting (Sen et al., 2004). Moreover, pgRNA contains specific features that are necessary for its retro-transcription and encapsidation in progeny capsids that can be either enveloped and secreted to form new infectious particles or recycled to the nucleus to maintain the cccDNA pool.

Despite a still poorly identified tertiary structure, the secondary structure of pgRNA is well known **(Figure 13)**. First, pgRNA harbors 5' and a 3' ε loops. ε is a stem-loop highly conserved in hepadnaviruses, harboring a bulge region and an apical loop. Only 5' ε loop serves as encapsidation signal. Noteworthy, the 3' ε loop is shared by all the viral transcripts and is not sufficient to mediate encapsidation (Rieger and Nassal, 1996). Moreover, pgRNA contains the ϕ sequence, complementary to the 5' half of ε. Finally, due to the transcript length, pgRNA also contains two copies of the Direct Repeat 1 (DR1) sequence, while only bearing one copy of the DR2 (Hu, 2009).pgRNA retro-transcription and encapsidation is a multi-step process, requiring structural reorganizations and *cis* and *trans*-acting factors **(Figure 14)**.

First, the TP domain of the viral polymerase, located in its N-terminal end, interacts with the bulge regions of the 5', but not the 3', ε loop. To interact, Pol has to be in its active conformation, implicating constant ATP supply, exchanged by different Heat Shock Proteins (Hsp), and the associated nucleotide exchange factors (Beck and Nassal, 2007; Hu et al., 2004). Stable Pol-ε loop association mediates the recruitment of HBc dimers leading to packaging initiation. In the meantime, protein priming to pgRNA 5' ε-loop drives the synthesis of a small (-)-DNA sequence of 4 nt-length, to which Pol will remain covalently bound *via* Y63 that provides the OH group for covalent bond with the first synthesized nucleotides, until its removal during cccDNA formation.

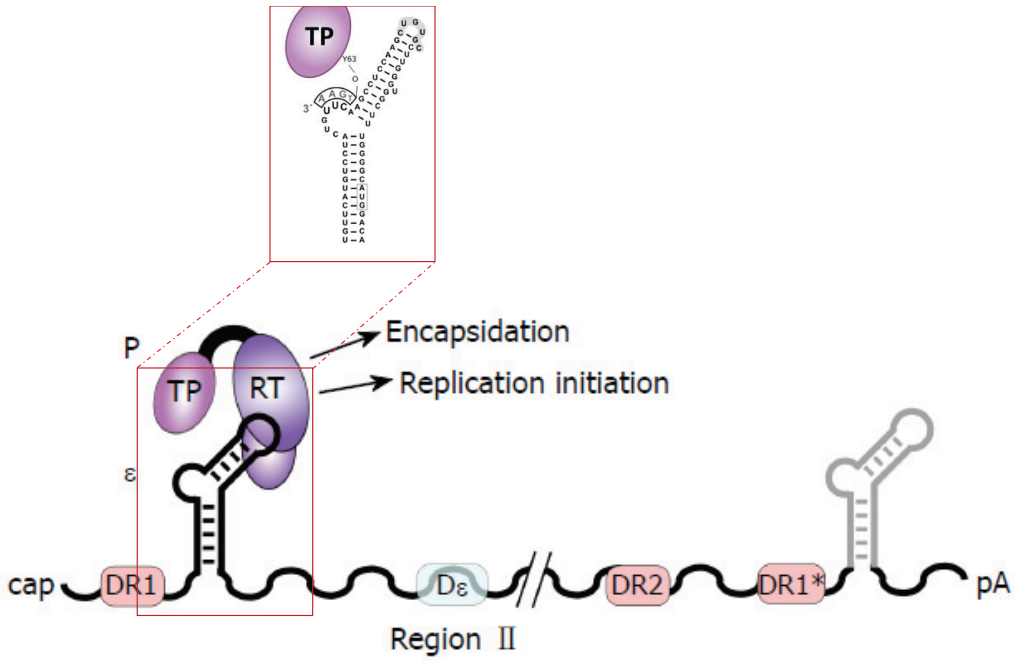


Figure 13: pgRNA secondary structures

pgRNA harbors a 5' and a 3' ε-loops that are secondary structures forming a bulge, to which the viral Pol-TP is recruited, and an apical loop. Direct repeat (DR1) is present twice on pgRNA sequence, while DR2 is only present once. pgRNA is 5'capped and polyadenylated.

-Adapted from Beck and Nassal, 20017

At this step, the first template switch occurs with the translocation of the neosynthesized 4-nt DNA primer (3'-AAGT-5') to the 3' DR1 sequence (5'-UUCA-3'), by sequence complementarity, together with the covalently attached viral polymerase. At that step, DNA primer hybridization to 3' DR1 is not the only factor involved in the translocation. Indeed, it has been proposed that pgRNA folding may be facilitated by i) the sequence complementarity of ϕ to the 5' half of ε; ii) yet-not-defined host factors, such as elongation initiation factor 4G (eIF-4G) that can bind both 5' cap and polyA tail. Interestingly, pgRNA 5' capping has been demonstrated to be essential for reverse-transcription but not the polyA tail (Jeong et al., 2000). (-)-DNA completion subsequently occurs through the 3' DR1 to the 5'end of the pgRNA, and pgRNA is concomitantly digested by the RNase H domain of the viral polymerase. The fate of the non-retro-transcribed 3' end of pgRNA, from downstream 3' DR1 sequence to the polyA tail, is still unknown.

RNase H-mediated pgRNA degradation is not complete, leaving about 15-18 nt in pgRNA 5'end, including 5' DR1 sequence, spared from digestion, probably protected by RT/RNase H domains steric encumbrance. This 5' capped RNA oligo is essential for (+)-DNA synthesis priming. It can be either extended from its initial location, in 10% of cases, leading to the so-called "*in-situ priming*", and initiating the synthesis of a dsL-DNA; or can be translocated to the DR2 sequence. This second template switch, essential for complete rcDNA synthesis, is mediated by complementarity to the 5' DR1 sequence, remaining in the RNA primer. From its new location, RNA primer is extended toward the viral Polymerase bound in 5' end of the (-)-DNA, including the 4-nt linker to Pol (Beck and Nassal, 2007).

A third template switch is finally required to fully synthesize rcDNA: circularization. Indeed, after having synthesized the (+)-DNA strand until the 5' end of the (-)-DNA strand, (i.e. where the Pol is still attached), the growing (+)-DNA is transferred from the 5' end to the 3' redundant ("r") sequence of the (-)-DNA from where it can be fully synthesized in mature rcDNA (Beck and Nassal, 2007). Although identity between 5' and 3' redundant sequences is important to fully achieve circularization, it is assumed that supplemental *cis*-acting elements are required to insure proper RNA primer translocation and circularization. Such elements have been identified for DHBV, and sequence similarity with HBV gives evidences on analogous events. Noteworthy, the synthesis of (+)-DNA strand is partially achieved and is

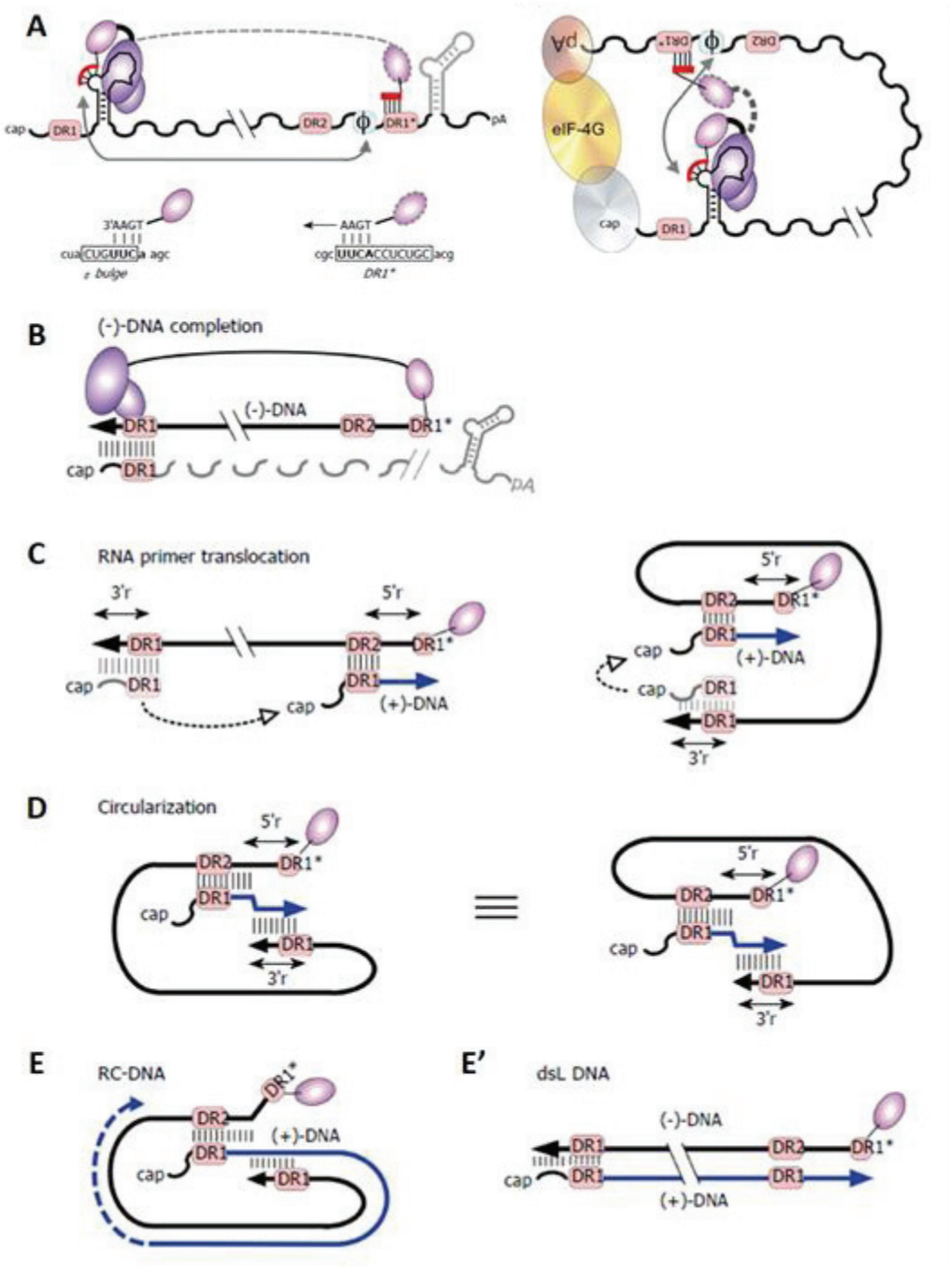


Figure 14 : pgRNA retro-transcription

The first step of pgRNA retro-transcription is the protein priming leading to the synthesis initiation of the (-)-DNA strand (A), followed by the negative strand elongation (B). After pgRNA partial degradation by the RNase H, the RNA primer is translocated to the DR2 sequence (C). The RNA priming leads to the synthesis of the (+)-DNA strand and the circularization (D) allows the proper rcDNA synthesis (E). In-situ priming can occur and lead to the synthesis of a dsL-DNA (E').
-Adapted from Beck and Nassal, 20017

heterogeneous in length, leading to the so-called “gap” that is further completed during cccDNA formation. This phenomenon could be explained by two hypotheses: i) a limited quantity of dNTPs trapped in the capsid shell, resulting in a more or less filled gap; ii) a space restriction constrained by the capsid shell. However, it is important to note that the capsid has a considerable structural plasticity (Böttcher et al., 2006).

Recently, pgRNA ε-loop has been demonstrated to be N6-methyl adenosine (m6A)-modified. Moreover, m6A modification of 5’ ε-loop, but not of 3’ ε-loop, has been proposed as regulator for pgRNA retro-transcription (Imam et al., 2018). However, precise rcDNA synthesis step and involved mechanisms have not been elucidated yet.

Finally, it is important to note that Core proteins undergo multiple phosphorylation/dephosphorylation on their Threonine and Serine residues, enriched in HBc C-terminal domain, during pgRNA retro-transcription and encapsidation. These dynamic (de)-phosphorylation processes play an essential role in pgRNA packaging and concomitant DNA synthesis (Hu, 2009). Besides, the phosphorylation status of the capsid is closely linked to rcDNA maturation (Gallucci and Kann, 2017). Indeed, during packaged (-)-DNA synthesis, HBc C-terminal domain is released outside the capsid shell, and exposed to phosphorylation by PKC (Kann and Gerlich, 1994).

E. Assembly and secretion

While HBV secreted virions, or Dane particles, are found in the serum of infected patients at around 10^9/mL, sub-viral particles (SVP) can be found 10^4 to 10^5 times more frequently. Among these SVP, the most secreted are spheres and filaments, but genome-free enveloped capsids can also be detected (Hu and Liu, 2017). More recently, HBV RNA-containing particles have been identified in the serum of infected patients, yet, in 100 to 1000 lesser extent than Dane particles (Rokuhara et al., 2006).

rcDNA-containing nucleocapsids are considered as mature nucleocapsids and have to acquire the host-derived lipid bilayer in which envelope proteins are embedded. They subsequently bud into the lumen of intracellular membrane compartments. Multi vesicular bodies (MVBs) are hypothesized to be compartments enclosing these particles (Hu and Liu, 2017). Interactions between the capsid and L-, and S-, (but not M-) envelope proteins rely on specific peptide domains. Indeed, in a very recent study based on an overexpression of mutated core

and envelope proteins, L-HBsAg was demonstrated to directly interact with HBc, through its preS1 domain, in its i-preS1 conformation, exposing a “matrix”-binding domain. L-HBsAg could thereby constitute an intermediate between S-HBsAg and core interaction. Regarding the core protein, residues L60, L95, K96, I126 and Y132 seem to be required for capsid interaction with envelope proteins (Pastor et al., 2019). However, these interactions have to be confirmed in an infectious model harboring a complete replication cycle.

The phosphorylation state of HBc C-ter domain, demonstrated to be associated with genome maturation and nuclear export, is not associated with envelopment and secretion processes (Hu and Liu, 2017). Eventually, Rab7 cellular protein was demonstrated to participate to viral secretion regulation. HBV infection leads to an increase of Rab7 activity leading to Rab7-dependent tubulation of MVBs and phagosomes. However, Rab7 knockdown induced an increased cytoplasmic viral particle accumulation and a subsequent secretion. By this mechanism, Rab7 probably constricts HBV particle release and contribute to its stealth (Inoue et al., 2015).

SVPs, spheres and filaments, are assumed to be secreted *via* the host cell constitutive secretory pathway, notably involving the endosomal sorting complex required for transport (ESCRT). L-, M- and S- proteins can efficiently bud from intracellular membrane compartments, between ER and Golgi Apparatus without any capsids. Noteworthy, SVPs can be stored in the ER lumen, and can cause cell toxicity in transgenic mice (Bruss, 2004).

Finally, genome-free enveloped capsids can also be secreted. As the envelopment and secretion processes were first demonstrated to be restricted to fully mature nucleocapsids, the hypothesis of a specific signal related to maturation emerged. A speculated “blocking-signal”, somehow induced by pgRNA or ssDNA-containing capsids, would block the envelopment process. During (+)-DNA synthesis, the “blocking-signal” transducing a “no-secretion” message would be removed. Thus, genome-free capsids should not possess this pgRNA- or ssDNA-related signal, thus explaining that genome-free capsids could also be enveloped and secreted (Hu and Liu, 2017). Nevertheless, this hypothesis is not applicable to RNA-containing particle envelopment and secretion.

4. Models for HBV replication study

HBV completes its life cycle only in fully differentiated hepatocytes of high primates. Due to this restriction in permissive hosts, only few *in vitro* and *in vivo* models are available to study complete HBV viral life cycle.

A. In vitro models

a. Hepatoma cell lines

Being the easiest cultured model, hepatoma-derived cells lines have been extensively used during the past decades. Most particularly, Huh7 and HepG2 cells lines are two cell lines widely exploited for viral-host interaction studies. However, these cell lines partially reproduce physiological hepatic functions and do not express the HBV receptor, requiring to bypass the entry step by transient or stable transfections, and thus preventing the study of full viral life cycle (E. Verrier et al., 2016). HepAD38 and HepG2.2.15 are two HepG2-derived cell lines stably expressing HBV genome and commonly used for HBV infectious particle production.

Since the identification of NTCP as a HBV receptor (Yan et al., 2012a), most of hepatoma-derived cell lines were stably transfected to overexpress human NTCP and became susceptible for HBV infection. These NTCP-overexpressing cell lines, such as Huh7-NTCP and HepG2-NTCP, allow *in vitro* study of the full HBV life cycle, yet HepG2-NTCP cells seem more susceptible to HBV infection than Huh7-NTCP (Ni et al., 2014).

HepaRG cells are non-transformed hepatic progenitors deriving from an HCV-induced hepatic tumor. In long-term dimethylsulfoxide (DMSO)-complemented culture conditions, HepaRG can differentiate either in cholangiocyte-like and hepatocyte-like cells. Differentiated HepaRG support both HBV entry and cccDNA production and are physiologically closer to primary hepatocytes than the previous-described hepatoma-derived cells lines, being a suitable model for studying HBV life cycle steps. In addition, HepaRG cells possess an efficient intrinsic innate immune response, rending this model suitable for innate immunity studies (Lucifora et al., 2010). Nevertheless, regardless of great advantages provided by HepaRG cell line, HBV infection levels are weak and cells need to be DMSO-differentiated for longtime. To overcome this issue, NTCP-overexpressing HepaRG cells have been developed, thus allowing HBV

infection at their undifferentiated stage, yet, with less infection efficiency compared to fully differentiated HepaRG cells (Ni and Urban, 2017).

b. Primary Human Hepatocytes

Being the only natural host for HBV, Primary Human Hepatocytes (PHH) cells are the most physiologically *in vitro* models for the study of HBV life cycle. However, PHH are isolated from liver resections from human donors, which renders the availability of this model highly limited. In addition, an important donor-to-donor variability is observed and even HBV infection efficiency depends on the donor genetic background (E. Verrier et al., 2016). *Tupaia belangari* Primary Hepatocytes (TPH), coming from a tree shrew, are the only non-primate animal cells susceptible to HBV infection. Noteworthy, TPH allowed the identification of NTCP as being the entry receptor for HBV (Yan et al., 2012a).

B. In vivo models

a. Models based on natural infection

First insights on HBV biology have been obtained *via* host-adapted viruses belonging to *Hepadnaviridae* family.

Among those models, DHBV that infects the Pekin Duck, has been extensively used for HBV cccDNA biology studies, due to its high cccDNA copy number per nucleus. However, DHBV shares only 40% homology with HBV and infected ducks do not develop liver diseases. Moreover, WHV that infects Woodchuck has a high degree of sequences homology to HBV and drives both acute and chronic infection, leading to HCC development in almost all infected individuals (Ortega-Prieto et al., 2019). This model is an interesting model, in particular for HCC development study, however, the lack of specific reagents makes it difficult to use.

Study of HBV biology is possible in chimpanzee and tupaia models, which are the best available models allowing infection by the human virus. Chimpanzees are the only primate model for HBV infection, sharing highly similar characteristics with human that allowed the development of HBV vaccines (McAuliffe et al., 1980). However, ethical constraints limit *in vivo* research on chimpanzees.

Besides, the northern tree shrews, or *Tupaia belangeri,* are the only non-primate model susceptible to HBV infection. Similarly to humans, tupaias develop chronic infections that can lead to HCC development (Ortega-Prieto et al., 2019). Due to the lack of tupaia specific reagents for experimentation, this model is not often used, yet it brought considerable insights in HBV fields, such as NTCP discovery (Yan et al., 2012a).

b. Transgenic Mice

To assess the potential oncogenic and virology roles of the different viral proteins, transgenic mice were initially constructed to express single viral proteins. Besides, transgenic mice expressing complete HBV replicon were generated and are able to produce infectious viral particles. However, being immunologically competent, these mice rapidly clear the infection and do not develop liver diseases. This model has been useful to test antiviral drug efficacy, yet, murine hepatocytes do not express NTCP receptor, thus limiting the complete study of viral infection and cccDNA biology (Hwang and Park, 2018).

c. Infectable models

In viral infection, immunocompetent animal models are required to fully reproduce a natural infection, and to study novel immunotherapeutic strategies.

To this aim, immunocompetent C57BL6 mice were transduced with an AAV containing HBV genome, and were demonstrated to support cccDNA persistence (Lucifora et al., 2017). Despite the presence of a functional immune system, being infected with an AAV, this model is not relevant to study viral entry and has several limitations also for the study of cccDNA. Indeed, cccDNA molecules co-exist with parental AAV constructs and it is not clear if they are produced by recombination or intracellular recycling of newly synthesized rcDNA (Lucifora et al., 2017).

In order to study the complete HBV viral cycle *in vivo*, infectable humanized chimeric mice were generated, according to two different strategies. The first one relies on SCID mice expressing urokinase-type plasminogen activator under control of albumin promoter (alb-uPA mice) that undergo liver failures compensated with PHH engraftment. This murine model allows HBV infection, yet, kidney and hematologic disorders constrains their usage to narrow period of time (Hu et al., 2019). PHH engraftment is also performed in a second model called

System	Virus and vector type	HBV replication	cccDNA	Infection Model	Immunology	Advantages	Limitations
Duck	DHBV	Yes	High copy number	Yes	Innate, adaptive, vaccine	High cccDNA copy number, infectivity, replication	Low similarity with human infection
Woodchuck	WHV	Yes	High copy number	Yes	Innate, adaptive, vaccine	High cccDNA copy number, carcinogenesis	Limited availability
Tupaia	HBV	Weak, transient	Low copy number	Weak, transient	Innate, adaptive, vaccine	Good alternative for infection studies	Limited availability, weak infection
Chimpanzee	HBV	Yes	Yes	Yes	Innate, adaptive, vaccine	Well characterized immune system. High similarity with human infection	Ethic and cost
Mouse	Tg mice – integrated HBV genome	Yes	No	No	Immune competent adaptive	Convenient inbred animals. Antiviral studies	No infection, no cccDNA, immune tolerance
	Adeno-AAV transduced mice_vector driven HBV DNA constructs	Yes	Yes	No	Immune competent acute infection	Studies in viral clearance	Transient vector driven infection – no spreading
	Human liver – chimeric mice	Yes	Yes	Yes	Yes	Long-term studies	PHH needed, Immune deficiency

Figure 15 : *Characteristics of in vivo systems used for HBV and HBV-related viruses*

FRG mice. These mice are immunodeficient Rag2 and IL2 γ receptor knockout. Moreover, FAH deficiency triggers liver injuries and enables liver repopulation with PHH. FRG mice can be HBV-infected and are a relevant model for investigating all stages of HBV infection, notably cccDNA establishment and maintenance, and also for direct antiviral agent studies (Hu et al., 2019). To counteract the lack of functional adaptive immune system, as described for alb-uPA and FRG mouse models, double-humanized mouse models have been generated. These models harbor both humanized innate and adaptive immune systems and a humanized liver, and support hepatotropic virus infections (HCV, HBV). For example, the HIS-HUHEP model develops human myeloid and lymphoid cell lineages and human liver tissue after human hematopoietic stem cells and human adult hepatocytes co-engraftement (Kremsdorf and Strick-Marchand, 2017).

In 2013, natural HBV infection was detected in Macaques (*Macaca fascicularis*) suffering from a chronic HBV infection. Due to its highly similar immune response with humans, *M.fascicularis* is a promising *in vivo* model to study HBV infection (Dupinay et al., 2013). Recently, *in vivo* HBV infection was achieved in macaques with viral-mediated expression of human NTCP on hepatocytes (Burwitz et al., 2017).

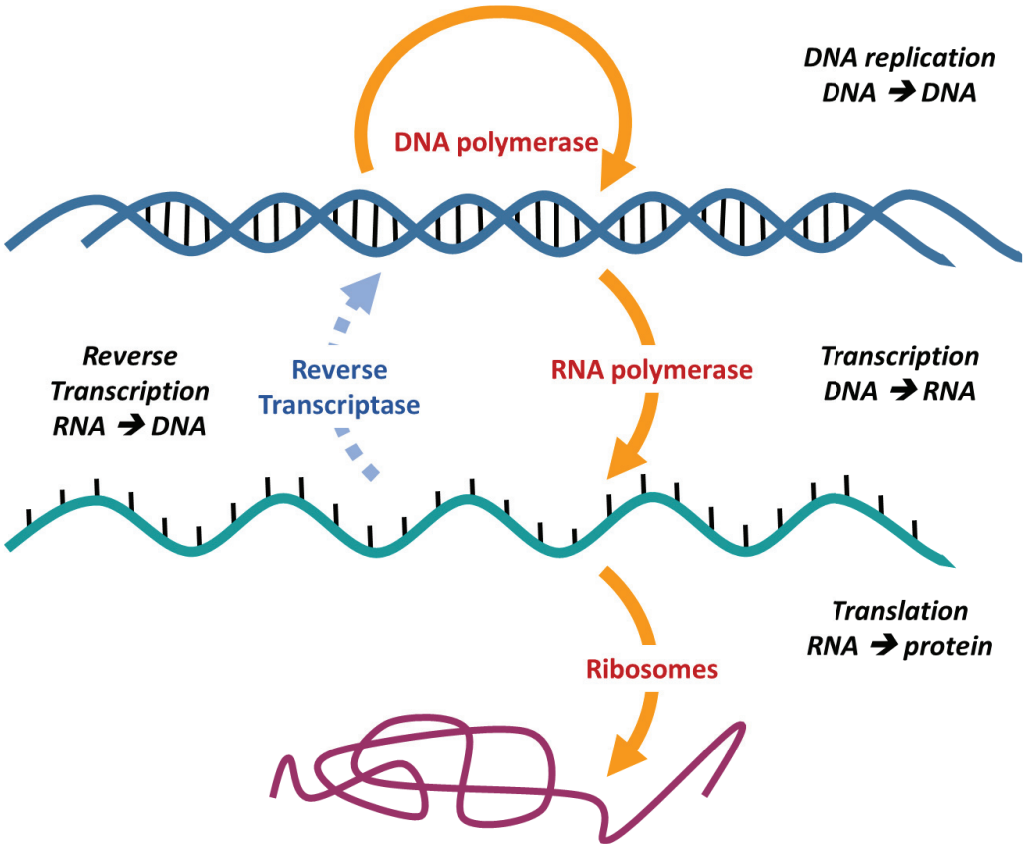


Figure 16 : Central dogrma of molecular biology

The first central dogma of molecular biology was assuming that a single gene was transcribed in a single mRNA by RNA polymerase, which in turn was translated into one protein, by ribosomes.

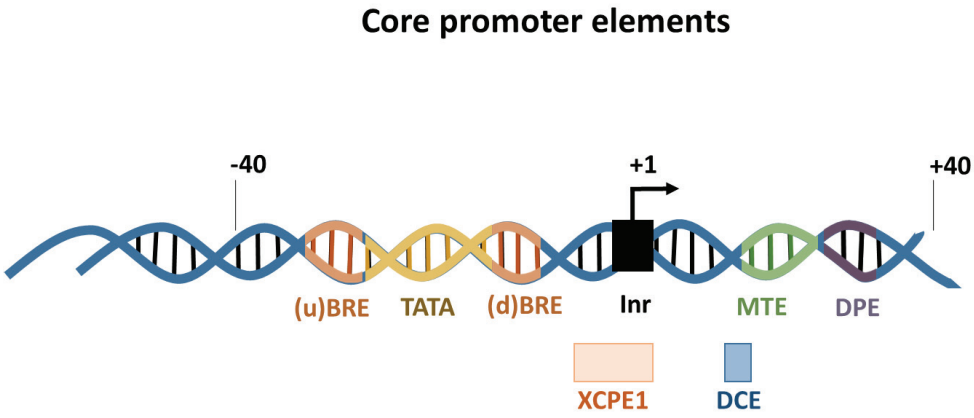


Figure 17 : Features of core promoter

The core promoter harbors different sequence motifs. First, the Inr motif contains the TSS, considered at the +1 position of the mRNA. It is preceded by a TATA-box 30/31 bp upstream, itself surrounded by the upstream (u)BRE and the downstream (d)BRE. MTE and DPE are located downstream the Inr +28 to +33 bp. DCE and XCPE1 can be present in core promoters. TATA, Inr, MTE and DPE are sufficient to initiate transcription.

III. Transcriptional events

1.Central dogma of molecular biology

The model for the central dogma of molecular biology, first enunciated by Crick in 1971, assumed that one specific DNA sequence, called gene, is transcribed by a DNA-dependent RNA polymerase to form the intermediate and complementary mRNA, which in turn is translated into one specific protein **(Figure 16)**. With the advent of high-scale genomic techniques, this simple model has turned into a highly complex multi-stepped process, regulated by several factors. Moreover, different subsets of genes were demonstrated to be transcribed by different RNA polymerases. Indeed, the cellular RNA Polymerase I is involved in the ribosomal RNA (rRNA) transcription, RNA polymerase III is transcribing transfer RNA (tRNA), 5S rRNA and other small non-coding RNA (ncRNA). Finally, protein-coding genes, and many non-coding genes, are transcribed by RNA Polymerase II (Pol II).

RNA Pol II-dependent transcription controls major cellular functions such as differentiation, identity maintenance and responses to changes in cell environment (Sainsbury et al., 2015). RNA Pol II is composed of 12 subunits, among which RPB1 is bearing the C-terminal domain (CTD). Pol II CTD consists in 52 repeats of the hepta-peptide (Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7) that can be phosphorylated at several sites. The most studied are phosphorylation on Ser2 and Ser5. However, Tyr1,Thr4 and Ser7 can be phosphorylated, though, their phosphorylation occurs at a frequency 100-fold lower than Ser2 and Ser5 (Suh et al., 2016). RNA Pol II transcription is regulated by thousands of protein and RNA factors responding to particular physiological, environmental and developmental signals, considered as direct factors or cogs in the transcription machinery. These factors interact to and around the target DNA sequence to determine which region and at what level have to be transcribed (Fuda et al., 2009).

The DNA sequence recognized by the transcription machinery is composed of three elements: i) the core promoter, which is the place of preinitiation complex (PIC) assembly, containing the general transcription factors (GTF) and the RNA Pol II; ii) the core promoter proximal region and iii) the distal enhancers (Fuda et al., 2009).

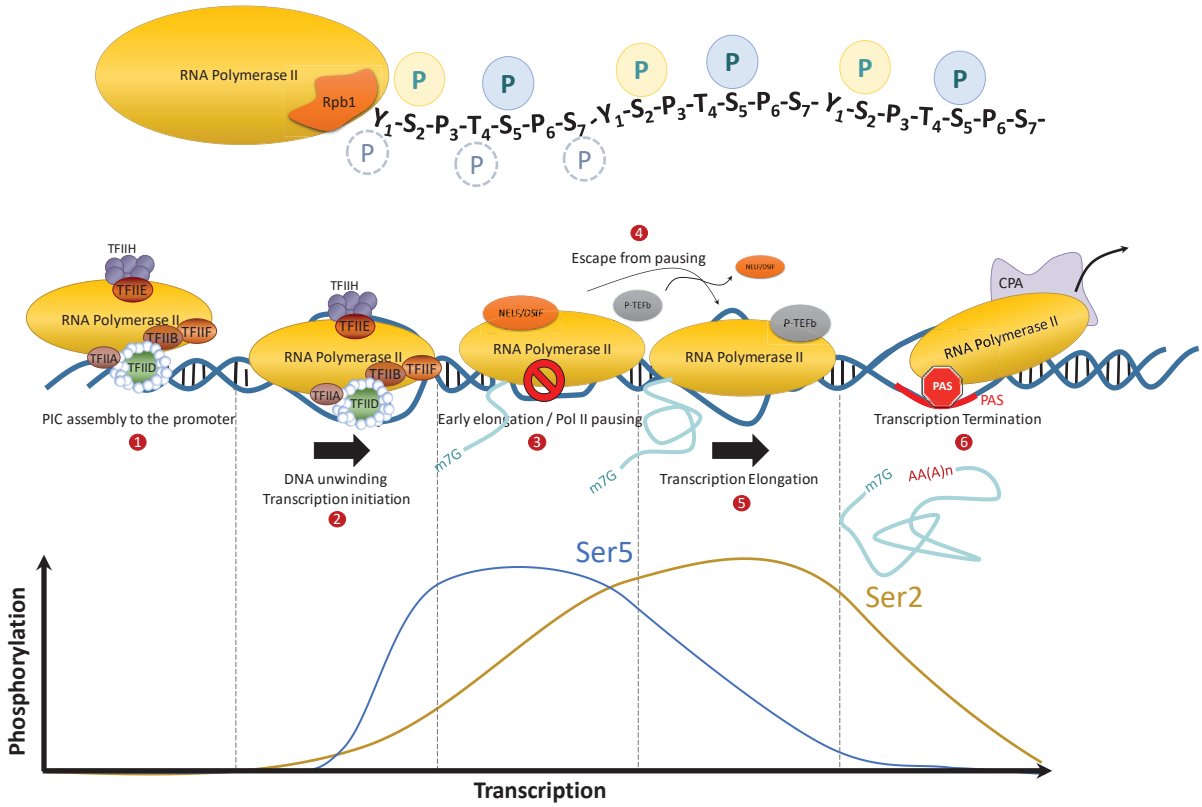


Figure 18 : Major steps of transcription process

A. RNA Polymerase II is composed of 12 subunits among which Rpb1 bears the heptameric repeat YSPTSPS forming Pol II C-terminal domain (CTD) that can undergo post-translational modifications (PTM). Full circles represent the well described CTD Phosphorylation sites (P), dashed circles represent suggested CTD P sites.

B. Transcription process can be divided in seven steps. Step 1 corresponds to preinitiation complex assembly (GTF + pol II) on the promoter. Pic assembly leads to DNA unwinding and transcription initiation (Step2). The early elongation is stopped by Pol II pausing in a promoter-proximal region, serving as 5' checkpoint (Step3). After this crucial stalling step, Pol II is hyperphosphorylated by P-TEFb and escapes from pausing (Step 4) to initiate a productive elongation (Step 5). At the polyA site (PAS), transcription ends and Pol II is released from the DNA template (Step 6).

C. Dynamics of the phosphorylation status of CTD Serine 5 and Serine 2 along with transcription process.

The core promoter is composed of different sequence motifs **(Figure 17)**. First, the initiator (Inr), encompasses the TSS, considered as the "+1" position of transcription initiation. Its conserved sequence, YYANWYY, bears the TSS in the "A" position. Then, TATA-box is located 30/31 bp upstream the TSS, and harbors the TATAWAAR consensus sequence. Downstream core promoter element (DPE) and Motif ten element (MTE) are located +28 to +33 bp downstream of the TSS. Spacing between both of these elements and the TSS is mandatory for a correct transcription initiation. The concomitant presence of TATA-box, Inr, MTE and DPE is sufficient to initiate transcription (Juven-Gershon et al., 2008). Besides, some other motifs can be found in core promoter and differentially associate to participate to promoter specificity. For example, the TATA-box can be immediately surrounded by the TFIIB recognition element (BRE), which can locate upstream (u)BRE or downstream (d)BRE and determines transcription directionality. Finally, downstream core element (DCE) and X core promoter element I (XCPE1) can be present in the core promoter. These motifs, and more generally the core promoter, recruit the general transcription factors (GTF) and the RNA Pol II.

The core proximal and the distal enhancer regions are involved in the recruitment of specific transcription factors (TFs), called activators and repressors. Specific TFs interact with multi-protein complexes termed co-regulators and regulate transcription mostly by nucleosome organization and chromatin structure modifications that in turn lead to specific transcription factor association and RNA Pol II CTD phosphorylation status (Juven-Gershon et al., 2008).

The transcription cycle is composed of six major steps: i) RNA Pol II association and PIC assembly to the promoter; ii) DNA unwinding and transcription initiation; iii) early transcription elongation and Pol II pausing by gripping both template DNA and nascent RNA, driving the clearing of PIC from the core promoter; iv) hyperphosphorylation of Pol II complex and escape of the enzyme from the pausing region; v) transcription elongation through the entire gene body; vi) transcription termination; (Fuda et al., 2009) **(Figure 18).**

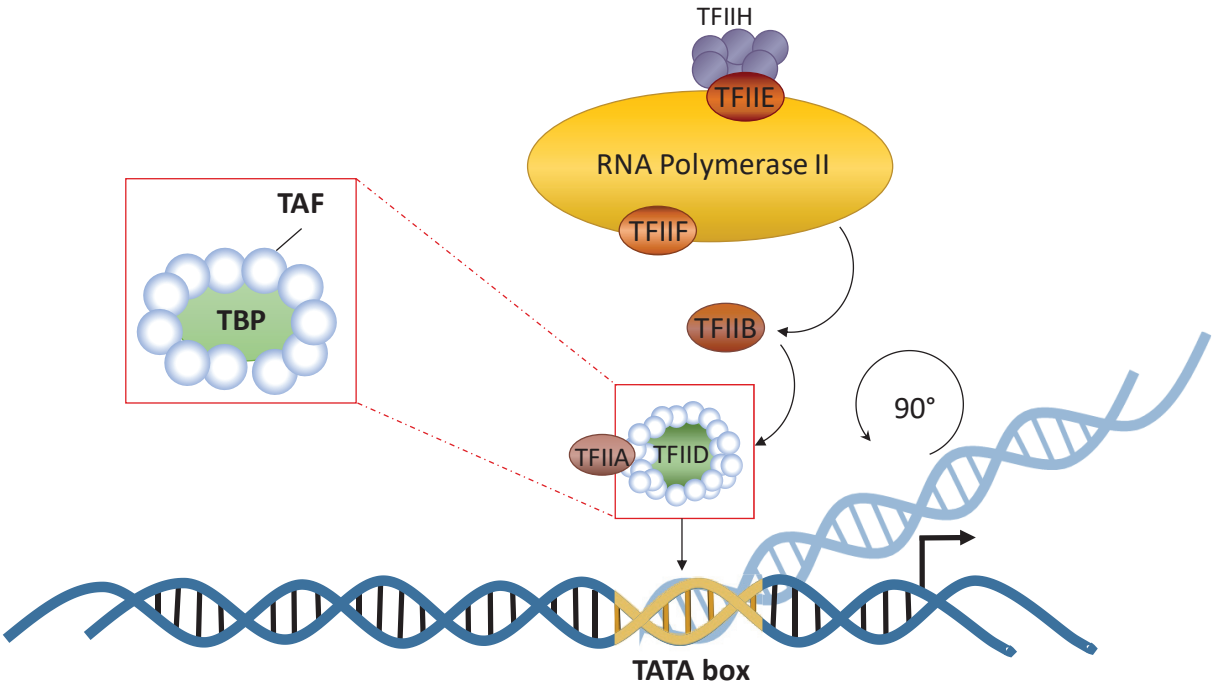


Figure 19 : Preinitiation complex assembly

PIC assembles to the promoter to initiate the transcription. TFIID and TFFIA first associate to the TATA-box and recruit TFIIB. TFIIB is the intermediate of recruitment for RNA pol II and TFIIF that further recruit TFIIE and TFIIH. "TFIIX" factors are the seven so-called General Transcription Factors (GTF). TFIID assembly to the promoter drives a DNA bending of around 90° that will further wrap around the RNA pol II. Once PIC is fully associated to the promoter, promoter sequence is unwound and the transcription bubble is formed, allowing RNA pol II to start transcribing.

2. Initiation

Transcription initiation is primed by the recruitment of the preinitiation complex to the promoter. PIC is composed of general transcription factors (GTF) together with the RNA Pol II. GTF comprise five RNA Pol II specific multi-protein complexes called: TFIIB, TFIID, TDIIE, TFIIF and TFIIH. PIC assembles through a classical pathway, which is detailed below **(Figure 19)**. TFIID complex is composed of the TATA-box binding protein (TBP) and 13 TBP-associated factors (TAF). While TBP is required for all the promoters, the different TAF insure promoter specificity for the complex recruitment. Altogether, these factors form the first assembled complex of PIC, leading to the bending of DNA promoter sequence of around 90° that will drive DNA wrapping around the further recruited RNA Pol II. Noteworthy, as only 10-20% of the promoters harbor the TATA-box motif, TFIID recognizes promoters principally through sequence architecture similarity to the consensual TATA-box sequence. TFIID further serves as a hub for the recruitment of TFIIA and TFIIB. These two factors are necessary for TFIID stabilization and TSS recognition. Moreover, they recruit RNA Pol II, together with TFIIF that also participate at TSS recognition. Next, TFIIE recruitment forms a bridge between RNA Pol II and the further associated TFIIH promoting DNA opening through TFIIH core subunit. TFIIH also contains a DNA-dependent ATPase activity critical for transcription initiation. Indeed, TFIIH triggers the phosphorylation of RNA Pol II-CTD Ser5 and the following phosphoSer5-mediated Pol II escape from the core promoter (Svejstrup, 2012). In parallel, TFIIF can initiate initiation by stimulating phosphodiester bond formation and thus RNA synthesis. The fully assembled PIC can now adopt a “close state” where the promoter is still entirely dsDNA, or an “open” state that requires ATP, where 15 bp of the promoter sequence are unwound, forming the so-called transcription bubble. This opening leads to the correct TSS selection and to RNA synthesis with the first phosphodiester bond formation (Sainsbury et al., 2015). At this step, the presence of (u)BRE or (d)BRE element is determinant for directionality of the transcription. During priming and transcription initiation, PIC undergoes several transformations, until the nascent RNA reaches a length of 25 nt, thereafter a stable elongation complex is formed. Noteworthy, during transcription initiation step, Pol II CTD is unphosphorylated (Nechaev and Adelman, 2011).

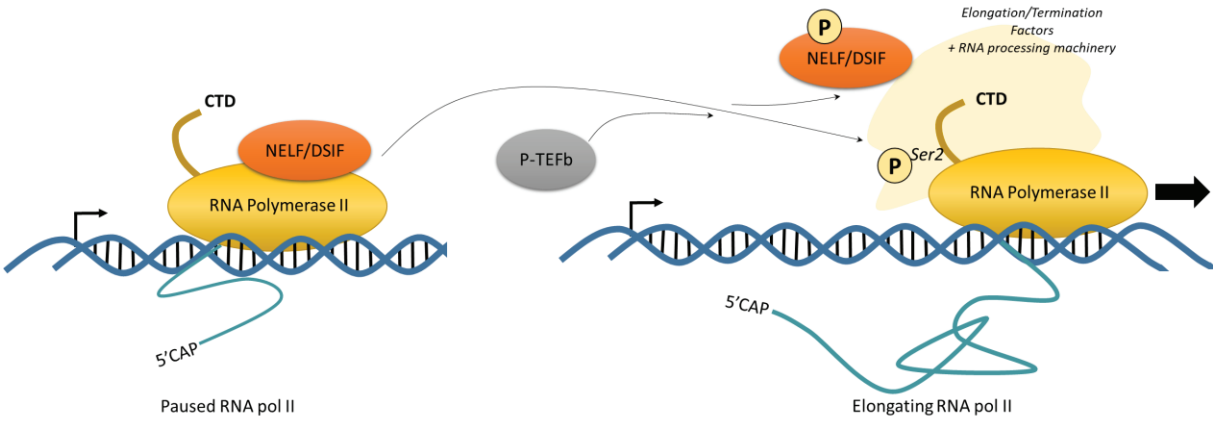


Figure 20 : Transcription elongation

RNA pol II is associated with NELF and DSIF that drive its pausing to a promoter-proximal region, leading to a 5'checkpoint for proper capping of the nascent RNA chain. P-TEFb is next recruited, and phosphorylates both NELF/DSIF to remove them from RNA pol II, and pol II-CTC-Serine 2. These phosphorylations trigger RNA pol II escape from the pausing site and the beginning of the productive elongation that can have different rates according to the transcribed gene. PhosphoSer2 serves as a hub for elongation and termination factors, and for machineries involved in RNA processing, travelling along the gene with the RNA pol II.

3. Elongation

When the neosynthesized RNA attains around 20-25 nt length, capping machinery insures the 5'end capping to protect the nascent RNA chain (*see below "Capping"*). At this point, GTF are released from the core-promoter. This early elongation step is a slow and non-productive process. After transcription initiation, RNA Pol II pauses and accumulates within 30-60 nts downstream of TSS and reaches the so-called promoter-proximal pause site. This 5' RNA Pol II pausing is thought to allow Pol II modification before productive elongation, and checkpoint for proper 5'capping (Jonkers and Lis, 2015). Pol II stalling in the core-proximal regions depends on core promoter features, notably specific TF and the negative elongation factor (NELF) in complex with DRB sensitivity inducing factor (DSIF). The protein complex NELF/DSIF stabilizes Pol II. Nucleosomes may also contribute to Pol II pausing by acting as a road block for Pol II progression. Pol II pausing at this step is a key step in transcriptional regulation to insure proper further elongation step. Indeed, Pol II pausing offers a wide time frame window for interaction and coordination of crucial elongation processes (Mayer et al., 2017).

After pausing, RNA Pol II is released from the promoter-proximal pause site by P-TEFb (positive transcription elongation factor b) composed of the cyclin T1 and the cyclin-dependent kinase (CDK) 9. P-TEFb is recruited to the promoter by specific TF and co-factors and phosphorylates both Serine 2 of RNA Pol II-CTD and NEFL/DSIF, leading to their eviction from RNA Pol II (Jonkers and Lis, 2015). This removal initiates the productive elongation step, which occurs all along the gene body, until termination, with different elongation rates modulated (until 3-fold) in a gene-dependent manner **(Figure 20)**

Pol II elongation rate can be regulated by several factors, such as: i) histone marks that loosen or tighten DNA around nucleosome (e.g. H3K56Ac, H3K79me2); ii) elongation factors (e.g. Elongin and TFIIS); iii) histone chaperones and nucleosome remodelers; iv) DNA sequence influencing DNA topology (G-rich sequences are more difficult to transcribe because more stable); v) gene feature, such as exons that have a strong negative effect on elongation rate

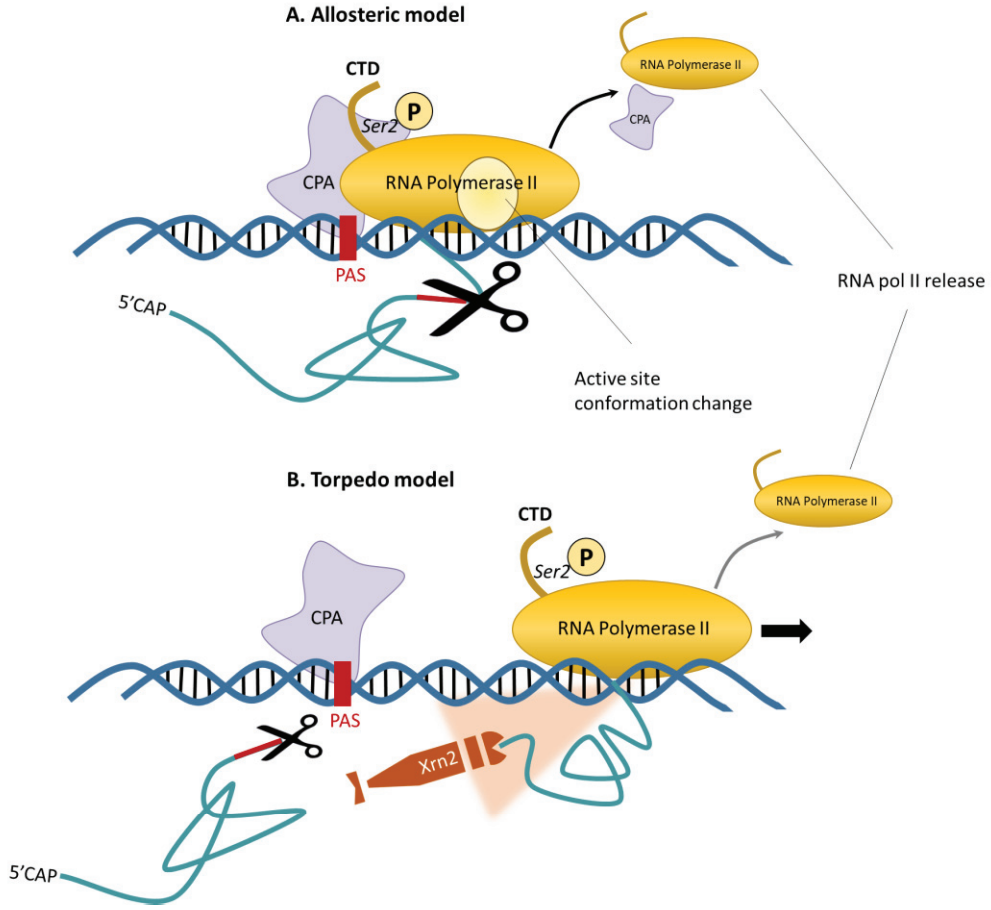


Figure 21 : Transcription termination

At the PAS, transcription ends. Transcription termination is proposed to occur according to two models.

A. The allosteric model assumes that pol II recognizes the PAS with the already CTD-recruited Cleavage and PolyAdenylation factor (CPA). Pausing at this site leads to active site conformational changes and RNA pol II release.

B. The torpedo model implies that, after cleavage of the neosynthesized RNA chain by CPA, RNA pol II continues transcribing a downstream RNA chain. The 5'-3' exonuclease Xrn2 digests the downstream transcript until colliding with RNA pol II, leading to its resolving from DNA template.

(due to their high nucleosomal occupancy slowing down the RNA Pol II crossing); and finally vi) chromatin tridimensional organization, notably involving mediators and cohesion complexes. Interestingly, Pol II elongation rates influences co-transcriptional processes, such as splicing and termination (Jonkers and Lis, 2015). Productive elongation is characterized by the progressive loss of phosphorylation of serine 5 and the phosphorylation of serine 2 on Pol II CTD (Svejstrup, 2012), which generates a platform for different factors that will "travel" with the Pol II along the gene and that regulate the different subsequent transcription steps (Nechaev and Adelman, 2011).

4. Termination

Transcription termination of protein-coding genes is a highly regulated process that can occur at different positions along the gene, leading to different regulatory 3' UTR length, or different encoded proteins. Similarly to the 5' checkpoint involving promoter-proximal RNA Pol II pausing, a 3' checkpoint insures the control of transcription termination, preventing the synthesis of erroneous mRNA. The 3' checkpoint is also regulated by CDK9 activity (Proudfoot, 2016).

Transcription termination occurs according to a mechanism dependent on the presence of a polyadenylation sequence (PAS) that leads to the reuse of the RNA Pol II for another round of transcription. During transcription termination, Pol II slows down nearby the PAS. Pol II pausing is induced by different factors. When it crosses the PAS, Cleavage and PolyAdenylation factors (CPA), notably CPSF and CstF, are recruited to Pol II-CTD consequently triggering RNA 3'-end processing and transcriptional termination. Moreover, chromatin structure itself is able to influence Pol II slowdown, especially the presence of core nucleosomes. Finally, one major mechanism implicated in Pol II deceleration is the hybridization between the nascent RNA and the template DNA, which is outside of the elongation complex. This DNA is unwound and nucleosome-free. RNA:DNA hybrids, also called R-loops, are thermodynamically more stable than dsDNA. It is important to note that R-loop accumulation results in increased DNA damages due to the mutagenic nature of the ssDNA. In mammals, Senatoxin is a protein that resolves R-loop structures. Senatoxin also directly promotes transcription termination (Proudfoot, 2016).

While RNA Pol II slows down, the termination process happens. To do so, the preceding Pol II-CTD-recruited CPA complex assembles on the pre-mRNA PAS that is extruding from the Pol II. Two different models, yet not mutually exclusive (Eaton et al., 2020), are proposed for transcription termination:

- *The allosteric model*: for this model, Pol II senses the PAS, helped by the associated CPA on its CTD. Sensing the PAS is sufficient to induce Pol II pausing and subsequent active site conformational changes triggering Pol II release. Evidences for this model suggest that the RNA Pol II release from the template DNA and the cleavage at the PAS are two independent mechanisms, occurring in subsequent steps. (Zhang et al., 2015) **(Figure 21-A).**

- *The torpedo model*: this model assumes that the transcript is still synthesized even downstream of the PAS by the Pol II, after cleavage of the upstream RNA chain in the PAS. Next, the 5'-3' exonuclease Xrn2 progressively digests the downstream RNA chain in processivity competition with the still elongating Pol II. When Xrn2 reaches Pol II, it creates a molecular trigger for Pol II release. Evidence for this model is that R-loop-mediated Pol II pausing enhances Xrn2-dependent termination (Proudfoot, 2016) **(Figure 21-B)**.

A non-productive transcription termination process can occur upon DNA damages due to UV or chemical exposure, or oxidation. It relies on irreversible RNA Pol II association to DNA template that has to be removed by proteolytic degradation.

Interestingly, most of Pol II-transcribed genes exhibit a distinct 3' terminus pattern that runs across multiple kilobases downstream the PAS site. This 3' readthrough can connect two neighboring genes, or can arise from fusion between two distal genes. ENCODE (The Encyclopedia of DNA Elements) identified 65% of human gene-derived RNA resulting from gene-fusion, called chimeric RNAs (He et al., 2018). This phenomenon can have consequences. Indeed, a non-stopped RNA Pol II can reach the promoter of a downstream gene, driving its repression, by a process named transcription interference. Moreover, RNA Pol II complexes transcribing two convergent genes can collide and lead to down-regulation of gene expression. Finally, if a transcribing Pol II attains a DNA replication region, Pol II and DNA polymerase can encounter, disrupting DNA synthesis and leading to genome instability

(Proudfoot, 2016). Readthrough is found more abundant in level and length under stress conditions (Vilborg et al., 2017).

5.Chromatin topology

In the nucleus, genome is spatially organized, in a non-random manner, at different levels of hierarchy. For a long time, genome structure has been studied separately from gene expression regulation. Nonetheless, many evidences demonstrate inter-connection and inter-regulation between genome organization and gene expression (Cavalli and Misteli, 2013). The spatial organization of chromosomes and genes is variable according to cell types, and modulated by physiological processes such as differentiation, development and by pathologies. Noteworthy, the nuclear organization is heterogeneous between cells, along with transcription process that also varies between cells (Ren et al., 2017) **(Figure 22)**.

First, the chromatin fiber, composed of DNA wrapped around histones, can fold in higher-order structures called chromatin domains. These domains can be partitioned in “repressive” and “open” domains, in accordance with their transcriptional state. Only three repressive domains have been identified, being polycomb-associated euchromatin, heterochromatin and chromatin with no strong enrichment for positive factors. Repressed regions, enriched for the repressive mark H3K27me3, are localized in the nuclear periphery. Contrarily, gene-rich and transcriptionally active regions, enriched for the H3K4me3 active histone mark, are more internal in the nucleus. Although precise subclasses are hard to define, four major groups of open chromatin are characterized: enhancers, promoters, transcribed regions and insulator-protein-associated regions. Interestingly, in chromatin domains, not all genes are in the same transcriptional state. Indeed, various levels of activity are supported in a gene-dependent manner, conserving however a more or less favorable environment for gene expression depending on the chromatin compartment (Cavalli and Misteli, 2013).

During gene expression regulation, and notably during transcription, intra-chromosomal interactions play a critical role through DNA-looping. Four different loops with specific functional consequences have been described **(Figure 22. A-D)**. 5’ end and termination site of a transcribed gene can drive DNA looping to efficiently recycle RNA Pol II and promote transcription directionality. Moreover, distal enhancer can contact gene promoter by long-range DNA looping. These interactions are subjected to regulatory changes that modulate

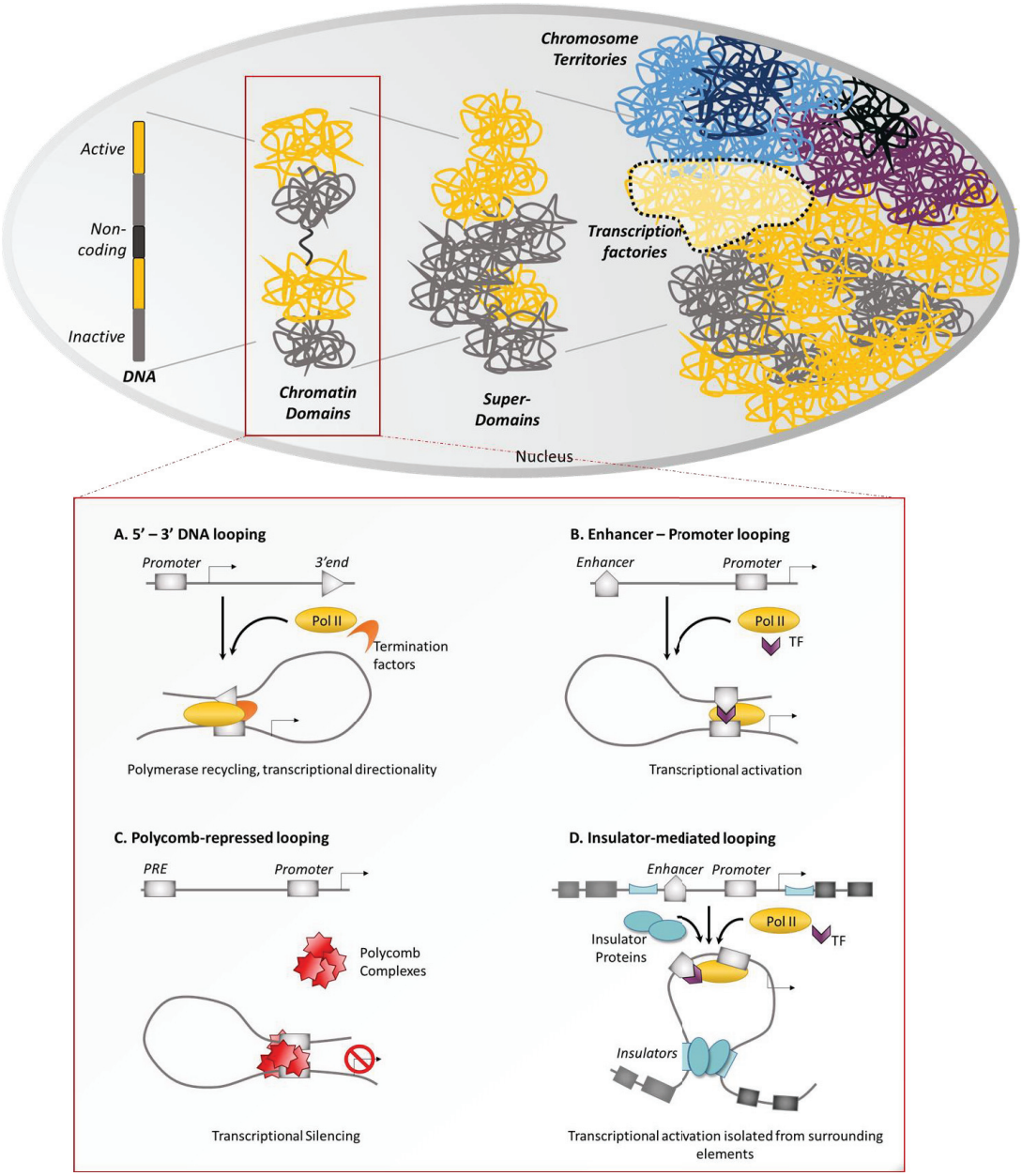


Figure 22 : Genome organization and DNA looping involved in transcription regulation

Active and inactive DNA regions organize in higher-order structures called chromatin domains that can further arrange in super domains. The last level of spatial organization is called chromosome territories, in which chromosomes preferentially organize. In this last level of organization, certain regions are more prompt for active transcription processes and are called transcription factories.

In chromatin domains, transcription can be regulated by DNA looping.
A. 5'-3'end DNA looping favors RNA pol II recycling by connecting both extremities of the coding sequence.
B. Enhancer can be enclosed to the promoter by DNA looping to approach specific transcription factors (TF) to the promoter and initiate transcription.
C. Polycomb-repressed DNA looping negatively controls transcription by polycomb complex deposition on DNA regulatory elements.
D. Insulator-mediated DNA looping allows a differential regulation of neighboring genes.

contacts between different enhancers and promoters to insure proper transcription regulation. Moreover, these contacts also promote transcription activation by negative histone mark removal. On the contrary, gene silencing can be regulated by polycomb-dependent repression that reaches distal promoter *via* DNA loops. This kind of loop can also exist for other regulatory proteins. Eventually, insulator-binding proteins can form DNA loops. For example, CCCTC-Binding Factor (CTCF) and cohesin proteins isolate differentially regulated regions (Cavalli and Misteli, 2013). 99% of CTCF interacting sites have been demonstrated to be co-occupied by cohesin. Moreover, CTCF foci are spatially associated with the so-called RNA Pol II transcription factories (Cavalli and Misteli, 2013). These factories consist in nuclear sites enriched for RNA polymerase and transcription factors that are the places where transcription occurs. These observations support the fact that CTCF-mediated loops are associated with transcription regulation (Tang et al., 2015). Importantly, many CTCF binding sites are encompassed by enhancers that are still capable to make contact with promoters. Remarkably, CTCF and cohesin seem not to be the exclusive insulator proteins as their knock-down does not drastically impact gene expression (Cavalli and Misteli, 2013).

Development of new technologies, such as (circular) chromosome conformation capture (4C and 3C) and High chromosome contact mapping (Hi-C), allow the identification of long-range interactions across the genome. Clustering of chromatin domains by looping through distal interactions forms super domains that further assemble into chromosome domains. Inter-chromosomal contacts are also implicated in gene expression regulation. Notably, long-range contacts between different genes or chromosomes may be mediated by TFs to co-activate contacting genes (Cavalli and Misteli, 2013).

Another layer of organization is represented by the so-called Topologically Associated Domains (TAD) harboring sharp boundaries defined by CTCF binding sites, or delimited by TSS. Long-range interactions among TAD influence the active or inactive TAD characteristics (Cavalli and Misteli, 2013).

Chromosome territories constitute the last order of chromatin organization, in which chromosomes preferentially occupy defined nuclear regions of around 2-4 µm diameter (Cavalli and Misteli, 2013).

More generally, chromatin topology seems to affect transcription regulation through two divergent effects: the accessible surface area that favors contact between DNA binding sites and proper regulatory factors, and the local density of reaction, depending on the crowded state of the regulated area (Almassalha et al., 2017).

6. HBV transcriptional regulation

HBV is the smallest DNA virus described and harbors a highly compact genome. The only viral enzyme of HBV is the DNA polymerase/retro-transcriptase, involved in rcDNA synthesis and effective virion production. However, in order to be efficiently transcribed, HBV hijacks the host cell machinery by harboring similar features to cellular genes, required for an effective transcription. Cell-free experiments demonstrated that α-amanitin treatments inhibited the viral transcription, highlighting that RNA Pol II is the cellular RNA polymerase in charge of the viral transcription (Rall et al., 1983). Moreover, these experiments revealed for the first time the presence of HBsAg and HBcAg promoters, harboring TATA-like structures. Later, these two promoters and two others were identified more precisely (Moolla et al., 2002).

HBV transcription is regulated by *cis*-regulatory elements, located in the genome, and embedded within protein coding sequences, and by *trans*-regulatory elements that are liver-specific and ubiquitous transcription factors (TFs). HBV transcription is under the control of four well-described promoters, whose activity is further regulated by two enhancers. Except for the preS1 promoter, HBV promoters do not contain the classical TATA-box sequence. Regulatory sequences in HBV genome possess binding sites for different cellular factors that are involved in proper viral transcription activation or repression **(Figure 23)**.

A. The core promoter

PreCore and pgRNA transcription is regulated by the core promoter, which is 232 bp long and divided in two elements. The upper regulatory region (URR) is located from nt 1613 to nt 1742, and the basic core promoter (BCP), located from nt 1742 to nt 1849. Choice in preC or pregenomic TSS is not liver-specific and does not involved URR element. Moreover, core promoter is devoid of TATA-box motif, yet, it harbors A/T-rich regions located 20 to 35 nt upstream of preCore Initiator (Inr) and binds cellular TATA-binding protein (TBP) (Kramvis and Kew, 1999). The BCP is preceded by the core upstream regulatory sequences (CURS) subdivided in CURS-α, CURS-β and CURS-δ. Among these sequences, only CURS-α is required

for BCP activity. BCP is also further upstream preceded by the negative regulatory element (NRE), divided in NRE-α; NRE-δ and NRE-γ that have a synergic repressive activity on viral transcription (Moolla et al., 2002). Although BCP is a weak promoter, it is sufficient to activate preC and pgRNA transcription, preferentially in differentiated hepatocytes, shedding light on the involvement of liver-specific factors. Notably, BCP contains two binding sites for the liver transcription factor HNF4. HNF4 has been demonstrated to be important for preC and pgRNA transcription in chronically infected patients (Zheng et al., 2004). Interestingly, HFN4 binding sites can differentially regulate the versatile transcription of either preC or pgRNA, maybe due to their proximity with both of the TSS. Recently, TARDBP has been demonstrated to associate with the core promoter and to play a role in viral replication (Makokha et al., 2019). Moreover, the histone acetyl transferase NuA4 subunit TIP60 was demonstrated to negatively regulate the core promoter and the consequent 3.5 kb RNA transcription by recruiting BrdD4 to acetylated H4 (Nishitsuji et al., 2018).

B. The surface antigen promoters

The 2.4 kb long RNA, encoding the L-HBsAg, is transcribed in lower abundance than the two 2.1 kb long RNAs encoding M- and S-HBsAg. This ratio is further detected at the protein level, where S- and M- surface proteins are found in higher quantity than the L-surface protein. This differential transcription is explained by the presence of two different promoters. The first one is the preS1 promoter, regulating the 2.4 kb RNA. PreS1 promoter is the only TATA-box-containing HBV promoter, and interacts with HNF1 and TBP. Moreover, preS1 promoter also harbors a negative regulatory sequence located downstream of the preS2 promoter. The preS2 promoter contains two different TSS leading to the transcription of the PreS2 and the S RNAs. This promoter contains seven regulatory elements that can be either positive or negative regulators according to their association with liver-specific and ubiquitous TF. Interestingly, this TATA-less promoter represses preS1 promoter activity and thus 2.4 kb RNA transcription (Moolla et al., 2002).

C. The X promoter

X promoter is located 140 nt upstream of X TSS and contains sites for liver-specific and ubiquitous TFs such as NF1, C/EBP, ATF, AP1/Jun-Fos and p53 (Moolla et al., 2002). Globally, X and Core promoters are more active than preS1 and preS2 promoters. Possibilities of this

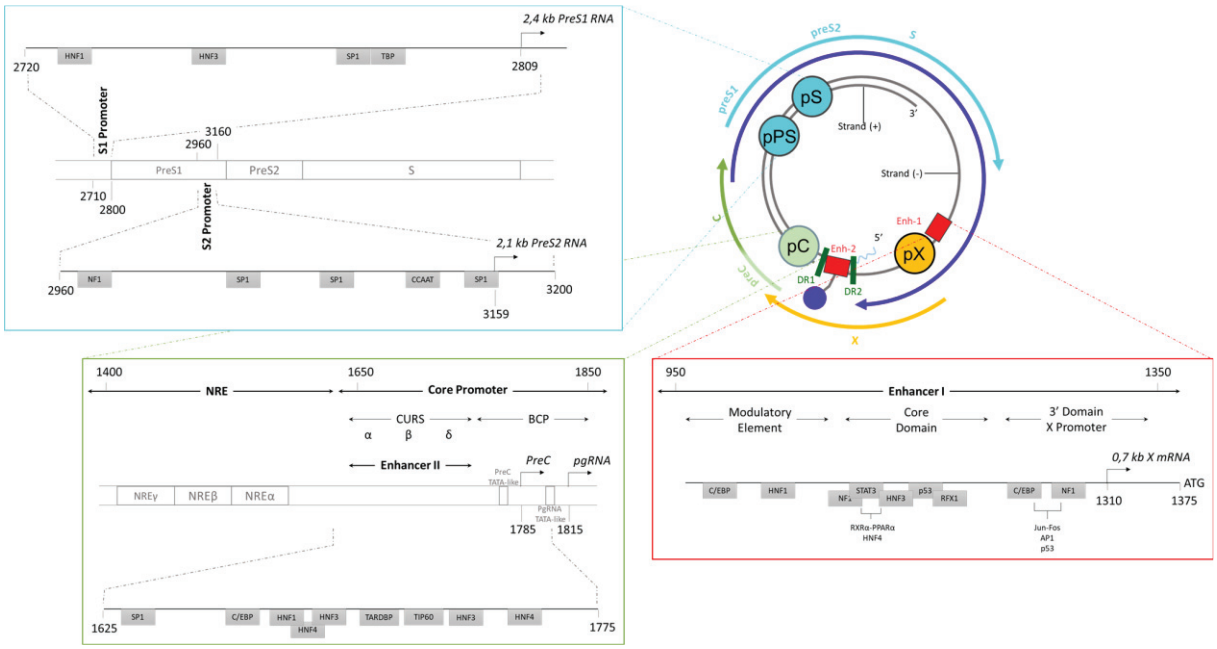


Figure 23 : Features of HBV regulatory elements

HBV genome encodes six mRNAs that are transcriptionally regulated by four promoters and two enhancers. Sequence features and associated cellular factors are represented for Enhancer I (EnhI), Core promoter (CP) and overlapping Enhancer II (EnhII), and the for the surface antigen promoters, S1 (pPS) and S2 (pS).

- *Adapted from Moolla et al, 2002*

differential activation is their ability to associate with several transcription activators and their location in high proximity, if not overlapping, with enhancer sequences I and II, named EnhI and EnhII.

D. The Enhancer I

EnhI favors transcription under control of core and X promoters to enhance the synthesis of preC, pregenomic and X transcripts. EnhI is 120 nt long and is organized in three domains. The 5'end domain is a modulator element able to interact with ubiquitous and liver-specific TF. The central domain, or enhancer core domain, displays a strong influence on the different promoters, notably by associating with HNF4. Finally, the 3'end region overlaps with the X promoter (Bock et al., 2000). Mutations in EnhI can occur during the natural course of HBV infection and impact negatively the viral transcription (Cho et al., 2013).

E. The Enhancer II

EnhII is 148 nt long and is located immediately upstream of BCP and controls the two surface antigen promoters, the X promoter and the BCP (Moolla et al., 2002). Moreover, EnhII associates with HNF1 and HNF3 liver-specific transcription factors (Kramvis and Kew, 1999).

Recently, a precise mapping of different viral TSS has been obtained in HBV-infected liver and HCC patient samples by a technic of Cap Analysis of Gene Expression (CAGE), based on the fact that HBV transcripts are capped in 5'. This experiment allowed the identification of the six major already well-described TSS, but also the uncovering of 11 supplemental TSS, among which two are in an anti-sense orientation (Altinel et al., 2016).

Proximity between EnhI and X promoter, and EnhII and core promoter easily explains their direct regulatory role in X and preC/pgRNA transcription. However, these two enhancers also regulate the transcriptional activity of the two further preS1 and preS2 promoters, located several hundred bp downstream in HBV genome. As demonstrated for cellular transcription regulation, even though being distally located, enhancers and promoters can physically interact *via* DNA looping to initiate transcription (Cavalli and Misteli, 2013). These intra-chromosomal interactions could be part of explanation on how HBV regulatory sequences act from distal sites. However, the average length of so far identified DNA loops in human genome is 100 kb, thus raising the question if DNA looping within a highly compacted and short 3.2 kb-long genome could happen and could be technically identified. Indeed, chromosome contact

technics, such as Hi-C, 3C and 4C, which could help answer this question achieve a resolution of 3-4 kb (Raviram et al., 2014). Resolution improvement for these technics would allow a better understanding on HBV intra-chromosomal transcriptional regulation.

Interestingly, these cutting-edge technics have been used to study HBV genome localization and potential inter-chromosomal long-range interactions with the host genome. Enrichment for HBV genome in transcriptionally active chromatin regions was demonstrated. Contacts with active gene transcription loci correlates with the presence of HBx (Hensel et al., 2018), but does not require the presence of HBx, as HBx-negative virus genomes are still mainly present in the active compartment (Moreau et al., 2018). Giving that HBx is essential for cccDNA transcriptional activity, the observation that HBx-negative cccDNA are present in chromatin active compartments raises the question about the biological relevance for cccDNA to be present in such compartments.

IV. Co-transcriptional events

Synthesis of mRNA is nuclear and physically separated from translation that occurs in the cytoplasm. Transcription is the process that leads to pre-mRNA synthesis. However, neosynthesized RNA is not a suitable template for protein translation and has to be further processed through different mechanisms. Moreover, mature mRNA has to be exported from the nucleus to reach the cytoplasm to be addressed to the translation machinery. Thereby, pre-mRNA undergoes series of maturation steps in a co-transcriptional way to acquire a 5' cap, to get rid of intronic sequences and be alternatively spliced in some circumstances, and to be lengthened with a polyadenylation tail. Moreover, other modifications such as methylation can be added to the neotranscribed RNA chain that implement a supplementary regulatory layer. All these steps constitute the so-called RNA processing and are inter-connected and inter-regulated. Remarkably, the mature mRNA is coated and interacts with several proteins belonging to the different RNA processing machineries, and could be considered as a messenger ribocucleoprotein (mRNP).

1.Capping

Capping in 5' end of the neosynthesized RNA is the first modification received by the transcript and occurs when transcription initiation has transcribed around 25-30 nucleotides. 5'cap is formed of a N7-methylated guanosine (m7G) associated with the first 5' nucleotide of the RNA chain. 5' m7G cap is highly conserved in eukaryotes and is generated by three enzymatic reactions. The first step consists in the removal of the γ-phosphate from the 5' end of the nascent RNA to generate a 5' diphosphate RNA by the RNA triphosphatase (TPase). This new extremity then supports the covalent bound to a guanosine monophosphate in a 5'-5' linkage, catalyzed by an RNA guanylyltransferase (GTase). The last step is the addition of a methyl group to the N7 amine group of the guanine cap by the guanine N7-methyltransferase (Guanine N7 MTase) (Ramanathan et al., 2016). Capping is a co-transcriptional event as the three capping enzymes are recruited on the Pol II-CTD, notably *via* a direct interaction between GTase and MTase with pol II-CTD phosphoserine 5, phosphorylated during the transcription initiation step (Ghosh and Lima, 2010). A 5' capping quality control is insured by other co-transcriptional RNA processing mechanisms, such as pre-mRNA splicing and

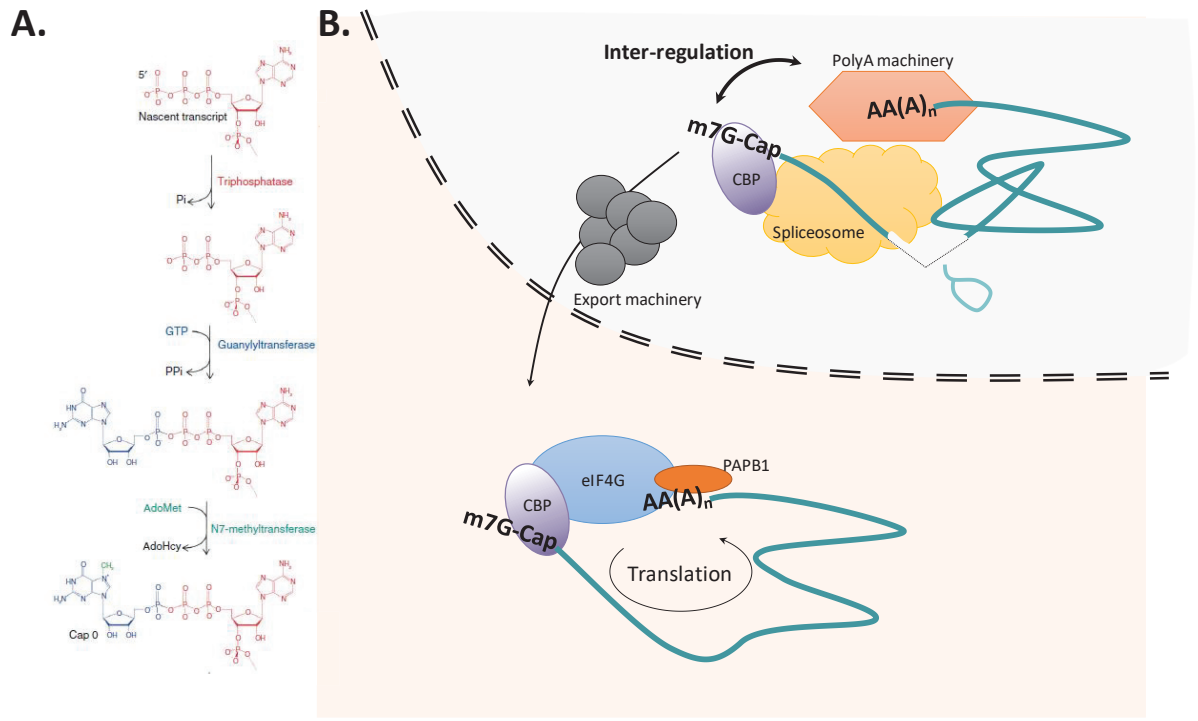


Figure 24 : mRNA capping

A. The addition of a N7-methylated guanosine is a 3-steps process mediated by three different enzymes, namely a triphosphatase, a guanylyltransferase and a mathyltrasnferase.

B. The m7G cap structure added in 5' of the neosynthesized pre-mRNA interacts with the polyA, the spliceosome and the export machineries *via* the Cap binding protein (CBP). In the cytoplasm, m7G cap is involved in translation regulation, notably by mediating mRNA circularization through the interaction of the cap and the polyA tail *via* eIFG4 connecting CBP and polyA binding protein PABP.

polyadenylation (See below *"Splicing"* and *"3' end processing*"). Moreover, DXO, a decapping, pyrophosphohydrolase and a 5'-3' exonuclease, is demonstrated to degrade unmethylated capped mRNA. Knockdown of DXO increases the accumulation of immature pre-mRNA in the nucleus, and strikingly affects splicing and leads to polyA defective cleavage. These observations, together with the involvement of splicing and polyadenylation in 5' capping quality control highlight a strong link between the different co-transcriptional events **(Figure 24)**.

Capping is firstly involved in RNA protection against 5'-3' exonuclease digestion and, thus, in RNA stability. For example, Dcp2 can remove the cap to generate a 5' monophosphorylated terminated RNA that can be in turn digested by 5'-3' exonucleases, such as Xrn1 (Ghosh and Lima, 2010). Furthermore, 5' cap is associated with the capping binding protein (CBP) that can interact with many cellular proteins and confer different cellular functions to this 5' modification. Notably, through CBP, m7G cap is in complex with U4/U6-U5 subunits of the spliceosome. CBP also contributes to the 3' end processing. Interestingly, deletion of CBP abolishes the endonucleolytic cleavage constituting the first step of polyadenylation process. In the nucleus, CBP interacts with RNA export factors such as Aly, and drives the way to nuclear export (Ramanathan et al., 2016). Afterward, in the cytoplasmic compartment, CBP recruits the translation initiation complex eIF4G that further enrolls the two ribosomal subunits to initiate translation. While eIF4G is associated to CBP in the mRNA 5'end, it also interacts with the polyA tail *via* the polyadenylation binding protein 1 (PAPB1) and rearranges mRNA structure into a pseudo-circular shape that favors ribosome processivity and thus translation efficiency (Ramanathan et al., 2016).

To escape immunity, viruses have adapted their transcripts to be capped as cellular mRNAs. Viruses can either synthesize their own cap (e.g. Alphaviruses, Rhabdoviruses, among others) or hijack 5' cap from cellular mRNA, a mechanism called "cap snatching". This is the case for influenza virus for example. Interestingly, in order to perpetuate the recognition of foreign RNA together with preventing self-reactivity, immune cells possess mRNA with a supplementary 2' O-methylation in their m7G cap (Ramanathan et al., 2016).

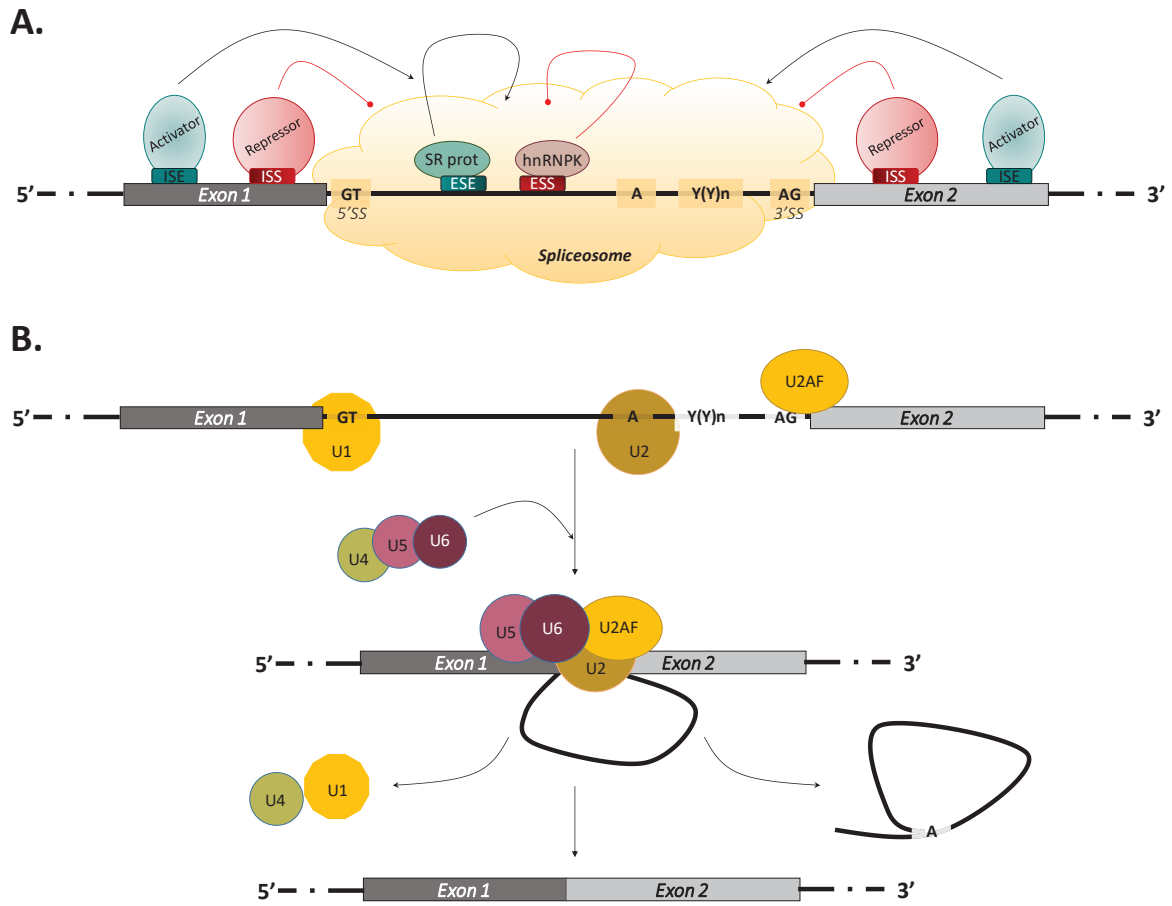


Figure 25 : pre-mRNA splicing

A. The neosynthesized pre-mRNA is composed of intronic and exonic sequences. Splicing signals are represented as yellow boxes in the intron. The *cis*-regulatory splicing elements, namely ISE/ISS and ESE/ESS, recruit specific activators or inhibitors that modulate the splicing process.

B. Removal of the intron (or alternative exon) occurs via the assembly of the spliceosome, composed of 5 factors: U1, U2, U4, U5 and U6 that are sequentially recruited to form a mature splicing machinery that will lead to intron (or alternative exon) skipping by two trans-esterification reactions.

2. Splicing

In 1970, coding sequences were described to be discontinuous and the process of removing non-coding in-between sequences was named “splicing”. Coding sequences were called exons whereas non-coding sequences were called introns. Later, consensus sequences were identified at the borders of intronic sequences, driving the splicing machinery for proper pre-mRNA splicing. Introns are flanked by a 5’ splice site (5’-C/A-A-G-G-T-A/G-A-G-T-3’), 3’ GT being the two first intron nucleotides, and by a 3’ splice site (5’-C/T-A-G-3’), AG being the two last intron nucleotides. In the last 3’ 40 nts of the intron, an adenosine residue constitutes the so-called “branch point”, crucial for the first splicing step, and a downstream polypyrimidine tract (Baralle and Baralle, 2018). All these splice motifs direct the splicing machinery assembly, also called “spliceosome”.

The spliceosome contains more than 100 components, and a ribozyme catalytic center, mainly driven by the U6 snRNA (Lee and Rio, 2015). The spliceosome assembles *de novo* onto the neosynthesized pre-mRNA without any pre-formed complex. It is composed of the NineTeen Complex (NTC) and five small nuclear ribonucleoproteins (snRNP) containing U-rich snRNA, namely U1, U2, U4, U5 and U6. Each snRNA is associated with a ring of seven Sm proteins. Spliceosome-mediated splicing process can be divided in six steps. The first step is the association of U1 snRNP to the 5’ splice site. Then, U2 snRNP is recruited to the branch point, together with U2 auxiliary factor U2AF recruitment onto the polypyrimidine tract. The third step is the recruitment of the tri-snRNP U4/U5/U6 (or pre-B complex) immediately followed by the removal of U1 and U4. At this step, the spliceosome (or B complex) enters its catalytically active conformation (or B^{act} complex) and performs the two subsequent trans-esterification reactions that lead to lariat dissociation and exon-exon ligation (Dvinge, 2018) **(Figure 25)**.

The splicing can either be constitutive, with the systematic exclusion of intronic sequences and exonic inclusion in the mRNA, or alternative by differential joining of 5’ and 3’ splice sites. Alternative splicing allows the generation of multiple RNAs, coding multiple protein isoforms with different functions. Alternative splicing regulation is mainly tissue-specific and is happening in more than 95% of genes (Baralle and Baralle, 2018). Moreover, exons are small sequences (50-250 bp) while introns are thousand bp long, and splice sites can be “strong” or

"weak" according to the sequence similarity with the consensus sequence. As the vast majority of genes undergo alternative splicing, the splicing machinery has to choose between the different splice sites available. This choice relies on different regulators.

First, *cis*-acting sequences are contained in the pre-mRNA substrate. These can act as exon splicing enhancers or silencers (ESE and ESS), or intron splicing enhancers or silencers (ISE and ISS), by recruiting specific splicing co-factors. Moreover, pre-mRNA sequences can harbor secondary RNA structures modulating the accessibility of splicing factors. For example, RNA-RNA pairing competes with splicing factor binding onto *cis*-regulatory sequences. RNA secondary structures can also overlap with 5' or 3' splice sites and prevent their recognition by the splicing machinery (Ramanouskaya and Grinev, 2017). Frequently, *cis*-acting sequences act with *trans*-acting factors by specific recruitment. *Trans*-acting factors are serine/arginine (SR)-rich proteins that contain one or two RNA recognition motif(s), and heterogeneous nuclear ribonucleoproteins (hnRNP) that form complexes with Pol II-transcribed genes. Briefly, SR proteins are mostly associated with splicing enhancement as they promote exon inclusion by recruiting U1 snRNP or U2AF. Contrarily, hnRNP are mostly splicing repressors. However, this model is simplistic and reality is much more complicated as the effect of *trans*-acting factors is essentially position-specific. The phosphorylation status of splicing factors is another layer of regulation. Finally, splicing events can also be modulated by tissue-specific RNA binding proteins (RBP) that can act as negative or positive splicing factors depending on the pre-mRNA substrate targeted (Dvinge, 2018).

Besides, the global transcriptional environment also controls splicing events. Indeed, two models have been proposed: the "kinetic model" relies on the splicing site recognition modulation by the RNA Pol II elongation rate, and the "recruitment model" that assumes splicing to be directly or indirectly regulated by splicing factor recruitment onto the pre-mRNA. Definitely, these two models are not mutually exclusive, and rather inter-connected.

In a first place, transcription rate affects splicing events. As previously described, the elongation rate that can be modulated by several mechanisms (cf. part "*transcription elongation"*) and the chromatin structure are intimately linked to splicing processes. Interestingly, exons are enriched in nucleosomes compared to introns and help the recognition of exon by splicing machinery together with contributing to RNA Pol II slowdown in exonic regions. For example, the SWI/SNF nucleosomal remodeling complex decreases RNA

Pol II elongation rate thus leading to weak exon inclusion. Like nucleosomes, histone marks are enriched at exonic sequences respect to introns and can contribute to the recruitment of splicing factors. H3K4me3 can recruit U2 snRNP, H3K36me3 induces exon skipping by recruiting SRSF1. Moreover, histone acetylation can recruit HDAC that in turn impact splicing by modulating RNA Pol II elongation rate (Dvinge, 2018).

Furthermore, DNA methylation also influences transcription and thus splicing. Indeed, DNA methylation modulates nucleosome positioning and can affect histone post-translational modifications. Remarkably, splice sites harbor higher levels of DNA methylation, and 22% of alternative exons are regulated by DNA methylation (Lev Maor et al., 2015). The insulator protein CTCF recruitment to its binding site depends on DNA methylation. Indeed, a methylated site does not allow CTCF recruitment. On *CD45* mRNA, when CTCF binding site in unmethylated, CTCF associates to the chromatin and leads to Pol II pausing and subsequent exon 5 inclusion (Shukla et al., 2011). In this study, a genome-wide analysis revealed that CTCF binding regulates the upstream exon inclusion through Pol II pausing. Moreover, intragenic loops were demonstrated to form between exon and cognate promoter to promote alternative splicing (Tim R Mercer et al., 2013). More recently, CTCF has been identified to mediate intragenic loops between promoter and gene body. Notably, exons with upstream regions predicted to be CTCF binding sites were enriched for differential splicing (Ruiz-Velasco et al., 2017, p.) However, besides correlating CTCF binding sites to alternative splicing-supporting regions, the direct correlation between CTCF and alternative splicing regulation has not been demonstrated in this study. CTCF is not the only protein able to mediate DNA methylation state to alternative splicing regulation. Other proteins have been identified to intermediate such mechanisms, and are well-summarized by Naftelberg and colleagues (Naftelberg et al., 2015). Eventually, alternative splicing can also be regulated by the promoter architecture and the recruitment of specific TFs and co-activators that can influence the subsequent RNA processing. As expected, promoter-sensitive exons are mostly located in 5' of the neosynthesized RNA chain. The core spliceosome itself can modulate alternative splicing, through differential snRNP biogenesis, or snRNA changes through abundances in sequence variants (Dvinge, 2018).

Alternative splicing is also important for RNA quality control, both in the nucleus and in the cytoplasm by co-translational processes. Briefly, a polyadenylated RNA with an incomplete

splicing, and thus retained intron is sensed and eliminated by the nuclear exosome machinery. Similarly, spliced mRNA with a still-associated lariat, not fully discarded, retains the spliceosome on the mRNA and thus leads to RNA degradation. Finally, the choice in the alternative terminal exon is also decisive for the mRNA fate. Indeed, as the 3' UTR bears miRNA target sequences, alternative exons in the 3' UTR can regulate the stability and the translation rate of mRNA. In the cytoplasm, alternative splicing and the choice in alternative exon inclusion/skipping can trigger nonsense-mediated decay (NMD) when mRNA contains a premature termination codon, non-stop decay (NSD), when the ribosome reaches the polyA tail without having encountered any stop codon, or non-go decay (NGD) when stable RNA secondary structures trigger ribosome stalling (Ramanouskaya and Grinev, 2017).

Finally, several non-canonical splicing events can occur, representing major events for new transcript emergence during evolution. Non-canonical splicing processes can generate cryptic exons, micro-exons, recursive splicing, circular RNA, retained introns (Sibley et al., 2016).

3.3' end processing and polyAdenylation

In order to be properly exported to the cytoplasm, the neosynthesized mRNA has to detach from the DNA template and the associated RNA Pol II complex. This disassembly occurs in two different steps: the cleavage of the transcript and the addition of a polyadenylated tail. Correct polyadenylation is crucial for protecting the transcript from degradation. Indeed, hypo- or hyper-adenylated RNA molecules are retained in the nucleus and subsequently degraded. The polyadenylation tail is also important in the cytoplasm, as it controls RNA fate, such as stability, availability for translation and subcellular localization (Neve et al., 2017).

On the nascent transcript, *cis*-acting motifs constitute the core polyA signal. 21 nt upstream the cleavage site, an A-rich hexameric motif harbors the conserved sequence A-A/U-U-A-A-A. Interestingly, non-canonical A-rich sequence variants exist in 10-20% of all the hexameric sequences. 10 to 30 nt downstream of the cleavage site, a G/U-rich region is present. Contrarily to the hexameric conserved sequence, the G/U-rich region does not possess any consensus but mutations in this sequence strongly impairs cleavage efficiency. PolyA regions can also contain auxiliary elements that vary in location and nucleotide composition, and act as polyA enhancers. These are mostly G-rich and U-rich sequences. Among the U-rich elements located upstream of the hexameric sequence, the conserved UGUA motif is

frequently found and is involved in polyA site choice through the binding factors that are recruited (Neve et al., 2017).

The polyA machinery is recruited co-transcriptionally and is composed of more than 80 proteins. Among them, 20 compose the core machinery divided in four multi-protein complexes. First, the cleavage and polyadenylation specificity factor (CPSF) is composed of six subunits and recognizes the 5'-AAUAAA-3' motif and bears the endonuclease activity. The cleavage stimulatory factor (CstF) is recruited concomitantly on the G/U-rich sequence. CstF contains three subunits and is essential for the cleavage step, yet it does not contain the endonucleolytic activity. The cleavage factors (CF) I and II are also part of the core polyA machinery. CFI is composed of four polypeptides among which CFIm68 (CPSF6) and CFIm25 (Nudt21) alone are sufficient for UGUA upstream element recognition. A second CF is also present in the machinery and is assumed to bridge CFI to CPSF, yet CFII is just partly purified and its functions still need to be uncovered (Neve et al., 2017). The enzyme responsible for the addition of the polyA tail is the monomeric polyA-polymerase (PAP) that catalyzes the following reaction: ATP + RNA-3'OH ➔ PPi + RNApolyA-3'OH **(Figure 26)**. PAP adds around 200 non-templated adenosines in the exposed 3'-OH. Some non-canonical PAPs have been identified, notably associated to mitochondrial RNA polyadenylation. The polyadenylation rate of PAP is enhanced by the recruitment of the polyA binding protein (PABPN1). Finally, RNA Pol II-CTD is also part of the polyA machinery as it constitutes a dynamic interaction platforms for the different polyA factors (Neve et al., 2017).

The model explaining the complete detachment of the transcript from the machinery proposed that while being polyadenylated, the RNA tail is concomitantly coated with PABPN1 and folds in a spherical RNP that constrains the number of PABPN1 recruited. When PABPN1 cannot integrate anymore, the RNA-PAPBN1-PAP complex disrupts from CPSF and terminates the polyadenylation (Neve et al., 2017).

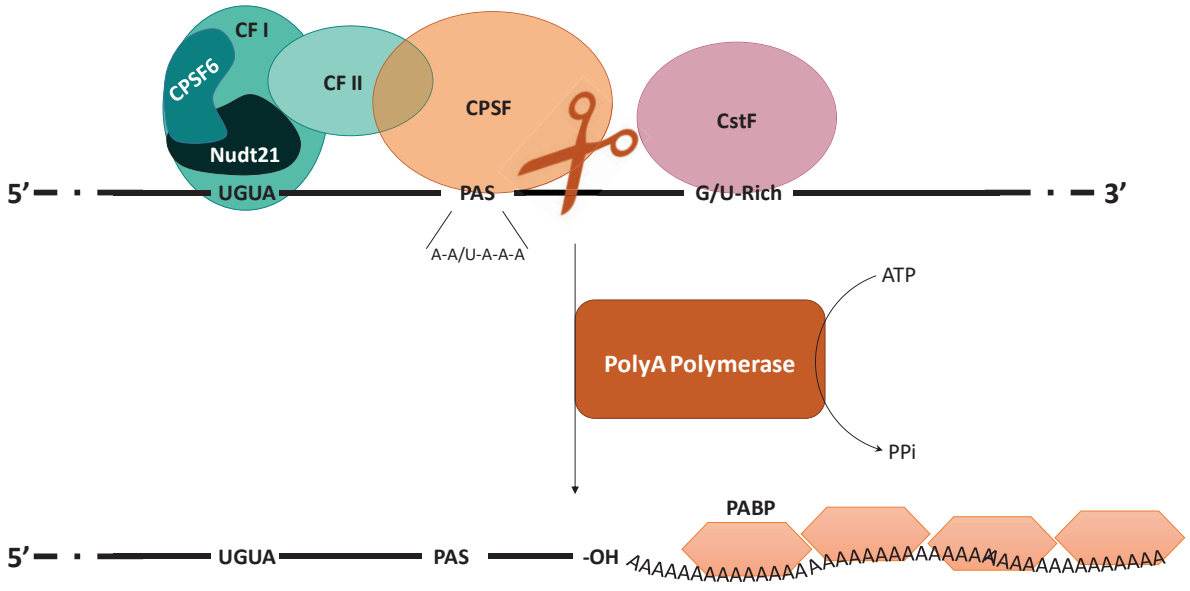


Figure 26 : 3' end processing and polyadenylation

The Polyadenylation site (PAS) is recognized by the Cleavage and Polyadenylation specific factor (CPSF), while the downstream G/U-rich region is recognized by the Cleavage stimulatory factors (CstF). Supplementary upstream regulatory motif, such as UGUA, recruit the Cleavage Factors I and II that mediate the choice of PAS. Only CPSF bears the RNA cleavage activity. Once mRNA is cleaved, the PolyA Polymerase is recruited and synthesizes a non-template polyA tail that is concomitantly coated by PABP protein.

In mammals, more than 70% of genes contain more than one PAS. The differential use of PAS leads to alternative transcripts. Alternative polyA (APA) can happen either upstream or downstream of the last exon. Upstream of the last exon, the use of an alternative PAS modifies the coding region of the transcripts and leads to different mRNA coding capacity and thus affects the cellular proteome. When APA happens downstream of the last exon, it triggers alternative length of the 3'-UTR leading to the presence or absence of regulatory regions involved in mRNA fate (Neve et al., 2017). Indeed, the 3'-UTR is targeted by microRNA and changes in 3'-UTR affects miRNA functions, notably stability and translation regulations (Tian and Manley, 2017). 3'-UTR size modifications and thus differential mRNA regulation have been reported during cell differentiation and in several pathologies such as cancer (Curinha et al., 2014).

The APA remains a poorly understood mechanism. However, it might principally relies on *trans*-acting RBP (Neve et al., 2017; Tian and Manley, 2017). Notably, CFI promotes the use of distal PAS by binding to the UGUA motif mostly enriched near distal PAS. Recently, CPSF6 and NUDT21, two subunits of CFI complex, were demonstrated to be enriched near distal PAS sites, consistently with the distribution of UGUA motif. Moreover, NUDT21 depletion triggers APA site usage of hundred genes upon depletion and increases a subset of protein levels. These observations were linked to a loss of miRNA target sequences and a subsequent upregulation of the translation (Brumbaugh et al., 2018). Indeed, when CPSF6 and NUDT21 are depleted, mRNA 3'-UTR are shortened (Hardy and Norbury, 2016).

Finally, the co-transcriptional polyA process is interdependent with capping and splicing processes. Capped RNAs are better substrate for polyadenylation, principally through the cap binding protein CBP. Furthermore, alternative splicing of exons and introns can stimulate or repress the use of PAS. Notably, 3' splice sites binding factors, such as U2 snRNP, directly contact polyA machinery *via* CPSF. On the contrary, factors that associate with 5' splice sites generally inhibit the polyA machinery to prevent premature polyadenylation (Neugebauer, 2002).

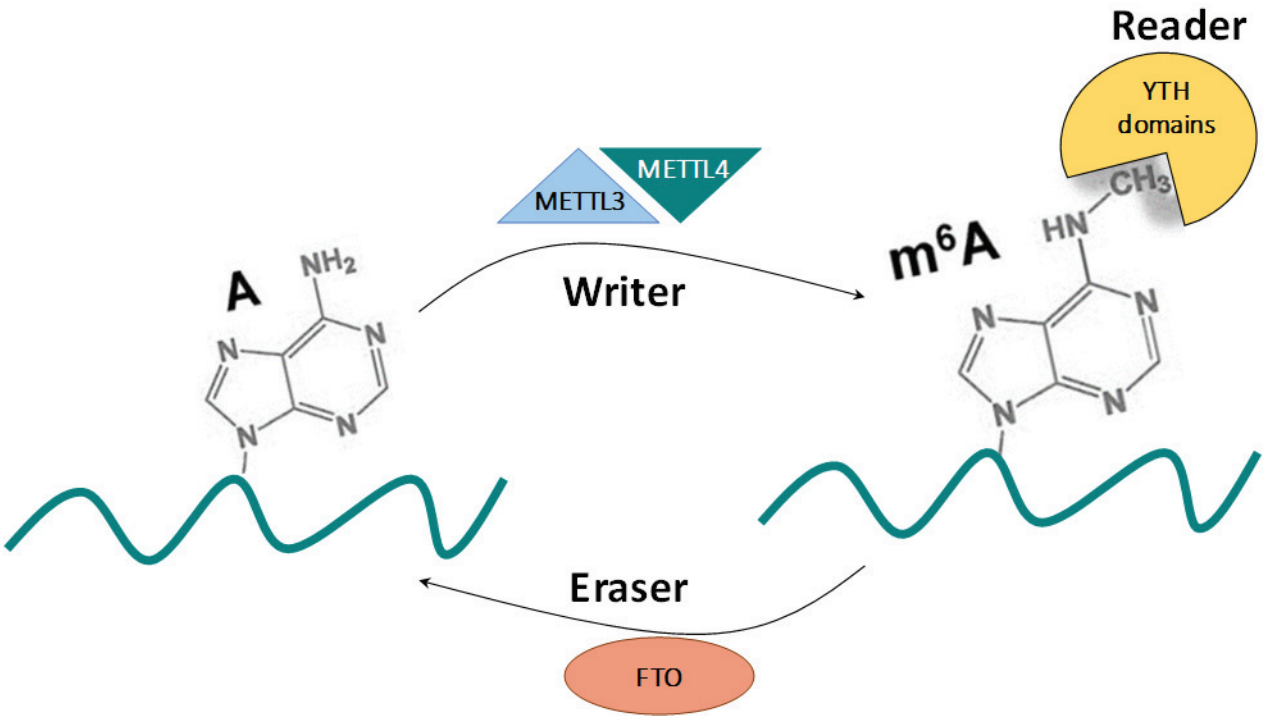


Figure 27 : RNA methylation

N6-methyladenosine is the most described RNA modification. It consists in adding a methyl group on an adenosine residue. This reaction is catalyzed by two methyltransferases, or writers, and can be reverted by and eraser that is a demethylase. Several readers bearing a YTH protein domain can recognize m6A and mediate subsequent nuclear or cytoplasmic processes.

4. RNA methylation

In eukaryotes, RNA molecules can undergo more than 100 modifications and raise the notion of “epitranscriptome”. The most studied modifications are RNA methylations that have been described to be key players in gene expression regulation. Six different types of RNA methylations have been identified: N1 and N6 methylation on adenosine, N3 and C5 methylation on cytosine, N7 methylation on guanosine and 2’OH methylation on the ribose. Nonetheless, the most abundant one is the N6-methyladenosine, or m^6A **(Figure 27)**. This mark is present on more than 7000 mammal genes in at least one copy. m^6A is found enriched in TSS, around stop codons and in the last exon of the mRNA. It has also been identified in non-coding RNA (ncRNA, lncRNA, snRNA). The RRACH motifs seems to be a conserved site on mRNA, yet, in the 5’-UTR, BCA is the conserved motif for m^6A. Methylation of adenosine, such as other modifications, is catalyzed by different enzymes: the writers are the methyl-transferases that deposit the methyl group on adenosine, the erasers are demethylases and remove the methyl group from the adenosine residues, and finally, several specific readers detect the m^6A and initiate the subsequent functions. These functions are nuclear, where m^6A regulates nuclear processes and cytoplasmic, where m^6A controls RNA stability and translation (Covelo-Molares et al., 2018).

On mRNA, m^6A writers are the homologous proteins METTL3 and METTL14 that act in the form of a heterodimer, and the METTL16, in association with different adaptors proteins. In the stable heterodimer METTL3/METTL14, only METTL3 bears the catalytic methyl-transferase activity, while METTL14 serves as a scaffold for the complex on the mRNA. This heterodimer is mostly associated to the consensus RRACH motif and methylates the “A” in the motif. METTL3/METTL14 are mainly found in nuclear speckles but their location depends on the adaptor proteins they are associated to (Covelo-Molares et al., 2018).

The methylation of N6-adenosine is a reversible mark that can be removed by two demethylases, namely FTO and ALKBH15 that co-localize with nuclear speckles.

m^6A has specific readers that drive specific functions. Most of the readers belong to the YTH protein family: YTHD1, YTHD2 and YTHD3. Besides, the translation initiation complex eIF3 also interacts with 5’-UTR m^6A. Readers can either directly recognize the methyl modification or detect the m^6A-mediated conformation change in RNA structure. Indeed, methylated RNA are

less folded than their unmethylated counterparts are. This folding modification due to m^6A is called "m^6A switch" and alters the association with RNA binding proteins, notably by modulating the accessibility of adjacent binding sites (Covelo-Molares et al., 2018).

Interestingly, deposition and removal of the methyl group principally occurs on the chromatin-associated pre-mRNA through co-transcriptional processes and further regulates several nuclear mechanisms. Firstly, m^6A is involved in pre-mRNA splicing by controlling the binding of specific splicing factors. For example, through conformation changes in the pre-mRNA structure, m^6A facilitates hnRNP accessibility. That is the case for hnRNPA2B1 that favors m^6A-METTL3-dependent alternative splicing by unknown mechanisms. In contrast, m^6A also acts indirectly *via* its writers and readers. Besides methylating N6-adenosine, METTL16 also adds a methyl group on the adenosine at position 43 on U6 snRNA that might be important for base pairing within the U4/U5/U6 tri-snRNP and subsequent spliceosome integrity. Moreover, the m^6A reader YTHDC1 recruitment on an m^6A mark in vicinity with a splice site can recruit SRSF3 and prevent SRSF10 association thereby enhancing alternative splicing (Covelo-Molares et al., 2018).

Secondly, m^6A is implicated in the formation of pre-mRNA 3' termini. Indeed, long last exons possess a higher probability of bearing m^6A sites that could positively or negatively modulate the alternative PAS selection. However, different studies described opposite regulation and the role of m^6A in PAS choice is still debated. The role of FTO in 3' length is more established as its depletion triggers longer 3' ends. Even though m^6A seems associated with 3' processing, a lot remains to be fully elucidated (Covelo-Molares et al., 2018).

Finally, reducing m^6A delays mRNA export out of the nucleus. Interestingly, SRSF3 forms an adaptor between YTHDC1 and the nuclear export factor 1 (NXF1) to facilitate the export of N6-adenosine methylated mRNA (Covelo-Molares et al., 2018).

Viral mRNAs have also been described to bear m^6A modifications. According to the virus, methylation does not have the same impact in the viral life cycle, and can be either pro- or anti-viral. For HIV, a loss of m^6A writers inhibits the viral replication. On the contrary, it has been demonstrated that the depletion of m^6A readers increases HCV replication (Patil et al., 2018).

5. Export

Once transcribed and fully processed, mRNAs reach the cytoplasm where they have different fates, such as translation to maintain or adapt the cellular proteome. The nuclear export machinery helps mRNA translocation through the nuclear pore complex (NPC) in order to exit the nucleus **(Figure 28)**.

The NPC is the largest and the most versatile transport channel existing in the cell, and is responsible of various protein and RNA carriage. The central channel of the NPC is composed of thousand FG-repeats. The cargo mRNA has to acquire an export receptor to be able to pass through FG-repeats. Most of nuclear cargos translocate through the NPC in a karyopherin-dependent manner. Interestingly, mRNAs are the only nuclear cargos that are not using karyopherin as export receptor. Differently, mRNA-specific export receptor is the heterodimer NXF1/NXT1, also called Tap15 factor. Tap15 recognizes both the proper mRNA to export, without any sequence specificity and the FG-repeat. Tap15 interaction with FG-repeat is weak in order to allow a rapid exchange from the nucleus to the cytoplasm (Xie and Ren, 2019).

Nuclear export can be summarized in three steps. The first one is the assembly of an export competent mRNP, composed of the mature mRNA associated to Tap15 factor through a DEAD-box ATPase UAP56. UAP56 mediates the binding of Tap15 to the mRNA *via* the transcription and export (TREX) complex, in the nucleus. Notably, THO, a subunit of TREX, was demonstrated to travel with RNA Pol II-phosphorylated CTD, and to recruit UAP56 onto the mRNA. The second step is the mRNA driving and translocation through the NPC. Finally, in the cytoplasmic face of the NPC, the mRNP cargo dissociates from its Tap15 receptor with the help of the DEAD-box ATPase DDX19 (Xie and Ren, 2019).

Only properly matured transcripts are exported. During transcription and concomitant RNA processing, plethora of RNA binding proteins (RBP) are loaded onto the nascent RNA, insuring an appropriate regulation of the following export. Indeed, the first RNA maturation step which is the decoration of the nascent RNA with a m7G cap, recruits TREX complex *via* its subunit

ALY that directly interacts with the capping binding protein (CBP). Moreover, TREX can export both intron-less and intron-containing mRNA. In order to only export correctly spliced transcripts, TREX is recruited on the mRNA in a splicing-dependent manner. Finally, TREX is

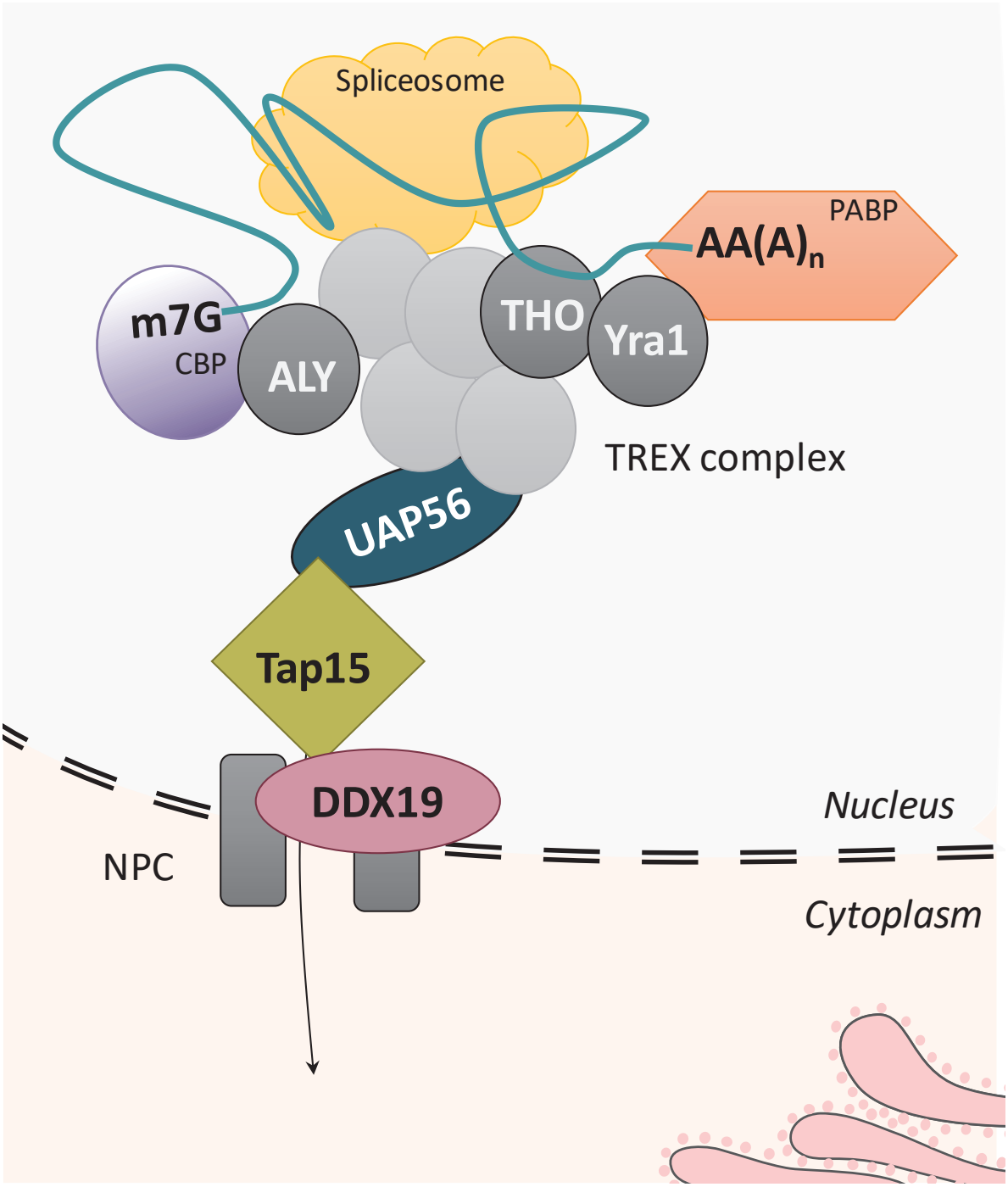


Figure 28 : mRNA export

The export machinery, or TREX complex, is composed of 7 subunits. The TREX complex is recruited by the CBP via its subunits ALY, THO and Yra1, the spliceosome and PABP, and mediates the nuclear export of properly processed mRNA. TREX interacts with the export receptor Tap15 to drive mature mRNA export, with the help of two RNA helicases, namely UAP56 and DDX19 that translocate the mature mRNA through the nuclear pore complex (NPC) to the cytoplasm.

also connected to the 3'end machinery, via its subunit Yra1 able to interact with a subunit of the CFI complex, associated to the RNA Pol II. Moreover, THO also interacts with polyA binding proteins (Xie and Ren, 2019).

6. HBV co-transcriptional regulation & RNA processing

HBV RNA transcription is addressed by the cellular RNA Polymerase II (Rall et al., 1983). It is thus natural to expect co-transcriptional RNA processing similar to cellular pre-mRNA **(Figure 29)**.

A. Capping

First, HBV RNAs are capped in their 5' termini. Capping of pgRNA is essential for effective encapsidation process (Jeong et al., 2000). Moreover, different studies are based on the presence of a 5' cap in HBV RNA for the precise identification of viral TSS (Altinel et al., 2016) and to discriminate HBV transcripts in cultured cells and patient samples (Stadelmayer et al., 2020a). However, besides taking advantage of HBV 5' capped RNA, no study explores the mechanisms involved in effective HBV RNA capping. Knowing that the capping machinery travels with the RNA Pol II and directly interacts with its CTD, it could be interesting to investigate the role of the three capping enzymes in HBV RNA capping. Moreover, our results demonstrate that X mRNA can exist in the form of uncapped transcripts in viral particles secreted by HBV-producing HepAD38 cell line (Stadelmayer et al., 2020a). Surprisingly, preC, pgRNA and S RNA are not detected in 5' RACE experiments uncovering uncapped X transcripts (Stadelmayer et al., 2020a). At present, the role of uncapped HBx transcripts remains unknown and raised the question about the fate of such RNA, notably concerning immune system recognition of un-capped RNA.

B. Splicing, export & stability

In the second hand, singly and multi-spliced variants have been identified in HBV. Predicted constitutive and cryptic splice sites have been found to vary in number and location according to the different HBV genotypes. To date, 17 variants have been identified to derive from pgRNA, and four to derive from Surface RNA (Candotti and Allain, 2016; Chen et al., 2015).

Interestingly, besides the fact that majority of HBV transcripts are unspliced, splicing of HBV RNA seems to be finely regulated by *cis*-acting sequence elements and *trans*-acting cellular and/or viral factors. Indeed, the conservation of predicted splice sites among genotypes A to D suggest a functional role of splicing in HBV life cycle. Among putative donor and acceptor sites, some have been found to be used in transfected cultured cells and in sera of infected patients. It is noteworthy to mention that the location of primers used to amplify spliced variants can affect the detection of some species. Furthermore, most of the variants have been identified in viral particles, thus introducing a selection bias toward variants that are still capable of being encapsidated (Candotti and Allain, 2016).

Nowadays, little is known about splicing regulation of HBV pre-mRNA. In HBV genome, some *cis*-acting elements have been identified to participate to splicing regulation. Two activators and one inhibitor of splicing have been identified within an intronic splicing regulatory domain in pgRNA, at the position 2951-2970, 3051-3070 and 3138-3143, respectively (Bhanja Chowdhury et al., 2011). Moreover, from nucleotide 1252 to 1288, a splicing regulatory element (SRE) has been identified to stimulate HBV splicing, notably *via* the splicing factor PSF. In this study, predictive tool identified this site as potential interactor for SR protein family (Heise, 2006). Two cellular factors have been reported as potential splicing *trans*-regulators. First, the multi-functional protein TARDBP, demonstrated to be involved in HBV transcriptional activation, was described to regulate HBV RNA splicing, and notably the single-spliced SP1, as its depletion triggers a 100% increase in splicing RNA ratio to full-length RNA. TARDBP was thus defined as a splicing inhibitor for HBV pgRNA (Makokha et al., 2019). On a second hand, PUF60, a member of the U2AF family, was reported to inhibit SP1 in favor to other alternatively spliced HBV RNA (Sun et al., 2017). The capsid protein itself might play a role in splicing as mutation in its C-terminal domain reduces encapsidated spliced isoforms (Candotti and Allain, 2016).

Remarkably, in eukaryotic cells, only correctly spliced mRNAs are exported as export factors interact with spliceosomal proteins. Surprisingly, majority of HBV transcripts are not spliced and still exported. This paradox lead to uncovering a post-transcriptional regulatory element, or PRE, in HBV genome, at the position 1217-1582, within the 3' end shared by all HBV RNA (Candotti and Allain, 2016). PRE serves as a splicing surrogate to allow the correct orientation-dependent nuclear export of the viral transcripts. Indeed, a 116 nt-long sub-element of PRE,

also called SEP1, interacts with a subunit of the TREX complex, the so-called ZC3H18 factor. Interestingly, the efficient recruitment of TREX on SEP1 is dependent on the 5' cap structure and the 3' end processing of viral mRNA (Chi et al., 2014). Evidences on the role of TREX, and Tap15 export factor have been reported, notably to interact with HBV 3.5 kb RNA and be involved in HBV RNA nuclear export (Yang et al., 2014). It is important to note that all these studies were performed in artificial models, such as full or partial HBV genome-containing plasmid transfection or HBV minigenes. Investigation in infectious models would confirm these results. Notably concerning TREX recruitment on viral transcripts.

In addition, PRE seems to be involved in pgRNA stability, but the mechanisms involved remain unknown (Candotti and Allain, 2016). Moreover, HBV transcript stability is also regulated by two non-canonical polyA polymerases, namely PAPD5 and PAPD7. These two proteins function in a redundant manner and their concomitant depletion leads to HBV RNA destabilization and degradation, without affecting the viral transcription (Mueller et al., 2019).

C. Polyadenylation

HBV transcripts are polyadenylated and all end at a common PAS. HBV genome element located at nucleotide 1916 and bearing the TATAAA sequence was first predicted to constitute a PAS for S transcript due to its great homology with canonical eukaryotic PAS (Pourcel et al., 1982). The year after, this PAS was confirmed to be used by the S transcripts. More precisely, transcripts were in reality identified to end 12 to 19 nts downstream of HBV PAS. Moreover, it was uncovered that the hexanucleotide composing HBV PAS was not sufficient to promote HBV RNA processing without the downstream 30 nts (Simonsen and Levinson, 1983). Interestingly, 3.5 kb RNA transcription starts around 100 nts upstream of HBV PAS, meaning that the longer-than-genome transcripts contain twice the PAS. In transgenic mice carrying the 3' portion of HBV genome, the use of the PAS was demonstrated to be tissue-specific. Indeed, in the liver, PAS was mainly used and transcripts stopped 10 nts downstream this site. Surprisingly, 500 nt-longer transcripts were identified in other tissues, such as kidney, spleen, brain and testis and identified to be read-through of the transgene. It was thus hypothesized that in other tissues than liver another PAS was used, yet, no more information was provided (Perfumo et al., 1992). Strikingly, HBV TATAAA PAS appears to overlap with a promoter region,

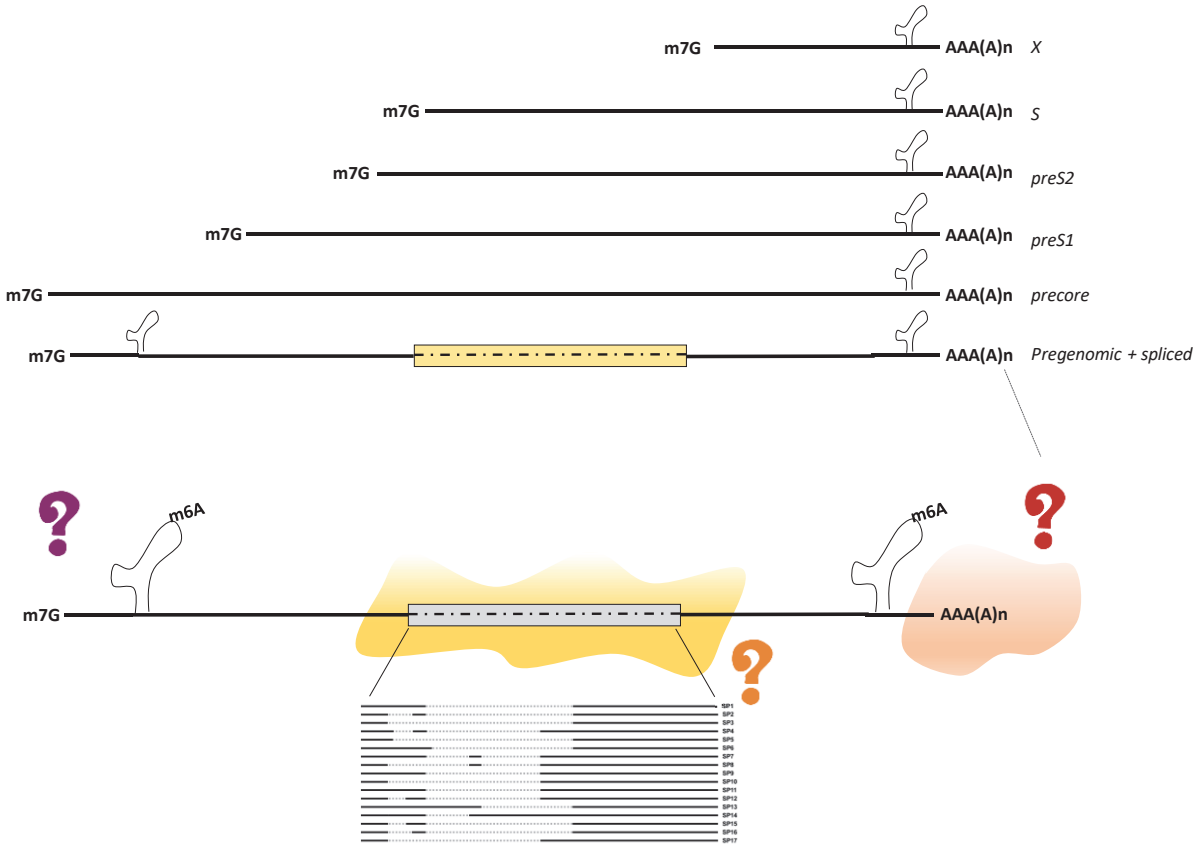


Figure 29 : Processing of HBV RNAs

HBV RNAs are capped in 5' and polyadenylated in 3'. Moreover, the two epsilon loops of the pgRNA can be methylated (m6A). Spliced variants deriving from the pgRNA have been identified. However, the mechanisms regulating 5' capping, splicing and polyadenylation of the viral transcripts are poorly understood and under-investigated.

notably due to its close sequence similarity with TATA-box promoter. HBV PAS was reported to bind general TFs such as TBP, TFIIA and TFIIB and to support transcription initiation in HBV-transfected cells. Mutation in TATAAA sequence to create the eukaryotic consensual AATAAA polyadenylation signal induces a loss of HBV DNA intermediates, thus suggesting an essential role of HBV PAS in the viral replication (Paran et al., 2000). The use of an upstream cryptic PAS was demonstrated in certain circumstances with DNA sequence restriction. Indeed, removal or inactivation of the HBV TATAAA PAS renders the cryptic CATAAA site functional (Schutz et al., 1996). Actually, this cryptic site is used by integrated HBV genome-deriving viral transcripts. In patients, integrated HBV genome lacks the canonical HBV PAS and produces truncated RNA polyadenylated at the cryptic site (Kairat et al., 1999; Su et al., n.d.; Wooddell et al., 2017).

Some defective circulating HBV DNA particles have been identified in patient samples. These shorter-than-WT HBV genomes were found in minority and harbored a premature polyA sequence immediately followed by 1.8 to 2.9 kb-long deletions. While harboring this internal polyA sequence and lacking long parts of the genome, the remaining sequences upstream of the aberrant the polyA tract contain the encapsidation signal together with DR1 and DR2 sequences. Despite being truncated and possessing an internal polyA, the presence of encapsidation elements in such sequences explains the fact that these shorter-than-WT HBV genomes could be found retro-transcribed and encapsidated in patient sera (Sommer et al., 1997).

D. Methylation

Finally, the viral transcripts are methylated in A residues belonging to ε-loop. Interestingly, 5' ε-loop and 3' ε-loop methylation displays different functions in HBV life cycle, the first one being involved in pgRNA retro-transcription while the second mostly interferes with translation. Noteworthy, the methylation writers (METTL3/14) and readers (YTHDF proteins) belonging to the cellular m^6A machinery are the ones regulating HBV RNA ε-loop methylation (Imam et al., 2018).

Despite the fact that HBV genome is well characterized to contain all the critical Pol II-related *cis*-elements to allow correct viral transcript maturation, little is known about how the different RNA processing mechanisms occur and are regulated. Indeed, alternative processing,

such as the presence of various spliced isoforms and choice in PAS, are conserved among HBV genotypes and seem to display essential functions in HBV life cycle. Recent studies aiming at elucidating such mechanisms do not rely on infectious models and artificial HBV gene expression can only partially explain how HBV RNA fully matures and the subsequent implications for viral replication.

V. Roles of DDX5/DDX17

In gene expression regulation, the highly conserved RNA helicases represent the largest family harboring an enzymatic activity and regulating all steps of RNA metabolism, from transcription to translation. Indeed, mRNA maturation and fate mostly rely on its association with RBP and the different subsequent mRNP generation. Dynamic reorganization of mRNP is controlled by RNA helicases (Bourgeois et al., 2016). The conserved core structure of RNA helicases is composed of two globular RecA domains involved in ATP binding and hydrolysis, and in sequence-independent RNA binding, through the sugar-phosphate backbone. N- and C-termini of helicases are variable and confer the specificity of each member, notably by interacting with different protein partners. RNA helicases are divided in six families, among which the most represented is the DEAD- (Asparagine-Glutamine-Alanine-Asparagine) box helicase family comprising 40 members (Bourgeois et al., 2016). In this part, we will focus on DDX5 (p68) and DDX17 (p72), two prototypic members of the DEAD-box helicase family.

1.Characterization

The multitasking RNA DEAD-box helicases DDX5 and DDX17 are paralogs that share 90% homology in their core domains and can associate in heterodimers in cells (Lamm, 1996; Ogilvie, 2003). The core domain consists in two RecA-like domains connected by a flexible linker domain, allowing an open or a closed conformation depending on the ATP binding status. ATP binding/hydrolysis and RNA annealing/unwinding activities rely on nine conserved motifs of the core domain, namely Q, I, Ia, Ib, II, III, IV, V and VI. *Via* their ATPase activity, DDX5 and DDX17 modulate local RNA secondary structures and RNA-protein association important for cellular processes, such as transcription, RNA maturation, export, degradation and translation. The amino and carboxyl ends of DDX5 and DDX17 share 60% and 30% homology, respectively (Fuller-Pace and Moore, 2011). Moreover, DDX17 possesses 2 alternative translation initiation sites generating 2 isoforms: p82 and p72 , the former being 79 amino acids longer at its N-Terminus (Lamm, 1996). As no distinction between these two isoforms will be made below, p72 and p82 will be referred as DDX17. Aberrant DDX5 and DDX17 protein expression is linked to some pathologies, such as cancer (Fuller-Pace and Moore, 2011).

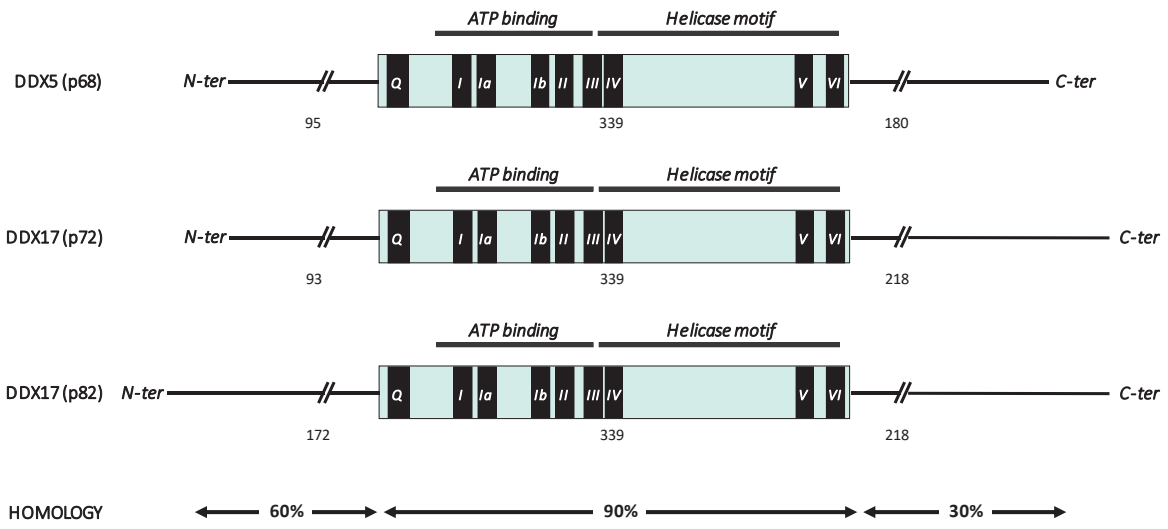


Figure 30 : Structure of the DEAD-box RNA helicases DDX5 and DDX17

DDX5 and DD17 share 90% homology in their core domain, composed of and ATP-binding domain and an helicase motif. The core domain is divided in 9 sub-domains, namely Q, I, Ia, Ib, II, III, IV, V and VI. Moreover, they share 60% homology in their N-ter domain, and 30% homology in their C-ter domain. DDX17 presents two isoforms: p72 and p82, the longer one produced from an alternative translation initiation codon.

Furthermore, the diverse functions of DDX5 and DDX17 are frequently investigated separately. However, whether they can act individually remains elusive.

2. Role of DDX5/17 in the transcription and co-transcriptional regulation

DDX5 and DDX17 have been associated to several processes of nuclear mRNA processing, mainly at the transcriptional level and alternative splicing regulation, and also in mRNA termination and export. DDX5 and DDX17 are also involved in microRNA maturation and have cytoplasmic functions that are well reviewed here (Bourgeois et al., 2016) **(Figure 31)**.

A. Transcriptional Regulation

DDX5 and DDX17 are co-factors of several transcription factors, such as p53 or estrogen receptor ERα. Transcriptional modulation mediated by DDX5 and DDX17 can be either positive or negative and depends on the nature of the TF stabilized and the target sequence (Giraud et al., 2018). Moreover, DDX5 and DDX17 interact with chromatin remodelers such as enzymes involved in histone PTMs. DDX5 and DDX17 binding can positively or negatively modulate histone acetyl transferases and HDACs. For example, DDX17 associates with CPB/p300 histone acetyl transferase and synergizes with it to promote transcription. On the contrary, helicases can also co-repress specific promoters by interacting with HDAC1 (Wilson et al., 2004). DDX5 can also modulate the methylation state of histone tails, notably in HBV-infected hepatocytes by interacting with SUZ12, a subunit of the PRC2 complex driving trimethylation of H3K27 and consequent transcriptional repression of cccDNA (Zhang et al., 2016). DDX5 and DDX17 are also associated with DNMT3A/B and might have a role in DNA methylation that is yet to be elucidated (Giraud et al., 2018).

In mammal cells, gene insulation mediated by CTCF is essential for allele-specific expression of IGF2/H19 locus. DDX5 associates to CTCF on IGF2/H19 locus *via* the long non-coding RNA steroid nuclear receptor activator (SRA). CTCF association to IGF2/H19 is not impaired by DDX5 nor SRA depletion, however, knock-down of SRA, DDX5 and potentially DDX17 affects CTCF-mediated insulation functions and consequent loss of IGF2/H19 imprinting. This imprinting is regulated by silenced expression of the maternal allele by CTCF-mediated contact inhibition between enhancer and promoter. DDX5 depletion increases the association between these

two regulatory regions and leads to an increase of the maternal-allele expression. Finally, in the IGF2/H19 locus, the cohesin complex also mediates insulation function of CTCF and is crucial for CTCF-mediated IGF2/H19 imprinting. On this locus, cohesin complex is also recruited in an interdependent manner with DDX5. Interestingly, genome-wide data revealed that DDX5 is present on around 20% of CTCF binding sites (H. Yao et al., 2010b). These latter results raise the question on the role of DDX5, and probably DDX17, in CTCF non-insulator functions, on other loci along the genome. A recent CTCF interactome and genome-wide study further identified overlapping binding of CTCF and DDX5, notably enriched at TSS that highly suggests a cooperation between these two proteins in transcription regulation (Marino et al., 2019). Noteworthy, DDX17 also belongs to CTCF interactome, but its presence on CTCF binding sites has not yet been assessed.

B. Splicing and alternative splicing regulation

DDX5 and DDX17 co-purify with the spliceosome, notably DDX5 co-immunoprecipitates with U1 snRNP at the 5' splice sites. These observations suggest a role of DDX5, and probably DDX17, in splicing machinery assembly. Indeed, DDX5 appears to favor the release of U1 snRNP from the core splicing machinery that renders the spliceosome catalytically active (Xing et al., 2019).

In a general manner, DDX5 and DDX17 depletion affects the global splicing pattern genome-wide (Xing et al., 2019). More precisely, DDX5/17 are able to drive an open structure of 5' splice site of *Tau* exon 10 leading to a facilitated access to U1 snRNP and following exon inclusion (Xing et al., 2019). Moreover, DDX5 and DDX17, *via* their helicase activity are able to resolve RNA secondary structures that might be important for alternative splicing. Notably, DDX5 and DDX17 are demonstrated to bind G-quadruplexes, but their ability to resolve this stable tertiary structure is yet to be elucidated (Xing et al., 2019). Interestingly, DDX5/17 depletion affects splicing of alternative exons that possess 5' splice site embedded in G-rich sequences (Dardenne et al., 2014). G-rich sequence and, more precisely, G-quadruplexes are known to be hnRNP H/F binding sites. During muscle cell differentiation and epithelial-to-mesenchymal transition, DDX5 and DDX17 appear to modulate splicing programs of exons predictably located upstream of G-rich sequences. Exon skipping correlates with decreased association of hnRNP H/F in G-rich sequences when DDX5 and DDX17 are downregulated (Dardenne et al., 2014). However, it is yet to be determined whether DDX5 and DDX17 directly

act on G-quadruplex structures or whether they mediate specific factor recruitment and subsequent alternative splicing changes.

In certain cases, alterations in splicing program may be implicated in human diseases. DDX5 knock-down leads to the inclusion of an alternative exon containing a premature stop codon in h-Ras mRNA, forming a shorter protein isoform that delays G1/S phase and consequently drive cells in a quiescent state (Xing et al., 2019). DDX5/17 are also involved in cell invasiveness as they promote the spliced isoform macroH2A1.1 which is linked to tumor-cell invasiveness by modulating expression of genes involved in redox metabolism (Dardenne et al., 2012). Remarkably, in this study, it is described that exons mostly skipped upon DDX5/17 depletion possess a 5' splice site enclosed in GC-rich sequences (Dardenne et al., 2012).

C. Termination

To date, few studies associate DDX5 to transcriptional termination regulation. During circadian clock, the PERIOD complex, containing Period (PER) and Cryptochrome (CRY) proteins, accumulates and transcriptionally represses its own Pre and Cry genes. PERIOD complex was identified to also contain DDX5 and RNA Pol II large subunit together with Pre and Cry pre-mRNA and Senataxin (Padmanabhan et al., 2012). As described in Termination Part, Senataxin in involved in transcription termination by resolving RNA:DNA duplexes thereby allowing Xrn2 digestion of the downstream RNA and subsequent Pol II release. During negative feedback, Pol II was demonstrated to accumulate near 3' end of Pre and Cry genes specifically. The PERIOD complex was found to associate with the elongating RNA Pol II complex and to inhibit Senataxin activity. By impairing Senataxin-mediated termination, the PERIOD complex impedes the RNA Pol II release and thus the recycling of the Pol II in a novel transcription cycle (Padmanabhan et al., 2012). In this study, the direct role of DDX5 has not been elucidated and its implication in transcription termination seems to rely on its belonging to the PERIOD complex. Recently, the involvement of DDX5 and DDX17 was demonstrated in transcription termination of HSP70 locus. Interactome of the termination and export complex THO identified Xrn2, CPSF6, DDX5 and DDX17 as protein partners and good candidates for transcription termination regulation (Katahira et al., 2019). Notably, the concomitant depletion of DDX5 and DDX17 lead to an accumulation of HSP70 transcripts in nuclear foci, concomitantly with an accumulation of Xrn2, CstF64 and THO at the HSP70 locus.

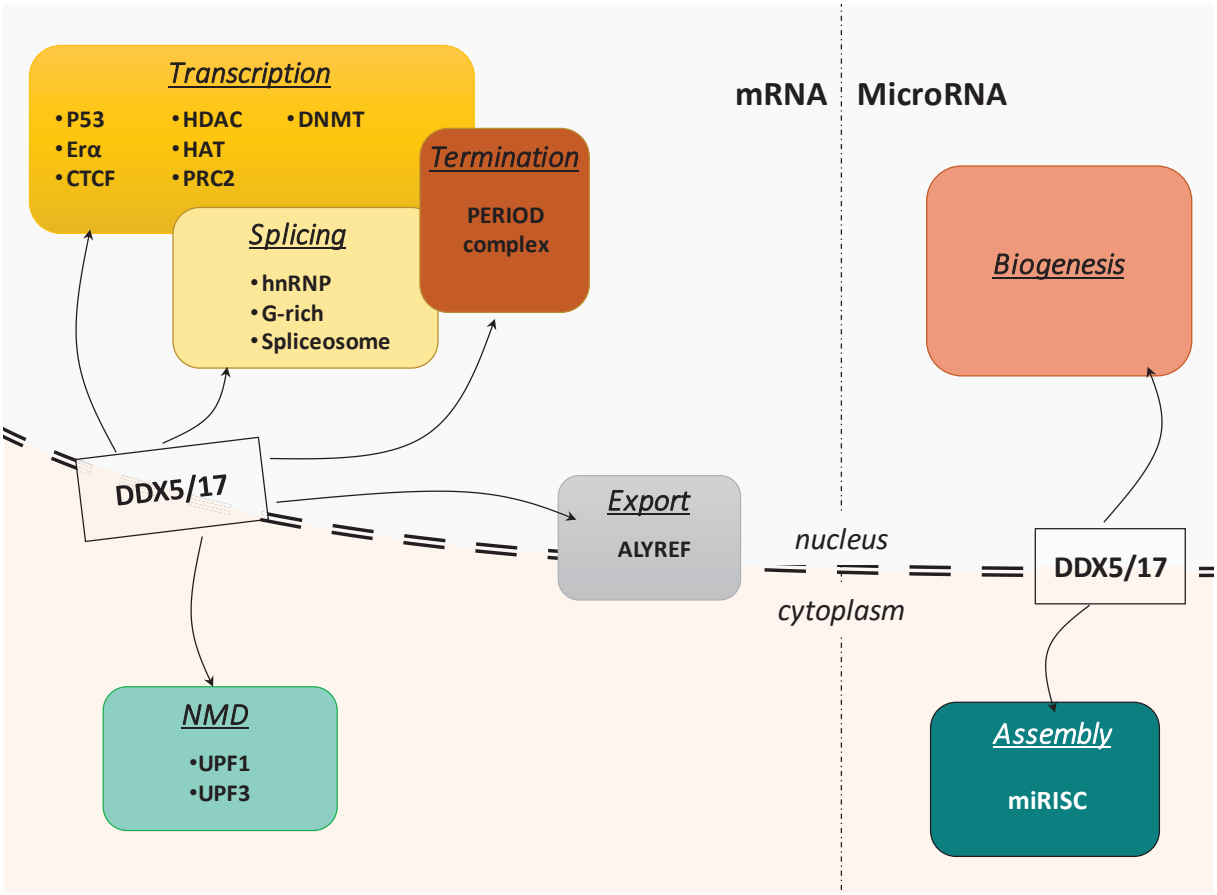


Figure 31 : DDX5 and DDX17 protein partners according to gene expression processes

DDX5 and DDX17 have several protein partners and are involved in several cellular processes implicated in gene expression regulation. First, in mRNA metabolism, DDX5 and DDX17 interact with factors to regulate transcription, and co-transcriptional events (splicing, termination, export). In the cytoplasm, DDX5 and DDX17 are involved in nonsense-mediated decay that leads to mRNA degradation.
DDX5 and DDX17 are also involved in microRNA biogenesis and assembly, notably within the miRISC complex, which is the functional unit of miRNA-mediated translation regulation.

Remarkably, the cleavage and the polyadenylation was not affected upon DDX5/17 silencing (Katahira et al., 2019). Knowing DDX5 role in R-loop resolution, the following model was proposed: DDX5/17 restrict R-loop formation thereby facilitating Xrn2-mediated transcriptional termination and release of the mature transcript. Evidence for this model was further brought by another study that reports that DDX5 unwinds R-loops thereby allowing a complete degradation of the read-through transcript by Xrn2 (Mersaoui et al., 2019). DDX5 interact with Xrn2 and mediate RNA Pol II pausing downstream of the PAS. DDX5 was thus demonstrated to be important for transcription termination and neosynthesized transcript release from the DNA template (Mersaoui et al., 2019). In this study, DDX17 has been identified as a protein partner of DDX5, yet, its role in transcriptional termination has not been addressed.

D. Other functions of DDX5 and DDX17

Besides transcription and pre-mRNA processing regulatory functions, DDX5 and DDX17 are also important for consequent mature RNA metabolism steps. Notably, DDX5 is a regulator of co-transcriptional splicing mRNP assembly and subsequent proper mRNA export. Indeed, correct mRNP gathering requires ATPase-mediated unwinding activity of DDX5. During export, ALYREF recruitment to the mRNP inhibits DDX5 activity to prevent mRNP rearrangement and favor an effective export (Bourgeois et al., 2016). Moreover, DDX5 associates with some factors of nonsense-mediated decay (NMD) process and interestingly enhances NMD-driven degradation of its paralog transcript DDX17 (Geißler et al., 2013, p. 3). Another important role of the DEAD-box helicases DDX5 and DDX17 is the regulation of microRNA biogenesis and their assembly in the miRISC complex (Bourgeois et al., 2016). This latter role in miRNA processing renders DDX5 and DDX17 important regulators of post-transcriptional gene regulation. Importantly, DDX5 and DDX17 directly promote the expression of some miRNA that in turn down-regulate their expression (Dardenne et al., 2014; Lambert et al., 2018).

Fine control of each of the paralogs, notably through regulatory feedback loop involving miRNA, maintains their stable expression at the protein level. Moreover, DDX5 and DDX17 share high homology and similar cellular functions that suggest a functional redundancy between these two proteins (Giraud et al., 2018). As their individual role is intricate to assess, concomitant deletion in cells appears to be a relevant way to examine their common functions and avoid functional overlying.

E. Role of DDX5/17 in the regulation of viruses

Ubiquitous expression of DDX5 and DDX17 in plant and animal cells make them essential regulators of gene expression, from transcription to translation. As viruses require host cell machineries to fully replicate, evidences for the role of DEAD-box helicases emerged in replication cycle of several viruses. DDX5 and DDX17, mostly independently, have been linked to be essential for many virus lifecycles, notably for RNA viruses, while DDX5 and DDX17 functions in DNA virus infections have been investigated to a lesser extent **(Figure 32)**.

First, DDX5 and DDX17 are implicated in transcriptional regulation of different RNA and DNA viruses (Wenyu Cheng et al., 2018). The (+)-RNA genome containing Several Acute Respiratory Syndrome Coronavirus (SARS-CoV) encodes a viral helicase called nsp13 that unwinds RNA and DNA duplexes. Nsp13 interacts with DDX5, which is assumed to act as a co-regulator of SARS-CoV transcription (Wenyu Cheng et al., 2018). Moreover, the Porcine Reproductive and Respiratory Syndrome Virus (PRRSV), another (+)-RNA genome-containing virus responsible of global swine diseases, encodes an RNA polymerase called nsp9 that interacts with DDX5 to promote viral replication (Wenyu Cheng et al., 2018). Moreover, the trimeric viral polymerase of Influenza virus interacts with both DDX5 and DDX17. The viral titer decreases upon DDX5/17 depletion, unravelling a pro-viral role of these two helicases in Influenza replication (Wenyu Cheng et al., 2018). However, the precise role of DDX5 and DDX17 in Influenza viral cycle remains elusive.

For the Human Immunodeficiency Virus (HIV-1), both DDX5 and DDX17 display several roles in the viral lifecycle, notably in HIV-1 RNA processing. HIV-1 is a 9 kb-long RNA virus whose replication relies on several host factors, and notably on DEAD-box helicases. DDX5 was found to complex with the viral regulatory protein Rev, which is crucial for nuclear export of unspliced HIV-1 transcripts. As DDX5 and Rev seem to belong to the same complex, a role of DDX5 in HIV-1 unspliced mRNA export was suggested but is yet to be demonstrated (Wenyu Cheng et al., 2018). In HIV-1 replication cycle, DDX5 and DDX17 were hypothesized to act independently (Sithole et al., 2018). DDX17 displays several roles in HIV-1 replication, notably in unspliced viral transcript export by interacting with Rev, degradation of HIV-1 RNA through its interaction with the NMD factor Upf1. Moreover, DDX17 regulates alternative splicing of HIV-1 pre-mRNA (Lorgeoux et al., 2013). Particularly, DDX17 depletion triggers an increase in HIV-1 spliced transcripts in line with a decrease in HIV-1 full-length mRNA. Opposite results

were observed in siDDX5 condition and a compensatory effect of each protein to its paralog has been proposed (Lorgeoux et al., 2013). However, the alternative splicing of the A4/5 cluster on HIV-1 RNA was shown to only depend on DDX17 (Sithole et al., 2018). In this study, Sithole and colleagues validated that DDX17 depletion triggers an increased level of full-length HIV-1 transcripts, and on the contrary a significant reduction of multiply spliced viral transcripts. Moreover, they identified U2AF65/35 and SRSF1/2 as protein partners of DDX17 associated to the 3' splice site of A4/5 cluster (Sithole et al., 2018). Besides uncovering a role of DDX17 in alternative splicing of HIV-1 mRNA, these two latter studies reported opposite effects of DDX17 on spliced to unspliced ratio. Furthermore, the independent role of DDX17 was demonstrated by a lack of phenotype rescuing in DDX5 overexpression upon DDX17 depletion. Noteworthy, no efficient overexpression of DDX5 was obtained in DDX17 depleted cells and further demonstrations would be required to conclude on the DDX5-independent function of DDX17. Further investigations are necessary in order to fully elucidate the role of DDX17 in HIV-1 RNA splicing regulation and to exclude or comprise a role of DDX5.

Finally, DDX5 and DDX17 also display regulatory role in viral translation processes. Hepatits C virus (HCV) and Japanese Encephalitis Virus (JEV) both belong to Flaviviridae family and possess a (+)-RNA genome. HCV NS5B protein that transcribe (+)- and (-)-RNA strands interacts with DDX5. Moreover, DDX5 also associates with HCV 5'-UTR and is thereby proposed to enhance IRES-dependent HCV translation (Wenyu Cheng et al., 2018). In HCV life cycle, DDX5 is important in viral genome synthesis. However, contradictory studies demonstrated that depleting DDX5 leads to higher HCV RNA levels and further investigations would help in elucidating the precise role of DDX5, and possibly DDX17, in HCV viral cycle (Wenyu Cheng et al., 2018). In JEV life cycle, DDX5 interacts with the core protein and the two non-structural proteins NS3 and NS5 implicated positive regulation of the viral replication through their helicase activity. NS3 and NS5, together with DDX5, bind JEV RNA 3'-UTR, probably mediating IRES-dependent JEV translation (Wenyu Cheng et al., 2018). Interestingly, besides HCV and JEV, other flavivirus members were found to hijack DDX5 activity. Among them, the pestivirus family, regrouping small (+)-RNA animal viruses, were found to interact with DDX5, notably *via* N^{pro} viral protein involved in viral RNA translation (Wenyu Cheng et al., 2018).

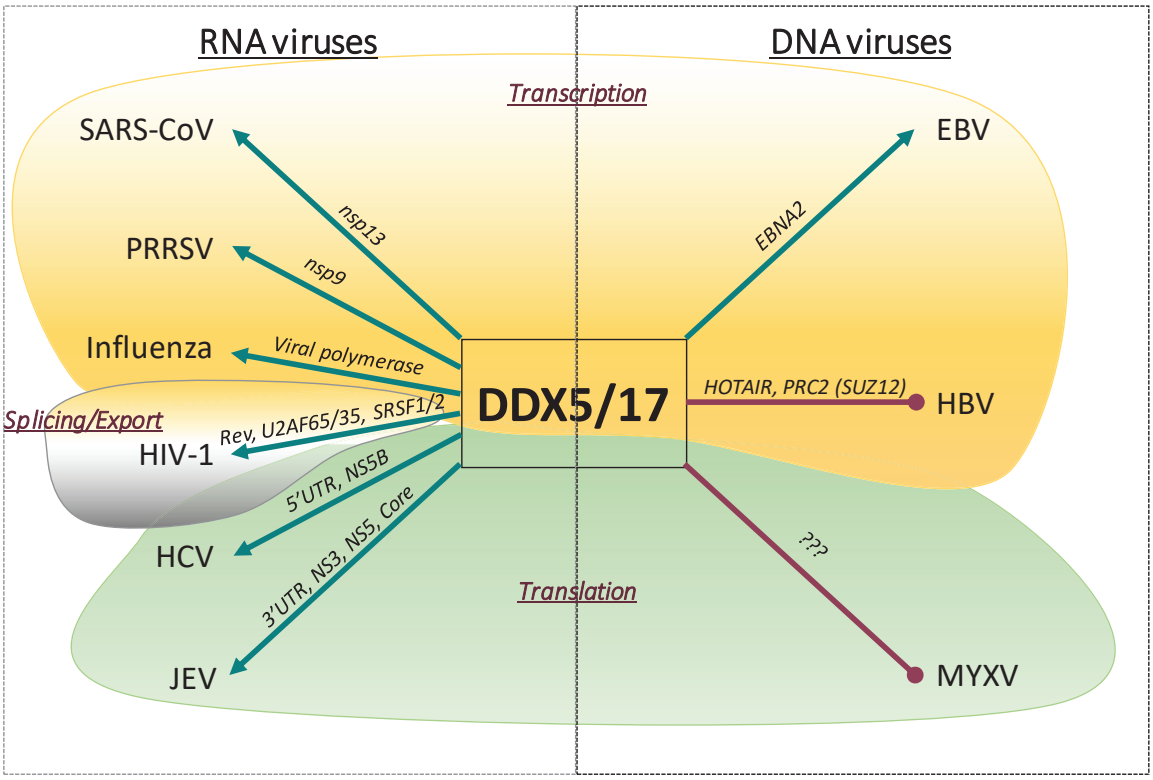


Figure 32 : DDX5/17 and viral replication

DDX5 and DDX17 are involved in viral replication of both RNA and DNA viruses at several steps, notably in viral transcription, RNA splicing/export and protein translation regulation. DDX5 and DDX17 can act as pro-viral factors (green arrows), or as anti-viral factors (purple lines).

DDX5 is also restrictive for Myxoma virus (MYXV) replication cycle. Indeed, DDX5 depletion triggers increased viral titer but seems not to be involved in gene expression, as suggested by luciferase reporter assay. In this study, DDX5 is suggested to act indirectly on MYXV replication repression through activation of antiviral innate responses (Rahman et al., 2017).

Concerning DNA viruses, DDX5 interacts with Epstein-Barr virus (EBV), a dsDNA, *via* its nuclear antigen 2 (EBNA2) during latency stage of EBV infection. EBNA2 regulates cellular and viral transcription by interacting with cellular TF and co-activators (Wenyu Cheng et al., 2018). Besides interacting with EBNA2, DDX5 role in EBV replication is still poorly understood. In HBV-infected cells, DDX5 seems to be downregulated at mRNA and protein levels thereby conferring an advantage for viral transcription. Indeed, DDX5 has been demonstrated to compete with the E3-ligase Mex3b for the binding to the lncRNA HOTAIR. When DDX5 is downregulated, HOTAIR acts as a scaffold between Mex3b and the PRC2 subunit SUZ12 to promote its proteasomal degradation and thereby a decrease in the repressive H3K27me3 mark deposition on cccDNA. Oppositely, when DDX5 is overexpressed in HBV-infected cells, it sequesters HOTAIR, thus preventing the degradation of SUZ12 and the persistence of H3K27me3 mark on cccDNA and subsequent viral transcription repression (Zhang et al., 2016).

In a general manner, RNA viruses seem to take advantage in DDX5 recruitment to facilitate their replication at different steps of viral gene expression (Wenyu Cheng et al., 2018). Interestingly, *via* its helicase activity, DDX5 participates in RNA secondary structure resolution. This property suggests a role of DDX5 in unwinding viral RNA genomes to promote efficient viral gene expression and proper following replicative cycles. Interestingly, in DNA viruses, DDX5 is mostly associated to restrictive activities (Wenyu Cheng et al., 2018). While its exact role is not fully elucidated in DNA viruses, DDX5 seems to be associated to latency control and is suggested to act as a restriction factor for MYXV and HBV by impairing viral transcription (Wenyu Cheng et al., 2018; Zhang et al., 2016). Noteworthy, besides transcriptional repression of HBV cccDNA and a poorly described role in EBV EBNA2-mediated transcription regulation, the role of DDX5 and DDX17 is understudied in episomal viruses.

Finally, the different abovementioned roles of DDX5 and DDX17 have been addressed individually and no compelling demonstrations on independent functions of these two helicases were depicted. As DDX5 and DDX17 share 90% sequence homology in their core domain (Lamm, 1996), can associate in heterodimer (Ogilvie, 2003), and have been described

to participate to similar gene expression processes (Bourgeois et al., 2016; Giraud et al., 2018), overlapping functions of these two helicases are extensively assumed. To fully appreciate the role of DDX5 and DDX17 in viral replication, and to avoid redundancy between the two paralogs, analysis by concomitant silencing appears to be a relevant approach.

Research Projects

I. Hypothesis and objectives

Despite the existence of effective antiviral treatments that maintain viral replication under control, cccDNA is still a major obstacle to achieve a definitive cure for HBV chronic infection (Lok et al., 2017). Indeed, cccDNA persistence in the nucleus of infected hepatocytes is responsible for HBV chronicity. cccDNA is the unique replicative intermediate used as template for viral transcription. Through its chromatinized structure, containing histones and non-histone proteins, together with the presence of regulatory sequences, cccDNA hijacks the host RNA Polymerase II machinery to ensure its efficient transcription (C.Thomas Bock et al., 2001; Moolla et al., 2002; Rall et al., 1983).

While HBx is well described to be essential in the transcriptional active state of cccDNA (Lucifora et al., 2011), little is known about the implication of cellular factors in cccDNA activity. Notably, HBV RNA metabolism is poorly studied and the processes leading to HBV RNA capping, splicing and polyadenylation are largely under-investigated.

In eukaryotic cells, RNA Pol II–transcribed genes are concomitantly processed by diverse cellular machineries. RNA maturation is crucial for proper generation of a suitable mRNA for translational processes and correct protein synthesis (Neugebauer, 2002). The largest family of enzymes involved in gene expression regulation, notably transcription and co-transcription modulation, are the so-called DEAD-box RNA helicases (Bourgeois et al., 2016). Among them, DDX5 and DDX17 are typical representatives, largely demonstrated to regulate alternative splicing together with several mechanisms involved in RNA processing (Bourgeois et al., 2016; Giraud et al., 2018). Furthermore, DDX5 and DDX17 have been demonstrated to participate to insulator function of CCCTC-binding protein (CTCF) (H. Yao et al., 2010). CTCF is an insulator protein involved in RNA metabolism, particularly in alternative splicing regulation (Shukla et al., 2011). Interestingly, both DDX5/17 and CTCF have been shown to be involved in the regulation of the replication of several viruses (Pentland and Parish, 2015; Wenyu Cheng et al., 2018).

Based on these observations, and previous results obtained in the laboratory, the main aims of my PhD thesis work were to:

i) Investigate the role of DDX5 and DDX17 in HBV RNA transcription and processing
ii) Investigate the implication of CTCF in cccDNA regulation

Research Project 1

Article in preparation

Supplementary Results

II. DDX5/17 are involved in HBV RNA processing

1.Article in preparation

DDX5 and DDX17 regulate HBV RNA processing

Fleur Chapus*[1, 2], Guillaume Giraud*[1], Caroline Charre[1, 2, 3], Hélène Polvèche[2, 4], Jean-Baptiste Claude[2,4], Maria-Guadalupe Martinez[1, 2], Judith Fresquet[1], Audrey Diederichs[1, 2], Maëlle Locatelli[1, 2], Lucie Dubrunfaut[4], Michel Rivoire[5], Cyril Bourgeois[2,4], Fabien Zoulim[1, 2, 3] and Barbara Testoni[1]

[1]Cancer Research Center of Lyon, INSERM U1052, CNRS UMR-5286, 69008, Lyon, France, [2]University of Lyon, Université Claude Bernard Lyon A (UCBL1), 69008, Lyon, France, [3]Hospices Civiles des Lyon (HCL), 9002, Lyon, France, [4]Ecole Normale Supérieure de Lyon, INSERM U1210, CNRS-UMR 5239, 69007, Lyon, France, [5]Centre Léon Bérard (CLB), INSERM U1032, 69008, Lyon, France

ABSTRACT

Background and aims. Hepatitis B virus (HBV) minichromosome, the covalently-closed-circular (ccc)DNA, is responsible for HBV chronicity and reactivation. Current antiviral treatments inhibit viral replication, but do not significantly affect cccDNA reservoir in the hepatocyte nuclei. cccDNA is a stable episome with a chromatin structure regulated by epigenetic mechanisms and the unique template for all viral transcripts. Overlapping open reading frames and regulatory elements on HBV genome have to be precisely and timely controlled to ensure correct cccDNA transcription, termination and viral mRNA processing. To better understand the mechanisms orchestrating cccDNA transcriptional regulation, we investigated the role of DDX5 and its highly related paralog DDX17, two DEAD-box helicases known to play a driving role in connecting cellular chromatin topology to correct transcriptional initiation and termination.

Methods. Chromatin Immunoprecipitation (ChIP), RNA-IP, 5' and 3' RACE, Next and Third generation sequencing and knock-down (KD) by RNAi were performed in HepG2-NTCP cells and Primary Human Hepatocytes one week after HBV infection, when cccDNA pool is stable.

Results. DDX5/17 knockdown by small interfering RNA (siRNA) only slightly affected global HBV RNA levels, while severely impacting the length of viral transcripts, which migrated faster in Northern Blot analysis. RNA-Seq under control and DDX5/17 KD conditions revealed alternative splicing of pgRNA sliced variants if DDX5/17 were depleted. Moreover, 3'RACE coupled to MinIONTM Nanopore single molecule sequencing identified a differential use of polyadenylation signal in the absence of DDX5/17. RNA-IP showed the binding of DDX5/17 to viral RNAs, and to preC/pgRNA in particular, and a role for DDX5/17 in mediating the recruitment of CPFS6 and NUDT21, two factors involved in transcriptional termination, to viral transcripts.

Conclusions. Altogether, our data highlight an essential role of DDX5 and DDX17 in fine tuning HBV RNA processing and ensure proper viral replication.

INTRODUCTION

Hepatitis B Virus (HBV) is the causative agent of a chronic hepatitis that affects more than 250 million people in the world. Chronic carriers have increased risks of developing severe liver diseases, such as liver cirrhosis and hepatocellular carcinoma (HCC) (Trépo et al., 2014). Despite effective prophylactic vaccines and antiviral treatments, chronic hepatitis B remains a major health burden, causing more than 800 000 deaths every year (WHO, 2019). A key determinant of the chronicity of viral infection is the persistence of HBV covalently closed circular DNA (cccDNA), which is its episomal replicative intermediate residing in the nucleus of infected hepatocytes. cccDNA contains the complete viral genetic information and is the template for full viral transcription and replication. Therefore, direct targeting of cccDNA is fundamental for a complete HBV cure (Nassal, 2015).

cccDNA harbors a chromatin structure regulated by epigenetic mechanisms, and is the unique template for transcription of the viral RNAs. cccDNA encodes six subgenomic RNAs: the 0.7 kb X RNA, the two 2.1 kb PreS2 and S RNAs, the 2.4 kb PreS1 RNA and the two longer-than-genome 3.5 kb preCore (pC) and pregenomic (pg) RNAs (Seeger and Mason, 2000). This latter is also the RNA intermediate necessary to achieve complete viral replication. pgRNA contains the whole genome information and is retro-transcribed into rcDNA by the viral polymerase/retro-transcriptase while being encapsidated to form nucleocapsids. These newly assembled nucleocapsids can be either recycled to the nucleus replenishing the cccDNA pool, or be enveloped and secreted to form new infectious particles (Seeger and Mason, 2000). Viral transcription is regulated by 4 promoters, 2 enhancers, and a common polyadenylation signal (PAS) and is orchestrated by the cellular RNA polymerase II (Rall et al., 1983). Each viral RNA starts at its own Transcription Start Site (TSS) and harbors a cap at its 5' end (Jeong et al., 2000). HBV genome contains four overlapping open reading frames (ORF), oriented in the same direction, with every base coding for at least one ORF (Nassal, 2015) and leading to the synthesis of the seven viral proteins. In addition to the above-described HBV transcripts, several spliced variants have been identified *in vivo*, in patient samples, and *in vitro*, in cell culture. Considering all HBV genotypes, seventeen spliced variants deriving from the pgRNA and four spliced variants deriving from the PreS2/S mRNA have been identified (Candotti and Allain, 2016; Chen et al., 2015). These spliced variants can be retro-transcribed and encapsidated, generating defective HBV DNA particles that are correlated with severe liver

diseases in patients (Soussan et al., 2008). The mechanisms and host factors involved in the fine-tuned regulation of cccDNA transcription are not yet fully understood.

The highly conserved DDX5 protein is a prototypic member of the DEAD-box RNA helicase family involved in several steps of RNA metabolism. This multitask protein regulates RNA processing from transcription to translation, through transcriptional regulation, pre-mRNA splicing, RNA export and decay (Bourgeois et al., 2016; Giraud et al., 2018). Several studies reported a role of DDX5 in promoting RNA virus replication (Wenyu Cheng et al., 2018). However, other studies demonstrated an antiviral role of DDX5 in DNA viruses, such as Myxoma Virus and HBV (Wenyu Cheng et al., 2018). In particular, DDX5 has been proposed to stabilize the PRC2 subunit SUZ12 promoting the deposition of the repressive H3K27me3 histone mark on HBV cccDNA subsequently inhibiting HBV RNA transcription (Zhang et al., 2016).

DDX5 shares more than 70% of sequence homology with its paralog DDX17 (Lamm, 1996) resulting in functional redundancy of these two proteins (Lamm, 1996). As DDX5, DDX17 is also implicated in gene expression regulation (Bourgeois et al., 2016) and has been shown to be involved in HIV-1 splicing regulation (Lorgeoux et al., 2013; Sithole et al., 2018). A major limitation of most of studies investigating the role of DDX5 and DDX17 in viral replication consists in analyzing them individually, without taking into consideration their redundant and reciprocal regulation role.

In this study, we aimed at investigating the role of both DDX5 and DDX17 in cccDNA transcription and RNA processing. The downregulation of both DDX5 and DDX17 protein expression in primary human hepatocytes (PHH) and in HBV-infected HepG2-NTCP cells indicated that DDX5 and DDX17 are mainly involved in viral RNA processing. Using classic molecular virology techniques, an adapted pipeline for HBV splicing detection by Next generation RNA sequencing (RNA-Seq) and Single molecule sequencing (Nanopore), we demonstrated that depletion of both DDX5 and DDX17 impairs exon inclusion of specific splicing junctions on HBV pre-mRNA together with favoring the choice of proximal polyadenylation site (PAS). For the first time, we precisely mapped the termination site of viral transcripts and uncovered a new co-transcriptional role of DDX5 and DDX17 in proximal PAS choice.

RESULTS

DDX5 and DDX17 are involved in correct HBV RNA processing in Primary Human Hepatocytes

To investigate the role of the DEAD-box RNA helicases DDX5 and DDX17, HBV-infected Primary Human Hepatocytes (PHH) and HepG2-NTCP cell line were transfected with siRNA concomitantly targeting DDX5 and DDX17 to avoid functional redundancy between these two proteins at 4 days post-infection, when cccDNA pool is formed and transcriptionally active (Figure 1, 2). Silencing efficiency in PHH and HepG2-NTCP was confirmed at both mRNA and protein levels (Fig1A and 2A) without any significant cell toxicity (data not shown). The levels of HBV replication markers were analyzed 2 days post-2nd transfection (i.e. 8 days post-infection) by ELISA and specific (RT)-qPCR. HBV antigen secretion was slightly reduced upon DDX5 and DDX17 silencing, with a higher decrease for HBe than for HBs antigen secretion, which was not affected in HepG2-NTCP cells (Fig1B and 2B). No changes in total HBV DNA and cccDNA levels were detected (Fig1C and 2C), while HBV total RNAs and 3.5 kb RNAs showed a non-significant increase trend upon DDX5 and DDX17 downregulation in both cell models (Fig1D and 2D). To determine whether this slight increase in HBV RNA levels was the consequence of an increased cccDNA transcriptional activity, we performed Run-on experiments and assessed the levels of nascent transcripts in absence of DDX5 and DDX17 in HepG2-NTCP infected cells (FigS1). Consistently with previous reported results (Zhang et al., 2016), we observed a slight increase in total HBV nascent RNA suggesting a modest increase in cccDNA transcriptional rate (FigS1). Since total HBV RNA specific primers do not allow discrimination between the different viral transcripts, Northern blot experiments using DIG-labelled HBV-specific probes were performed in PHH and HepG2-NTCP cells to characterize each viral RNA. Interestingly, all HBV transcripts migrated faster in siDDX5/DDX17 condition suggesting a shortening of the viral transcripts (Fig1E and 2E).

DDX5 and DDX17 are necessary for correct HBV RNA splicing

DDX5 and DDX17 are important regulators of co-transcriptional pre-mRNA splicing (Dardenne et al., 2014). Interestingly, several HBV spliced isoforms deriving from pgRNA have been identified (Candotti and Allain, 2016). These observations led us to hypothesize that the shortening of HBV RNA observed upon DDX5 and DDX17 depletion could be a consequence of

an altered production of some splicing variants. We therefore investigated HBV transcriptome of DDX5/17-depleted HepG2-NTCP cells by RNA-seq at 8 days post-HBV infection. All sequenced RNA samples showed good clustering in the principal component analysis (PC1 and PC2), which clearly separated the two biological replicates from each experimental condition (FigS2A). Sequencing reads were mapped to the *H.sapiens* reference genome (hg19) and to a chimeric genome consisting in the combination between *H.sapiens* and HBV (NC_003977.2) genomes.

An analysis of DDX5/17-dependent differential splicing of human transcripts was carried out using the FaRLine tool (Benoit-Pilven et al., 2018), and was validated by RT-PCR on candidate alternative exons (FigS2C). For example, we confirmed in cultured hepatocytes the previously described inclusion of *RFC5* exon 2 upon DDX5/17 silencing (Dardenne et al., 2014). We also carried out a DESeq2-based global expression analysis of human genes, as described previously (Lambert et al., 2018), and we validated experimentally the effect of DDX5/17 depletion by RT-qPCR in four independent experiments (FigS2C).

We then asked whether sequenced reads could be aligned to HBV genome. 284 000 reads were mapped to HBV genome when pooling the 4 sequenced samples. Such number of reads allowed a sufficient coverage of the virus genome to achieve a whole HBV genome sequence analysis (Fig3A). We created a novel HBV annotation file that included the previously described nomenclature for HBV spliced variants (Chen et al., 2015), allowing the first precise annotation of spliced and unspliced HBV transcripts that provides important information for HBV transcriptome analyses (Fig3B).

The development of an original tool allowing to detect and quantify *de novo* splicing junctions, combined with our novel annotation file, allowed us to investigate the influence of DDX5 and DDX17 depletion on HBV transcripts. A precise identification of 5' and 3' splice sites on HBV genome (Fig3C) uncovered the presence of differentially spliced products in siDDX5/17 condition. The splicing junctions corresponding to five distinct exonic fragments were confidently identified at a reduced frequency upon DDX5/17 depletion, with a reduced usage of their donor and acceptor sites (Fig3A). As a consequence, exon 2350-2447 from SP2, exon 2237-2447 from SP12 and exon 2350-2471 from SP16 were significantly skipped upon DDX5/17 depletion, leading to an increased switch in favor of the spliced variant SP3 (donor site: 2067; acceptor site: 489). Moreover, exon 2350-2447 from SP10 and exon 2237-2447

from SP15 were significantly spliced when DDX5 and DDX17 were silenced, inducing a switch to the SP17 isoform.

Thus, the RNA-seq data highlighted an important role of DDX5 and DDX17 in HBV RNA alternative splicing.

DDX5 and DDX17 are required for HBV RNA 3'end maintenance

RNA sequencing allowed the identification of differentially spliced isoforms in HBV transcripts, not entirely justifying the shortening of all the viral transcripts in absence of DDX5 and DDX17. The shortening of all the viral RNAs could be explained by three non-mutually exclusive phenomena: i) a delayed transcriptional initiation leading to a truncated 5' end; ii) a 3' end shortening due to the use of an alternative polyadenylation signal (PAS); iii) or a shorter polyA tail. To tackle these hypotheses, we accurately characterized the viral transcript, by performing specific HBV 5' RACE and 3' RACE, thus allowing to cover viral transcript total lengths (Fig4A). As previously described by Stadelmayer and colleagues (Stadelmayer et al., 2020a), full length 5' RACE technique allows the discrimination of each viral transcript, including the 0.7 kb RNA, not detectable in Northern-blot. 5' RACE allows the identification and the characterization of each viral transcript from its 5' initiation to a common position shared by all the transcripts and defined by a gene specific primer, positioned at nt 1584.Migration of 5' RACE products showed a similar profile in siDDX5/17 and control conditions, suggesting that the absence of DDX5/17 did not affect the 5'end of the viral transcripts (Fig4B).

Besides its role in transcriptional and alternative splicing regulation, DDX5 is part of a complex involved in transcriptional termination (Katahira et al., 2019; Mersaoui et al., 2019; Padmanabhan et al., 2012). Therefore, we then asked whether DDX5 and DDX17 could affect HBV 3' end processing. To validate this hypothesis, precise termination sites of HBV transcripts, including the length of the polyadenylated tail was evaluated by 3' RACE. 3' RACE technics relies on anchor ligation to the free 3'-OH chemical group exposed at the 3' end of all RNA chains, followed by a HBV specific retro-transcription. PCR amplification detects amplicons starting at a gene specific primer, common to all HBV transcripts at the position 1564, and ending with the ligated anchor. Agarose gel electrophoresis of 3' RACE products showed a different band profile in siDDX5/17 compared to the control condition (Fig4C).

Major bands from 400 to 700 bp are visible in the control condition, while a major smear of around 500 bp is detectable in siDDX5/17. These results suggest that DDX5 and DDX17 are involved in the 3' end HBV RNA processing and that HBV transcripts seem more homogeneous in their 3' termini upon DDX5/17 silencing compared to the non-depleted conditions.

DDX5 and DDX17 are decisive for the choice of the polyadenylation site

To further characterize HBV RNA 3' end in absence of DDX5 and DDX17, 3' RACE products were sequenced by 3rd generation sequencing, using MinION™ Nanopore technology. For each condition, more than 150 000 reads were obtained among which approximately 90 000 aligned to the reference HBV genome. As expected, when combining all the results, in control conditions, most of the transcripts stop at the canonical PAS, namely 5'-TATAAA-3' (position 1916 to 1921). For the first time, we demonstrated that the exact position of transcription end is located 13 +/- 1nucleotides downstream of the PAS. However, the coverage revealed that approximately half of the transcripts do not end at the canonical PAS, but present an extensive 3' readthrough, running along approximately 560 nucleotides (Fig4D and FigS3A). Strikingly, the 3' readthrough almost completely disappeared upon DDX5 and DDX17 depletion (Fig4D), where almost 100% of the transcripts end at canonical PAS. It is important to note that 3' RACE product sequencing does not allow the discrimination of the different HBV transcripts.

DDX5 and DDX17 recruit CPSF6 and NUDT21 on HBV RNA

Alternative polyadenylation (APA) site usage is occurring in 70% of mammal genes (Neve et al., 2017). While APA can occur between two PASs allowing for generation of transcripts with distinct coding potential from a single gene, most APA occurs within the untranslated region (3'UTR) and changes the length and content of these non-coding sequences. We then explored whether potential alternative PAS were present along the readthrough sequence. We first confirmed that the described canonical HBV PAS was conserved within the major HBV genotypes using HBVdb database (FigS3B) and then carried out *in sillico* analysis with PAS prediction tool (PolyApred, http://crdd.osdd.net/raghava/polyapred/). No putative PAS among the most represented ones in human genome was identified in HBV RNA 3' readthrough sequence (FigS3C). However, we identified two 5'-UGUA-3' motifs downstream of the canonical PAS. These motifs are auxiliary sequence elements recognized by members of the Cleavage Factor I (CFI) complex, a key regulator of polyadenylation site choice during

APA. Notably, CFI subunits Cleavage and polyadenylation specificity factor subunit 6 (CPSF6) and nudix hydrolase 21 (NUDT21) bind the 5'-UGUA-3' motif localized in the 3' UTR of most of the pre-mRNA. CPSF6 and NUDT21 enhance the 3' processing at distal PAS (Hardy and Norbury, 2016). These observations led us to hypothesize that CPSF6 and NUDT21 could be responsible for the 3' readthrough observed in control conditions.

Both DDX5 and DDX17 have been demonstrated to directly associate with cellular RNA (Fuller-Pace, 2006, p.). DDX5 and DDX17 implication in HBV RNA processing drove the hypothesis of an association to HBV transcripts. We therefore investigated the recruitment of these two factors on HBV RNA in HepG2-NTCP cells infected with HBV at a MOI of 250 during 7 days. By UV cross-linking immunoprecipitation (CLIP) we assessed the recruitment of DDX5 and DDX17 directly on viral RNAs. CLIP-RT-qPCR experiments revealed that DDX5 and DDX17 bind HBV RNAs when using primers recognizing all HBV transcripts (total HBV RNA) and when using primers specific for 3.5 kb RNA (Fig6A). Interestingly, DDX5 and DDX17 also associate to cccDNA both in HBV-infected HepG2-NTCP cells and in PHH and suggest that DDX5 and DDX17 are located at the transcriptional interface between cccDNA template and HBV transcripts (FigS4). Then, we investigated the direct binding of CPF6 and NUDT21 on HBV RNA in siCTL- or siDDX5/17-transfected HBV-infected HepG2-NTCP cells. CLIP-RT-qPCR experiments revealed a strong decrease in CPSF6 and NUDT21 recruitment onto HBV 3.5 kb and total RNA upon DDX5 and DDX17 silencing (Fig6C). This decrease is not accompanied by any differences in total CPF6 and NUDT21 mRNA levels upon DDX5/17 silencing, as shown by our RNA-Seq data (data not shown). Altogether, these data strongly suggested a direct role of DDX5 and DDX17 in HBV transcription termination regulation by recruiting CPSF6 and NUDT21 directly on the viral transcripts.

DISCUSSION

HBV cccDNA is a stable episome with a chromatin structure regulated by epigenetic mechanisms and the unique template for full viral lifecycle. Current antiviral treatments inhibit viral replication, but do not significantly affect cccDNA reservoir in the hepatocyte nuclei. As a consequence, cccDNA is responsible for HBV persistence in chronic infections, which can develop in sever liver diseases and still represent a major worldwide health burden.

Overlapping open reading frames and regulatory elements on HBV genome have to be precisely and timely controlled to ensure correct cccDNA transcription, termination and viral mRNA processing. In this study, we identified RNA helicases DDX5 and DDX17 as novel regulators of HBV RNA processing, describing a shortening of HBV RNAs upon depletion of the helicases in infected cells. DDX5/17 are master regulators of alternative splicing during human transcription (Bourgeois et al., 2016; Dardenne et al., 2014). Several studies uncovered spliced isoforms deriving from HBV pgRNA using semiquantitative PCR amplification using primers flanking the donor and acceptor splicing sites, thus discriminating between the different variants only by size, or by full-genome reverse PCR amplification of encapsidated DNA, assuming that defective molecules were originating from spliced RNAs (Bayliss et al., 2013; Chen et al., 2015; Duriez et al., 2017). We used a non-biased technique to investigate HBV RNA splicing by RNA sequencing. RNAseq has already been used for detecting viral transcripts in infectious models, with the main limitation that, when cccDNA transcription is active, the short read sequences cannot be allocated to a particular transcript, due to the overlapping nature of HBV RNAs (Niu et al., 2017). For the first time, we used RNAseq-generated data for unbiased identification of splice junctions on HBV RNAs. We identified splice junctions, corresponding to splice variants, which have been annotated according to previously published literature (Chen et al., 2015). Under DDX5/17 depletion, a less usage of specific donor/acceptor sites was highlighted. Splice junction skipping from position 2337 to 2447 from SP12, 2350 to 2447 from SP2 and 2350 to 2471 from SP16 leads to variant switch to SP3 isoform (donor site: 2067; acceptor site: 489). SP3 isoform constitutes 70% of spliced variants in circulating HBV particles, together with SP1 and SP2 (Chen et al., 2015). However, their implication in HBV replication remains to be elucidated. Moreover, exon skipping from position 2350 to 2447 from SP10 and 2237 to 2447 from SP15 leads to a switch to SP17 variant (donor site: 2067; acceptor site: 282). SP17 is a newly identified spliced isoforms, described

for the first time by Chen and colleagues in the serum of CHB patients (Chen et al., 2015). Interestingly, pgRNA splicing has been correlated with HBV pathogenesis, and notably with HCC development (Bayliss et al., 2013; Soussan et al., 2003). However, exact identification of spliced isoforms positively associated with HCC development was not possible due to the technical limitations of using primers flanking the splice donor/acceptor sites common to several variants. Since DDX5 seems to be downregulated in a subset of HBV-associated tumors (Zhang et al., 2016), investigating functions of the favored SP3 and SP17 variants in the absence of DDX5/17 could bring further information on HBV-related pathogenesis and tumorigenesis.

To fully characterize HBV transcripts upon DDX5/17 silencing, we combined 5' RACE and 3' RACE methods to cover the entire HBV RNA length. 5' RACE adapted for HBV transcripts was set up in our team (Stadelmayer et al., 2020a) and allows the precise identification of the main viral transcripts. 3' RACE allows precise identification of 3' end of viral transcripts. We uncovered phenotype changes in the 3' end of viral transcripts. Deeper characterization was performed by single-molecule sequencing using the 3rd generation sequencing with MinION^{TM} Nanopore technology. We showed that HBV transcripts precisely end 13 +/-1 nucleotides downstream of HBV PAS 5'-TATAAA-3' (position 1916-1921). Moreover, we uncovered that almost 50% of HBV transcripts in control conditions harbor an extensive 3' readthrough, which is completely abolished in DDX5/17 depleted condition, where all HBV transcripts end at canonical PAS. This phenomenon could explain the phenotype of "RNA shortening" that we saw by Northern Blot analysis. Given the fact that DDX5 and DDX17 were already suggested to be part of the termination and export complex THO, together with CPSF6 (Katahira et al., 2019), and were involved in transcription termination process (Mersaoui et al., 2019; Padmanabhan et al., 2012), we hypothesized that the 3'readthrough could be ascribed to an alternative polyadenylation (APA) mechanisms. Despite the 3' readthrough sequence did not contain any additional PAS, it contained 2 UGUA motifs, which are auxiliary sequences recognized by cleavage factors involved in transcription termination, namely CPSF6 and NUDT21. By CLIP, we demonstrated a direct recruitment of CPSF6 and NUDT21 to HBV transcripts, which is abolished when DDX5/17 are downregulated, suggesting a role for the helicases in promoting the recruitment of factors involved in PAS choice, notably by favoring the use of distal PAS (Hardy and Norbury, 2016). However, the precise binding site of these

factors on HBV transcripts remains to be elucidated to confirm their role in HBV RNA 3' end processing. Depletion of CPSF6 and NUDT21 and investigation of consequent viral parameters, in particular HBV RNA transcription termination, would bring to light a potential direct role of such factors in viral co-transcriptional processes.The biological significance of HBV RNA 3' readthrough remains to be elucidated. The 3' UTR significantly determines the stability, localization, translation and degradation of mRNA. Approximately half of mammalian genes use alternative cleavage and polyadenylation to generate multiple mRNA isoforms differing in their 3'UTRs. Mirroring mammalian genome transcription, we could speculate that longer HBV transcripts may contain target sequences for microRNAs, thus adding a layer of complexity to the post-transcriptional regulation of HBV RNAs. Notably, by bioinformatics prediction tools, we found target sequences for miR222, miR500 and miR548, all demonstrated to be modulated in HCC patients (Bao et al., 2018; Hu et al., 2016; Ogawa et al., 2012). The length of 3' UTRs seems to be correlated to the proliferative status of the cell, being shorter in actively dividing cells and highly proliferative tumors (Curinha et al., 2014). We could speculate that when HBV infects terminally differentiated hepatocytes, its transcription would be submitted to the preferential use of distal PAS, as expected in non-dividing cells. After HBV infection, we and others (Zhang et al., 2016) observe a decrease in DDX5/17 protein expression and this could be instrumental for the virus to limit the 3' UTR lengthening and optimize viral RNA processing rate and protein translation.

To conclude, we identified DDX5 and DDX17 as regulators of HBV RNA processing and demonstrated that HBV transcripts do not systematically end at the expected PAS. Rather, HBV transcripts are lengthened in their 3' end. This observation has to be deeper investigated to better understand processes at the basis of cccDNA transcriptional activity and co-transcriptional regulation. Indeed, development of accurate non-invasive biomarkers is crucial to precisely reflect cccDNA transcriptional activity and to further develop new treatments toward a cure of HBV. Further investigations are required to elucidate the role of such 3' readthrough in HBV replication. Notably, confirming these observations in HBV-infected patients is crucial. Indeed, uncovering such readthrough in CHB patients would question on the accuracy of current available biomarkers for cccDNA transcriptional activity.

MATERIAL AND METHODS

PHH Isolation

Primary human hepatocytes (PHHs) were isolated from surgical liver resections, after informed consent of patients (IRB agreements #DC-2008-99 and DC-2008-101) as previously described (LeCluyse and Alexandre, 2010) and plated in complete William's supplemented with 1% penicillin/streptomycin (Life Technology), 1% glutamine (Life Technology), 5 µg/ml Human insµlin (Sigma-Aldrich), 25 µg/mL Hydrocortisone hemisuccinate UPJOHN (SERB), and 5% Fetal Calf Serum (FCS; Fetalclone IITM, PERBIO). PHH were maintained in William's medium supplemented with 1.8% DMSO (Sigma, St Quentin, France). All PHH-related data were obtained from at least two distinct patients. All experiments on PHH were prerformed as described for HepG2-NTCP.

Cell culture and HBV infection

HepG2-NTCP cells were seeded at 10^5 cells/cm^2 in DMEM medium supplemented with 1% penicillin/streptomycin (Life Technology), 1% sodium pyruvate (Life Technology), 1% glutamine (Life Technology), 5% Fetal Calf Serum (FCS; Fetalclone II™, PERBIO). The day after, medium was renewed and complemented with 2.5% DMSO (SIGMA). After 72h, cells were infected at a mµltiplicity of infection of 250 in the presence of 4% PEG800 for up to 16h and then extensively washed with PBS and maintained in complete DMEM medium containing 2.5% DMSO until harvesting. Intracellµlar accumµlation of viral RNA and DNA, were monitored by RT-qPCR, qPCR, Northern-blotting and Southern blotting. Total DNA was purified from infected cells using MasterPure™ Complete DNA Purification Kit (Epicentre). Total RNA was extracted using ExtractAll TRI-Reagent (MRC), precipitated in isopropanol, washed in ethanol and resuspended in RNase-free water.

siRNA transfection

siRNA (siDDX5 catalog# 4392420; siDDX17 catalog# 4390824; ON-TARGETplus LUC siRNA, Dharmacon) were transfected either concomitantly (12.5 nM siDDX5, 12.5 nM siDDX17), or alone (25 nM siRNA control: siLUC) in HepG2-NTCP cells and PHH four days post-infection, using Lipofectamine RNAiMAX (ThermoFisher), following manufacturer's protocol. The day after transfection, culture medium was changed, and cells were transfected a second time six days post-infection. The culture medium was changed the day after the second transfection. The supernatants were collected and cells were harvested two days after the second transfection.

qPCR and RT-qPCR

Extracted RNA were digested with RNAse-free DNase I (Qiagen) and retro-transcribed into cDNA using SuperScript III reverse transcriptase according to manufacturer's instructions (Invitrogen, Carlsbad, USA). cccDNA was quantify after ExoI + ExoIII endonuclease (Epicentre) digestion of total extracted DNA for 2 hours at 37°C, followed by 20 minutes inactivation at 80°C. Real-time qPCR for total HBV RNA/DNA and cccDNA was performed using an Applied QuantStudio 7 machine (BioSystem) and TaqMan Advanced Fast Master Mix or SYBR Green Master Mix (see primers and probes below, table 1). Serial dilutions of a plasmid containing an HBV monomer (pHBV-EcoR1) served as quantification standard for HBV DNA and cccDNA. The number of cellular genomes was determined by using the β-globin gene kit (Roche

Diagnostics, Manheim, Germany). RT-qPCR were analyzed using the ΔΔCt method were ΔCt = Ct (target) − Ct (GUS), where GUS is the housekeeping gene.

Western-Blot

Cells were lysed in RIPA buffer supplemented with 1X PIC and 1X PMSF. 20 µg of protein were migrated in 4-20% mini-PROTEAN@ TGX stain-Free™ Precast Gel (Bio-Rab Laboratories) and transferred onto a nitrocellulose membrane (Bio-Rab Laboratories). Membranes were blocked 1 hour with 5% milk in TBS (1 x Tris Buffer Saline (Sigma)) and stained with primary antibodies in blocking buffer overnight at 4 °C (see Table XX). After primary antibody incubation, membranes were washed 3 times with TBS-T 0.1% (1X TBS with 0.1% Tween 20), stained with HRP-conjugated secondary antibodies (1/5000) for 1 hour at room temperature and washed again 3 times with TBS-T 0.1%. The detection was done using Clarity Western ECL and the ChemiDoc XRS system (Biorad).

Northern-Blot

Purified RNA were denatured at 50°C for one hour with glyoxal reagent (Life Technologies) and then subjected to electrophoresis through 1X phosphate buffer in a 1.2% agarose gel and transferred to a positively charged nylon membrane (Amersham N+, GE). Membrane-bound RNA were hybridized overnight at 42°C to DIG-labeled HBV-specific probes (Table 1). The membrane was washed twice in low-stringency wash buffer (1X SSC, 0.1% SDS) for 30 min at room temperature and twice in high-stringency wash buffer (0.1X SSC, 0.1% SDS) for 30 min at 65°C. Detection was performed using anti-DIG alkaline phosphatase (1/20 000) and CDP-Star reagent (Roche) according to the manufacturer's recommendation and imaged using ChemiDoc MP imaging system (BioRad). 18S and 28S rRNA were used as loading and quality control.

Nascent RNA Capture

Cultured cells were pulsed for 2 hours with 0.2 mM Ethynil-Uridine. Total RNA were extracted as above. 1 µg of RNA was used to perform nascent RNA capture using Click-iT® Nascent RNA Capture Kit (ThermoFisher), according to the manufacturer's instructions. Precipitated RNA were retro-transcribed using Superscript IV Vilo (ThermoFisher), and qPCR was performed as described above.

5'RACE

5'RACE was performed as previously described in (Stadelmayer et al., 2020a).

3'RACE

RNA were extracted as described above. 1 µg of extracted RNA was digested with RNAse-free DNase I (Qiagen) and purified with acid phenol (pH 4.3), and then precipitated with 100% ethanol, 10% Sodium Acetate (Sigma), and 1 µl Glycogen, overnight at -20°C, then washed with 70% ethanol and resuspended in water. A 3'-DNA-Adaptor Ligation was performed with T4 RNA ligase 2 truncated (NEB) in presence of RNase-OUT (ThermoFisher) for 1 hour at 25°C. After acid phenol precipitation, gene specific reverse-transcription of anchored RNA was performed using ProtoScript II (NEB). The final PCR reaction was done with Platinum SuperFi II DNA Polymerase (ThermoFisher) (see primers Table X). PCR products were deposited and migrated on a 1% ultrapure agarose gel.

RNA-seq

RNA was extracted as above. Extracted RNA was digested with RNAse-free DNase I (Qiagen) and tested for RNA quality control using a Bioanalyzer Instrument (Agilent). mRNA libraries were prepared using the TruSeq Stranded Total RNA Library Prep Gold (Illumina) kit to increase the depth of sequencing. Each library was paired-end sequenced to an average of 50 million reads per sample using 75x2 cycles on a NextSeq 500 Illumina platform. On average, approximately 90 million reads were generated per sample.

Sequencing reads were mapped to the H.sapiens reference genome (hg19) and to a merged artificial genome combining H. sapiens and HBV (NC_003977.2)

For the analyses of human reads, the raw reads were mapped on GRCh37.75 EnsEMBL reference annotation by Tophat2 (Kim et al., 2013) (v2.0.13). Only reads mapping to exonic junctions were used analysed using the FaRLine pipeline (Benoit-Pilven et al., 2018) and the FasterDB database (Mallinjoud et al., 2014). The reads below a mapping score of 10 or mapped to multiple locations were filtered out using samtools (Anders et al., 2015) (v0.1.19). The gene expression level in each sample was analysed as described previously (Lambert et al., 2018).

For the HBV analysis, a novel progenomic reference that started with preC and ended with the full HBV sequence was derived from NC_003977.2. We concatenated the segments (nt) 1820-3182 and 1-1932 of NC_003977.2. Consequently, former segment (nt) 1820-1932 is represented twice. Raw reads were mapped on the progenomic reference with bowtie2 (Langmead and Salzberg, 2012) in the local mode (option --local). The resulting mapping file was analyzed with our home-made tool ClipStick (unpublished) to detect all splicing events *de novo*. A GTF file compiling all known and new splicing events was created. Alternative splicing of the different segments was calculated using STAR (v2.7.3) (Dobin et al., 2013) and rMATS (v3.1.0) (Shen et al., 2014). The level of HBV transcripts was estimated with kallisto (v0.44.0) (Bray et al., 2016).

Nanopore

DNA from 3'RACE PCR and products of Complete genome (CG) amplification performed with Platinum™ *Taq* DNA Polymerase High Fidelity (Invitrogen) (see primer sequences table X) were prepared for Nanopore Sequencing with SQK-PBK004 kit, according to the manufacturer's instructions. Analysis was performed with Epi2me (epi2me.nanoporetech.com), NanoPipe (http://bioinformatics.uni-muenster.de), and aligned to a reference HBV genome, corresponding to full-length genome starting at forward PCR primer (position 1564). (Fig Ref genomes)

Cross-Link Immunoprecipitation

RIP experiments were carried out 7 days post-infection. Cells were washed in 1X PBS and UV cross-linked at 200 mJ/cm^2. Total cell lysate was obtained by 20 min RIPA lysis buffer incubation on ice, followed by sonication. Total cell lysate was pre-cleaned in RIPA buffer with Protein G-coupled magnetic beads (Dynabeads, Invitrogen), and then subjected to overnight immunoprecipitation at 4 °C using 2–4 µg of antibodies (see Table 2). Chromatin incubated with no antibody or anti-H3.3 antibody was processed as other conditions and was used as negative controls. Immune complexes were then incubated 2 hours with Protein G-coupled

magnetic beads at 4 °C, washed, and eluted in ChIP elution buffer. The flow-through from the no antibody condition was used as input. Immunoprecipitated RNA were extracted using ExtractAll TRI-Reagent (MRC), precipitated in isopropanol, washed in ethanol and resuspended in RNase-free water. RT-qPCR was performed as described above. Samples were normalized to input RNA using the ΔΔCt method were ΔCt = Ct (input) − Ct (immunoprecipitation) and calculated as percentage of the input.

Chromatin Immunoprecipitation

ChIP experiments were carried out 7 days post-infection. Cells were washed in 1X phosphate buffered saline (PBS), and incubated for 15 minutes with 1% formaldehyde at 37 °C and quenched with 0.125 M Glycine for 10 minutes at 37°C. For nuclear extracts preparation, cells were lysed in lysis buffer (5 mM PIPES, 85 mM KCl, 0,5% NP-40, 1 mM PMSF, 1X Protease inhibitor cocktail (PIC) (ThermoFisher)). After "tight" douncing (10 times) and 5 minutes centrifugation at 3000 rpm, nuclei were resuspended in sonication buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH8, 1 mM PMSF, 1X PIC). After sonication, chromatin was pre-cleaned in RIPA buffer with Protein G-coupled magnetic beads (Dynabeads, Invitrogen), and then subjected to overnight immunoprecipitation at 4 °C using 2–4 µg of antibodies (see Table 2). Chromatin incubated with no antibody or anti-E2F antibody was processed as other conditions and was used as negative controls. Immune complexes were then incubated 2 hours with Protein G-coupled magnetic beads at 4 °C, washed, and eluted in 10 mM Tris-HCl pH8, 5 mM EDTA, 50 mM NaCl, 1% SDS, 50 µg Proteinase K, 1X PIC. The flow-through from the no antibody condition was used as input. Immunoprecipitated DNA was extracted with phenol-chloroform isoamyl (25:24:1) and quantified by qPCR using cccDNA specific primers (see table 1). Samples were normalized to input DNA using the ΔΔCt method were ΔCt = Ct (input) − Ct (immunoprecipitation) and calculated as percentage of input.

Target	Primer Forward (5'-3')	Primer Reverse (5'-3')
cccDNA	CCGTGTGCACTTCGCTTCA	GCACAGCTTGGAGGCTTGA
3.5 Kb RNA	GGAGTGTGGATTCGCACTCCT	AGATTGAGATCTTCTGCGAC
total HBV DNA/RNA	GCTGACGCAACCCCCACT	AGGAGTTCCGCAGTATGG
DDX5	AGAGAGGCGATGGGCCTATTT	CTTCAAGCGACATGCTCTACAA
DDX17	GATGTTTGTCCTAAACCCGTGT	CCAACGGAAATCCCTGGCA
GUS	CGTGGTTGGAGAGCTCATTTGGAA	TTCCCCAGCACTCTCGTCGGT
GC PCR	CCACCGAATGTTGCCCAAGG	GCGTTCACGGTGGTCTCCAT
3' DNA-Adaptor	[5rApp]CACTGTCATGCCGTTACGTAGCGTATCGTTGACAGC[3ddC]	
3'RACE RT	GCTGTCAACGATACGCTACGTAACG	
3'RACE PCR	CATCTGCCGGACCGTGTGCAC	CGATACGCTACGTAACGGCATGAC
	Taqman Probes	
cccDNA	[6FAM]CATGGAGACCACCGTGAACGCCC[BBQ]	
3.5 Kb RNA	[6FAM]AGGCAGGTCCCCTAGAAGAAGAACTCC[BBQ]	
	DIG-Labeled Probes	
	Forward (5'-3')	**Reverse (5'-3')**

HBV-1	TAGCGCCTCATTTTGTGGGT	CTTCCTGTCTGGCGATTGGT
HBV-2	TAGGACCCCTTCTCGTGTTA	CCGTCCGAAGGTTTGGTACA
HBV-3	ATGTGGTATTGGGGGCCAAG	GGTTGCGTCAGCAAACACTT
HBV-4	TGGACCTTTTCGGCTCCTC	GGGAGTCCGCGTAAAGAGAG
HBV-5	GTCTGTGCCTTCTCATCTG	AGGAGACTCTAAGGCTTCC
HBV-6	TACTGCACTCAGGCAAGCAA	TGCGAATCCACACTCCGAAA
HBV-8	AGACGAAGGTCTCAATCGCC	ACCCACAAAATGAGGCGCTA
Mitochondrial Fw-huND1	CCCTACTTCTAACCTCCCTGTTCTTAT	CATAGGAGGTGTATGAGTTGGTCGTA
Mitochondrial Fw-huND5	ATTTTATTTCTCCAACATACTCGGATT	GGGCAGGTTTTGGCTCGTA
Mitochondrial Fw-huND6	CATTTACACCAACCACCCAACTATC	CGAAAGCCTATAATCACTGTGCC

Table 1: Primers and probes

Antibody	**Quantity/Dilution**	**Reference**
	Immunoprecipitation	
HBc	2 ug	Thermo
DDX5	4 ug	Bethyl A300-523A
DDX17	4 ug	Proteintech 19910-1-AP
E2F	x ug	
	Western-Blotting	
HBc	1/500	abcam ab115992
DDX5	1/7500	Bethyl A300-523A
DDX17	1/4000	Proteintech 19910-1-AP
β-Actin	1/10 000	Abcam ab6276

Table 2: Antibodies

Acknowledgement

We thank M. Michelet and J. Molle for Primary Human Hepatocyte isolation. Chloé Goldsmith for help in Nanopore data analysis. We also gratefully aknowledge support from the PSMN (Pôle Scientifique de Modélisation Numérique) of the ENS de Lyon for computing ressource.

REFERENCES

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. https://doi.org/10.1093/bioinformatics/btu638

Bao, L., Zhang, M., Han, S., Zhan, Y., Guo, W., Teng, F., Liu, F., Guo, M., Zhang, L., Ding, G., Wang, Q., 2018. MicroRNA-500a Promotes the Progression of Hepatocellular Carcinoma by Post-Transcriptionally Targeting BID. Cell Physiol Biochem 47, 2046–2055. https://doi.org/10.1159/000491472

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. Journal of Hepatology 59, 1022–1028. https://doi.org/10.1016/j.jhep.2013.06.018

Benoit-Pilven, C., Marchet, C., Chautard, E., Lima, L., Lambert, M.-P., Sacomoto, G., Rey, A., Cologne, A., Terrone, S., Dulaurier, L., Claude, J.-B., Bourgeois, C.F.,euf, D., Lacroix, V., 2018. Complementarity of assembly-first and mapping-first approaches for alternative splicing annotation and differential analysis from RNAseq data. Sci Rep 8, 4307. https://doi.org/10.1038/s41598-018-21770-7

Bock, C.T., Schwinn, S., Locarnini, S., Fyfe, J., Manns, M.P., Trautwein, C., Zentgraf, H., 2001. Structural organization of the hepatitis B virus minichromosome. Journal of Molecular Biology 307, 183–196. https://doi.org/10.1006/jmbi.2000.4481

Bourgeois, C.F., Mortreux, F., Auboeuf, D., 2016a. The multiple functions of RNA helicases as drivers and regulators of gene expression. Nature Reviews Molecular Cell Biology 17, 426–438. https://doi.org/10.1038/nrm.2016.50

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34, 525–527. https://doi.org/10.1038/nbt.3519

Candotti, D., Allain, J.-P., 2016a. Biological and clinical significance of hepatitis B virus RNA splicing: an update. Annals of Blood 2, 6–6. https://doi.org/10.21037/aob.2017.05.01

Chen, J., Wu, M., Wang, F., Zhang, W., Wang, W., Zhang, X., Zhang, J., Liu, Yinghui, Liu, Yi, Feng, Y., Zheng, Y., Hu, Y., Yuan, Z., 2015. Hepatitis B virus spliced variants are associated with an impaired response to interferon therapy. Sci Rep 5, 16459. https://doi.org/10.1038/srep16459

Curinha, A., Oliveira Braz, S., Pereira-Castro, I., Cruz, A., Moreira, A., 2014. Implications of polyadenylation in health and disease. Nucleus 5, 508–519. https://doi.org/10.4161/nucl.36360

Dardenne, E., Polay Espinoza, M., Fattet, L., Germann, S., Lambert, M.-P., Neil, H., Zonta, E., Mortada, H., Gratadou, L., Deygas, M., Chakrama, F.Z., Samaan, S., Desmet, F.-O., Tranchevent, L.-C., Dutertre, M., Rimokh, R., Bourgeois, C.F., Auboeuf, D., 2014. RNA Helicases DDX5 and DDX17 Dynamically Orchestrate Transcription, miRNA, and Splicing Programs in Cell Differentiation. Cell Reports 7, 1900–1913. https://doi.org/10.1016/j.celrep.2014.05.010

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Duriez, M., Mandouri, Y., Lekbaby, B., Wang, H., Schnuriger, A., Redelsperger, F., Guerrera, 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. Journal of Hepatology 67, 687–699. https://doi.org/10.1016/j.jhep.2017.05.025

Fuller-Pace, F.V., 2006. DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. Nucleic Acids Research 34, 4206–4215. https://doi.org/10.1093/nar/gkl460

Giraud, G., Terrone, S., Bourgeois, C.F., 2018. Functions of DEAD box RNA helicases DDX5 and DDX17 in chromatin organization and transcriptional regulation. BMB Reports 51, 613–622. https://doi.org/10.5483/BMBRep.2018.51.12.234

Hardy, J.G., Norbury, C.J., 2016. Cleavage factor Im (CFIm) as a regulator of alternative polyadenylation. Biochemical Society Transactions 44, 1051–1057. https://doi.org/10.1042/BST20160078

Hu, X.-M., Yan, X.-H., Hu, Y.-W., Huang, J.-L., Cao, S.-W., Ren, T.-Y., Tang, Y.-T., Lin, L., Zheng, L., Wang, Q., 2016. miRNA-548p suppresses hepatitis B virus X protein associated hepatocellular carcinoma by downregulating oncoprotein hepatitis B x-interacting protein: miR-548p suppresses HBx-HCC by inhibiting HBXIP. Hepatol Res 46, 804–815. https://doi.org/10.1111/hepr.12618

Jeong, J.-K., Yoon, G.-S., Ryu, W.-S., 2000. Evidence that the 5J-End Cap Structure Is Essential for Encapsidation of Hepatitis B Virus Pregenomic RNA. J. VIROL. 74, 7.

Katahira, J., Ishikawa, H., Tsujimura, K., Kurono, S., Hieda, M., 2019. Human THO coordinates transcription termination and subsequent transcript release from the HSP70 locus. Genes Cells 24, 272–283. https://doi.org/10.1111/gtc.12672

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36. https://doi.org/10.1186/gb-2013-14-4-r36

Lambert, M.-P., Terrone, S., Giraud, G., Benoit-Pilven, C., Cluet, D., Combaret, V., Mortreux, F., Auboeuf, D., Bourgeois, C.F., 2018. The RNA helicase DDX17 controls the transcriptional activity of REST and the expression of proneural microRNAs in neuronal differentiation. Nucleic Acids Research 46, 7686–7700. https://doi.org/10.1093/nar/gky545

Lamm, G., 1996. p72: a human nuclear DEAD box protein highly related to p68. Nucleic Acids Research 24, 3739–3747. https://doi.org/10.1093/nar/24.19.3739

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

LeCluyse, E.L., Alexandre, E., 2010. Isolation and Culture of Primary Hepatocytes from Resected Human Liver Tissue, in: Maurel, P. (Ed.), Hepatocytes. Humana Press, Totowa, NJ, pp. 57–82. https://doi.org/10.1007/978-1-60761-688-7_3

Lok, A.S., Zoulim, F., Dusheiko, G., Ghany, M.G., 2017. Hepatitis B cure: From discovery to regulatory approval: Lok et al. Hepatology 66, 1296–1313. https://doi.org/10.1002/hep.29323

Lorgeoux, R.-P., Pan, Q., Le Duff, Y., Liang, C., 2013. DDX17 promotes the production of infectious HIV-1 particles through modulating viral RNA packaging and translation frameshift. Virology 443, 384–392. https://doi.org/10.1016/j.virol.2013.05.026

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. Journal of Hepatology 55, 996–1003. https://doi.org/10.1016/j.jhep.2011.02.015

Mallinjoud, P., Villemin, J.-P., Mortada, H., Polay Espinoza, M., Desmet, F.-O., Samaan, S., Chautard, E., Tranchevent, L.-C., Auboeuf, D., 2014. Endothelial, epithelial, and fibroblast cells exhibit specific splicing programs independently of their tissue of origin. Genome Research 24, 511–521. https://doi.org/10.1101/gr.162933.113

Mersaoui, S.Y., Yu, Z., Coulombe, Y., Karam, M., Busatto, F.F., Masson, J., Richard, S., 2019. Arginine methylation of the DDX 5 helicase RGG / RG motif by PRMT 5 regulates resolution of RNA:DNA hybrids. EMBO J 38. https://doi.org/10.15252/embj.2018100986

Moolla, N., Kew, M., Arbuthnot, P., 2002. Regulatory elements of hepatitis B virus transcription. Journal of Viral Hepatitis 9, 323–331. https://doi.org/10.1046/j.1365-2893.2002.00381.x

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

Neugebauer, K.M., 2002. On the importance of being co-transcriptional. Journal of Cell Science 115, 3865–3871. https://doi.org/10.1242/jcs.00073

Neve, J., Patel, R., Wang, Z., Louey, A., Furger, A.M., 2017. Cleavage and polyadenylation: Ending the message expands gene regulation. RNA Biology 14, 865–890. https://doi.org/10.1080/15476286.2017.1306171

Niu, C., Livingston, C.M., Li, L., Beran, R.K., Daffis, S., Ramakrishnan, D., Burdette, D., Peiser, L., Salas, E., Ramos, H., Yu, M., Cheng, G., Strubin, M., Delaney IV, W.E., Fletcher, S.P., 2017. The Smc5/6 Complex Restricts HBV when Localized to ND10 without Inducing an Innate Immune Response and Is Counteracted by the HBV X Protein Shortly after Infection. PLoS ONE 12, e0169648. https://doi.org/10.1371/journal.pone.0169648

Ogawa, T., Enomoto, M., Fujii, H., Sekiya, Y., Yoshizato, K., Ikeda, K., Kawada, N., 2012. MicroRNA-221/222 upregulation indicates the activation of stellate cells and the progression of liver fibrosis. Gut 61, 1600–1609. https://doi.org/10.1136/gutjnl-2011-300717

Padmanabhan, K., Robles, M.S., Westerling, T., Weitz, C.J., 2012. Feedback Regulation of Transcriptional Termination by the Mammalian Circadian Clock PERIOD Complex. Science 337, 599–602. https://doi.org/10.1126/science.1221592

Patient, R., Hourioux, C., Sizaret, P.-Y., Trassard, S., Sureau, C., Roingeard, P., 2007. Hepatitis B Virus Subviral Envelope Particle Morphogenesis and Intracellular Trafficking. Journal of Virology 81, 3842–3851. https://doi.org/10.1128/JVI.02741-06

Pentland, I., Parish, J., 2015. Targeting CTCF to Control Virus Gene Expression: A Common Theme amongst Diverse DNA Viruses. Viruses 7, 3574–3585. https://doi.org/10.3390/v7072791

Rall, L.B., Standring, D.N., Laub, O., Rutter, W.J., 1983. Transcription of hepatitis B virus by RNA polymerase II. Molecular and Cellular Biology 3, 1766–1773. https://doi.org/10.1128/MCB.3.10.1766

Rat, V., Seigneuret, F., Burlaud-Gaillard, J., Lemoine, R., Hourioux, C., Zoulim, F., Testoni, B., Meunier, J.-C., Tauber, C., Roingeard, P., de Rocquigny, H., 2019. BAY 41-4109-mediated aggregation of assembled and misassembled HBV capsids in cells revealed by electron microscopy. Antiviral Research 169, 104557. https://doi.org/10.1016/j.antiviral.2019.104557

Seeger, C., Mason, W.S., 2000a. Hepatitis B Virus Biology. Microbiology and Molecular Biology Reviews 64, 51–68. https://doi.org/10.1128/MMBR.64.1.51-68.2000

Seeger, C., Mason, W.S., 2000b. Hepatitis B Virus Biology. Microbiology and Molecular Biology Reviews 64, 51–68. https://doi.org/10.1128/MMBR.64.1.51-68.2000

Shen, S., Park, J.W., Lu, Z., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., Xing, Y., 2014. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci USA 111, E5593–E5601. https://doi.org/10.1073/pnas.1419161111

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., Oberdoerffer, S., 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 479, 74–79. https://doi.org/10.1038/nature10442

Sithole, N., Williams, C.A., Vaughan, A.M., Kenyon, J.C., Lever, A.M.L., 2018. DDX17 Specifically, and Independently of DDX5, Controls Use of the HIV A4/5 Splice Acceptor Cluster and Is Essential for Efficient Replication of HIV. Journal of Molecular Biology 430, 3111–3128. https://doi.org/10.1016/j.jmb.2018.06.052

Soussan, P., Pol, J., Garreau, F., Schneider, V., Pendeven, C.L., Nalpas, B., Lacombe, K., Bonnard, P., Pol, S., Kremsdorf, D., 2008. Expression of Defective Hepatitis B Virus Particles Derived from Singly Spliced RNA Is Related to Liver Disease. The Journal of Infectious Diseases 198, 218–225. https://doi.org/10.1086/589623

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. Journal of Hepatology 38, 343–348. https://doi.org/10.1016/S0168-8278(02)00422-1

Stadelmayer, B., Diederichs, A., Chapus, F., Rivoire, M., Neveu, G., Alam, A., Fraisse, L., Carter, K., Testoni, B., Zoulim, F., 2020. Full-length 5'RACE identifies all major HBV transcripts in HBV-

infected hepatocytes and patient serum. Journal of Hepatology. https://doi.org/10.1016/j.jhep.2020.01.028

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

Wenyu Cheng, Guohua Chen, Huaijie Jia, Xiaobing He, Zhizhong Jing, 2018. DDX5 RNA Helicases: Emerging Roles in Viral Infection. International Journal of Molecular Sciences 19, 1122. https://doi.org/10.3390/ijms19041122

Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R.D., Felsenfeld, G., 2010. Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. Genes & Development 24, 2543–2555. https://doi.org/10.1101/gad.1967810

Zhang, H., Xing, Z., Mani, S.K.K., Bancel, B., Durantel, D., Zoulim, F., Tran, E.J., Merle, P., Andrisani, O., 2016a. RNA helicase DEAD box protein 5 regulates Polycomb repressive complex 2/Hox transcript antisense intergenic RNA function in hepatitis B virus infection and hepatocarcinogenesis: HEPATOLOGY, Vol. XX, No. X, 2016 ZHANG ET AL. Hepatology 64, 1033–1048. https://doi.org/10.1002/hep.28698

Zhang, H., Xing, Z., Mani, S.K.K., Bancel, B., Durantel, D., Zoulim, F., Tran, E.J., Merle, P., Andrisani, O., 2016b. RNA helicase DEAD box protein 5 regulates Polycomb repressive complex 2/Hox transcript antisense intergenic RNA function in hepatitis B virus infection and hepatocarcinogenesis: HEPATOLOGY, Vol. XX, No. X, 2016 ZHANG ET AL. Hepatology 64, 1033–1048. https://doi.org/10.1002/hep.28698

MAIN FIGURES & LEGENDS

Figure 1: DDX5 and DDX17 are involved in correct HBV RNA processing in Primary Human Hepatocytes

PHH were infected with HBV at MOI 250 and transfected with siRNA targeting both DDX5 and DDX17 (siDDX5/17, gold bars) or with non-targeting siRNA (siCTL, grey bars) 4dpi. 4 days post-transfection (i.e. 8 days post-infection), cells were harvested and the supernatant was collected.

A.Levels of DDX5 and DDX17 mRNA were assessed by RT-qPCR and normalized to siCTL condition. GUSB was used as housekeeping gene. Extinction was confirmed by Western-blotting with anti-DDX5 or anti-DDX17 antibodies. Actin was used as loading control.

B. HBeAg and HBsAg secretion was measured by ELISA. Levels are expressed as ratio of siCTL.

C. Total DNA was extracted and digested or not by ExoI/III. Total HBV DNA and cccDNA levels were then quantified by specific qPCR. Levels are expressed as ratio of siCTL.

D. Total RNA was extracted with Trizol and digested with DNase I. HBV total RNA and 3.5 kb RNAs were quantified by RT-qPCR with primers targeting a region common to all HBV transcripts or primers specific for a sequence unique to 3.5 kb RNAs. Values are expressed as ratio of siCTL.

E. 10 ug of total RNA were migrated on agarose gel. 3.5, 2.4 and 2.1 kb HBV RNAs were detected after transfer and incubation with HBV-specific DIG-labeled probes (top panel). 28S and 18 S rRNA used as loading and quality controls were detected by UV imaging (bottom panel).

The bars represent the average of 3 or 6 independent experiments, indicated as n=x. The error bars represent the standard error or the mean. Mann-Whitney test was use to compare siCTL and siDDX5/17 conditions; α-threshold = 0.05, *p-value<0,05, **p-value<0,01

Figure 2: DDX5 and DDX17 are involved in correct HBV RNA processing in HepG2-NTCP

HepG2-NTCP were infected with HBV at MOI 250 and transfected with siRNA targeting both DDX5 and DDX17 (siDDX5/17, gold bars) or with siRNA negative control (siCTL, grey bars) 4dpi.

4 days post-transfection (i.e. 8 days post-infection), cells were harvested and the supernatant was collected.

A. Levels of DDX5 and DDX17 mRNA were assessed by RT-qPCR and normalized to siCTL condition. GUSB was used as housekeeping gene. Extinction was confirmed by Western-blotting with anti-DDX5 or anti-DDX17 antibodies. Actin was used as loading control.

B. HBeAg and HBsAg secretion was measured by ELISA. Levels are expressed as ratio of siCTL.

C.Total DNA was extracted and digested or not by ExoI/III. Total HBV DNA and cccDNA levels were then quantified by specific qPCR. Levels are expressed as ratio of siCTL.

D. Total RNA was extracted with Trizol and digested with DNase I. HBV total RNA and 3.5 kb RNAs were quantified by RT-qPCR with primers targeting a region common to all HBV transcripts or primers specific for a sequence unique to 3.5 kb RNAs. Values are expressed as ratio of siCTL.

E. 10 ug of total RNA were migrated on agarose gel. 3.5, 2.4 and 2.1 kb major HBV RNAs were detected after transfer and incubation with HBV-specific DIG-labeled probes (top panel). 28S and 18 S rRNA used as loading and quality controls were detected by UV imaging (bottom panel).

The bars represent the average of 3 or 6 independent experiments, indicated as n=x. The error bars represent the standard error mean. Mann-Whitney test was use to compare siCTL and siDDX5/17 conditions; α-threshold = 0.05, *p-value<0.05

Figure3: DDX5 and DDX17 are necessary for correct HBV RNA splicing

HepG2-NTCP were infected with HBV at MOI 250 and transfected with siRNA targeting both DDX5 and DDX17 (siDDX5/17) or with siRNA negative control (siCTL) 4dpi. 4 days post-transfection, cells were harvested and RNA was extracted.

A. Mean coverage for the duplicates siCTL (in blue) and siDDX5/17 (green). Reads are aligned to chimeric HBV genome genotype D starting at position 1820, where EcoRI site = +1 position. A redundant sequence (1820-1932) is added in 3’ and indicated with a ligh grey square. PAS : Polyadenylation Site (position 1916-1921).

B. Schematic representation of spliced variants, named SP "X". Annotation starts at position 1814, corresponding to the preCore ATG. Red lines represent exons found to be more skipped upon DDX5/17 depletion.

C. Spliced isoforms SP2, SP12, SP16, SP10 and SP15 are representated as colored. Green: 1814-2067; orange: 2350-2447; Red: 489-1932; Yellow: 2237-2447; Grey: 2350-2471, Blue: 282-1932 Positions of donor and acceptor sites are annotated with dashed lines. Skipped exons in siDDX5/17 is represented as orange (2350-2447), yellow (2237-2447) and grey (2350-2471) squares with dashed line. The number of reads corresponding to splicing junctions for both the exon included isoform (green) and the skipped exon isoform (red) are given for each replicate. The mean PSI value (percent spliced in) was calculated for each exon (n=2) in both conditions (in blue), and the deltaPSI value corresponds to PSIcontrol-PSIsiDDX5/17 condition is given with its associated p-value. SP3 and SP17 are the variants obtained after exon skipping.

Figure 4: DDX5 and DDX17 are required for HBV RNA 3'end maintenance and are decisive for the choice of the polyadenylation site

HepG2-NTCP were infected with HBV at MOI 250 and transfected with siRNA targeting both DDX5 and DDX17 (siDDX5/17) or with siRNA negative control (siCTL) 4dpi. 4 days post-transfection (i.e. 8 days post-infection), cells were harvested and RNA was extracted.

A. Schematic representation of 5' RACE and 3' RACE amplification strategy. PAS corresponds to canonical Polyadenylation signal, position 1916-1921. Gene specific primers used for 5'- and 3'-RACE are represented as « Gsp 5' » and « Gsp 3' » and their position is indicated. Numerotation is referred to EcoRI site = +1 position.

B. 5' RACE PCR products deried from siCTL and siDDX5/17 samples were migrated on 1% agarose gel. HBV transcripts are annotated according to their full-length size.

C. 3' RACE PCR products from siCTL and siDDX5/17 conditions were migrated on 1% agarose gel. 1 kb and 0,5 kb are indicated.

D. 3' RACE PCR products were sequenced with MinIONTM Nanopore Technology and aligned to HBV reference genome. The coverage corresponds to siCTL (grey) and siDDX5/17 (yellow) aligned reads visualized on IGV. Coverage scale is 90 000 reads. +1 position correspond to 3' Gsp primer. UGUA motifs are indicated with black arrow heads. 5'-TATAAA-3' indicated in bold

red corresponds to HBV canonical PAS. Positions 1, 350, 500 and 900 are indicated in blue. The 3' readthrough position and length are represented with blue annotations.

Figure 5: DDX5 and DDX17 recruit CPSF6 and NUDT21 on HBV RNAs

A. HepG2-NTCP cells were infected with HBV at a MOI of 250. 7 days post-infection, CLIP experiments were performed using the anti-DDX5 or anti-DDX17 antibody (yellow bars) or the anti-H3.3 antibody or no-antibody (NoAb) (grey bars) followed by RT-qPCR using primers targeting a region common to all HBV transcripts or primers specific for a sequence unique to 3.5 kb RNAs. H3.3 and NoAb conditions were used as negative controls. Bars represent the average of 2 independent experiments. The error bars represent the standard error of the mean.

B. HepG2-NTCP were infected with HBV at MOI 250 and transfected with siRNA targeting both DDX5 and DDX17 (siDDX5/17) or with non-targeting siRNA as negative control (siCTL) 4dpi. 4 days post-transfection (i.e. 8 days post-infection), cells were harvested and RNA was extracted. CLIP experiments were performed using anti-DDX5, anti-DDX17, anti-CPSF6 or anti-NUDT21 antibodies, followed by RT-qPCR using primers targeting a region common to all HBV transcripts (left) or primers specific for a sequence unique to 3.5 kb RNAs (right). Results are calculated as fold enrichment of DDX5/17 antibodies over H3.3 antibody and normalized to siCTL condition. Yellow and grey bars indicate siDDX5/17 and siCTL condition, respectively.

The bars represent the average of 3 independent experiments. The error bars represent the standard error of the mean.

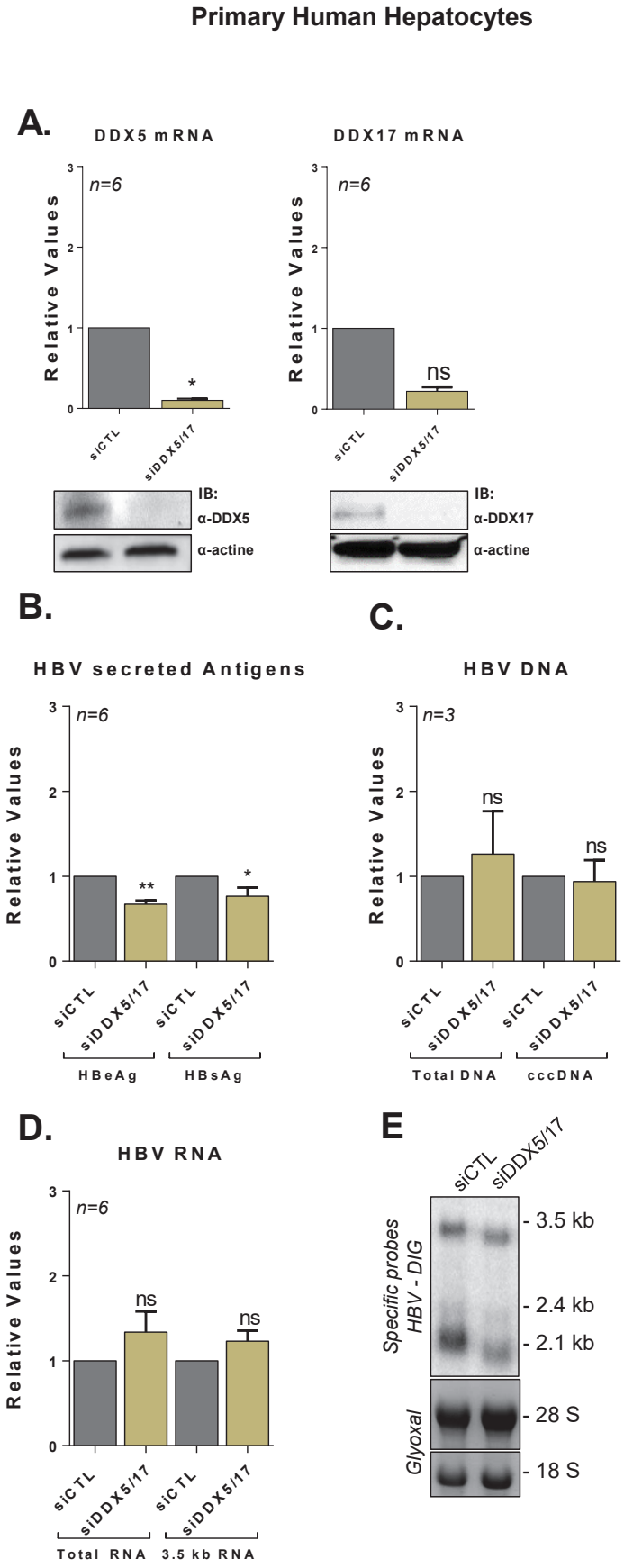

Primary Human Hepatocytes
A.
DDX5 mRNA
DDX17 mRNA
n=6
n=6
Relative Values
*
ns
siCTL
siDDX5/17
IB:
α-DDX5
α-DDX17
α-actine
B.
HBV secreted Antigens
n=6
**
*
HBeAg
HBsAg
C.
HBV DNA
n=3
ns
ns
Total DNA
cccDNA
D.
HBV RNA
n=6
ns
ns
Total RNA
3.5 kb RNA
E
siCTL
siDDX5/17
Specific probes HBV - DIG
Glyoxal
- 3.5 kb
- 2.4 kb
- 2.1 kb
- 28 S
- 18 S

Figure 1.

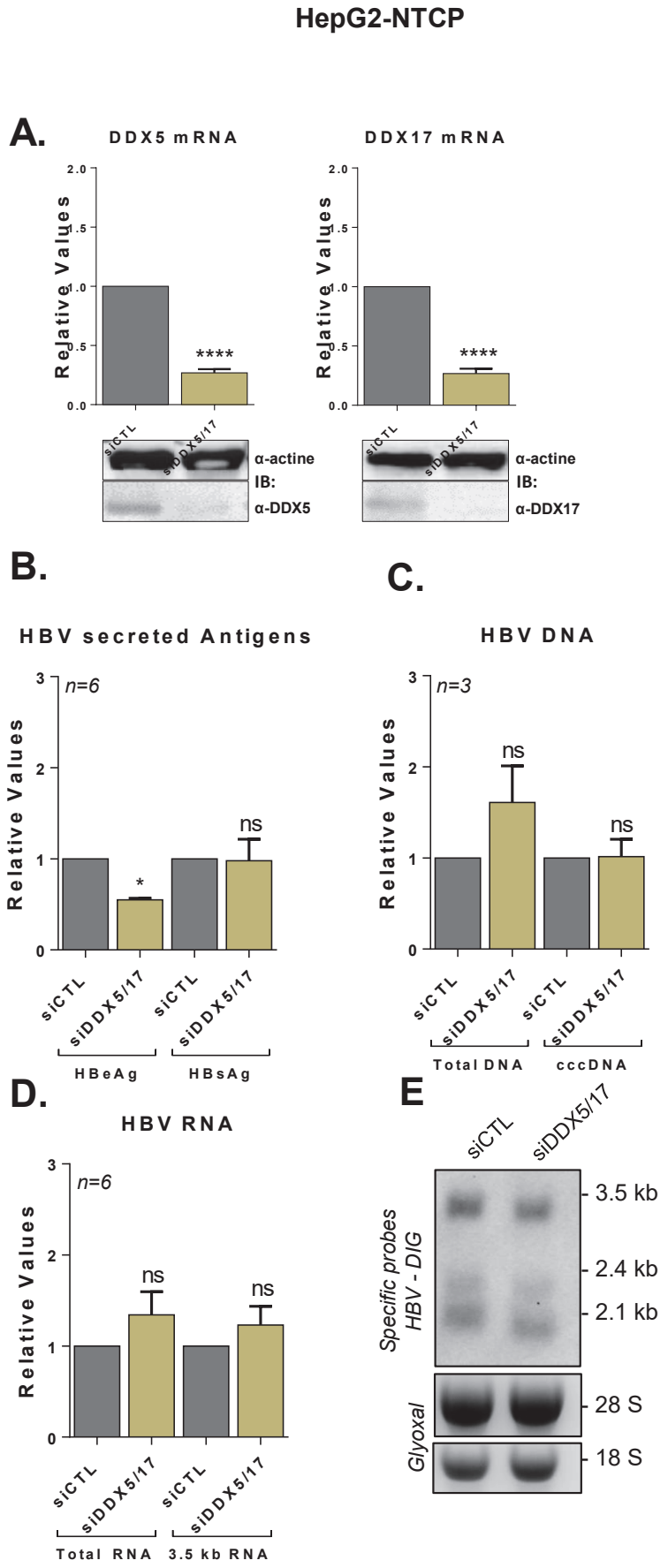

HepG2-NTCP
A.
DDX5 mRNA
DDX17 mRNA
Relative Values
****
siCTL
siDDX5/17
α-actine
IB:
α-DDX5
α-DDX17
B.
HBV secreted Antigens
n=6
*
ns
HBeAg
HBsAg
C.
HBV DNA
n=3
Total DNA
cccDNA
D.
HBV RNA
Total RNA
3.5 kb RNA
E.
Specific probes HBV - DIG
Glyoxal
3.5 kb
2.4 kb
2.1 kb
28 S
18 S

Figure 2.

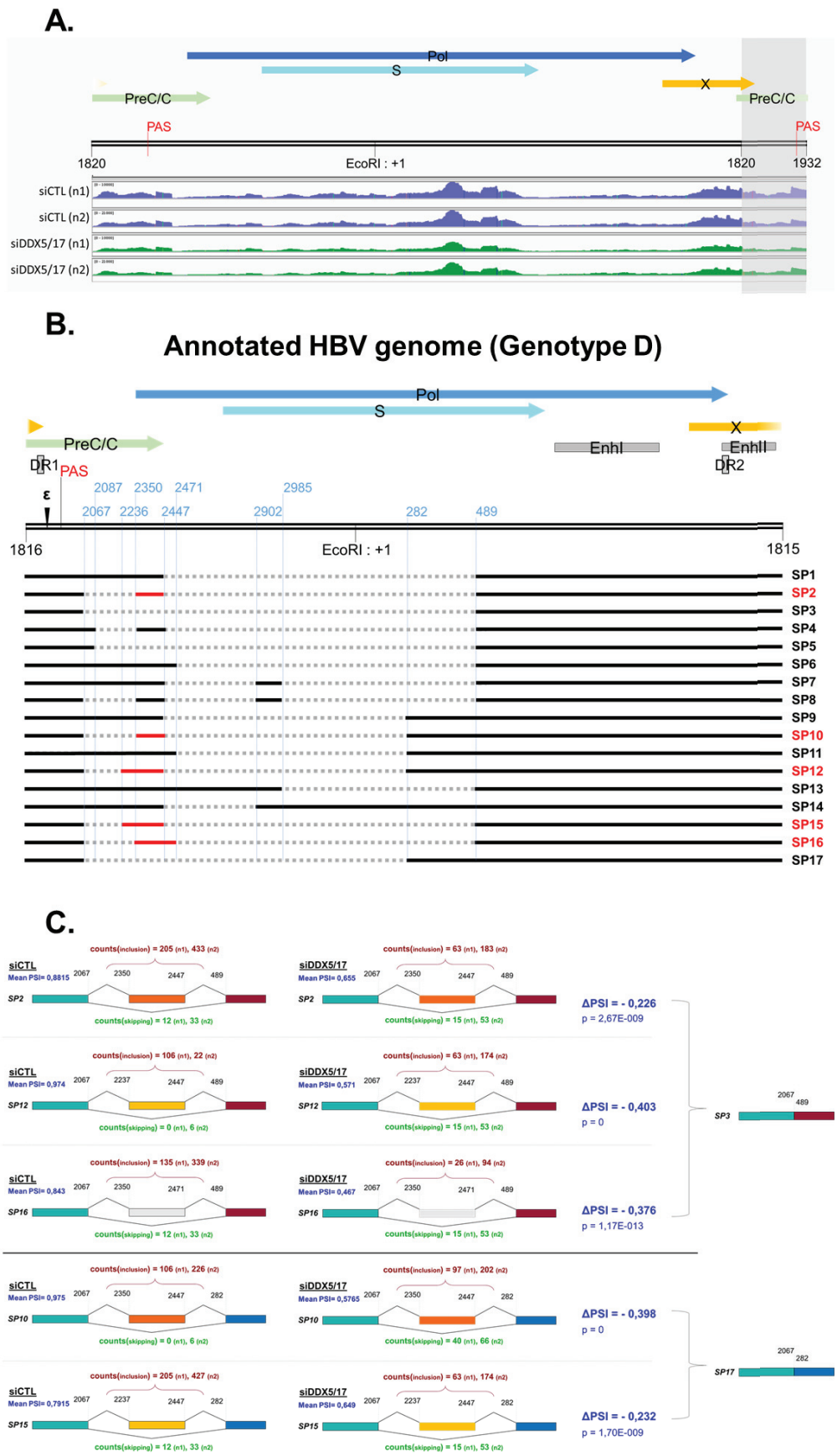

A.
Pol
S
X
PreC/C
PreC/C
PAS
PAS
1820
EcoRI : +1
1820
1932
siCTL (n1)
siCTL (n2)
siDDX5/17 (n1)
siDDX5/17 (n2)
B.
Annotated HBV genome (Genotype D)
Pol
S
X
PreC/C
EnhI
EnhII
DR1
PAS
DR2
ε
2087 2350 2471
2985
2067 2236 2447
2902
282
489
1816
EcoRI : +1
1815
SP1
SP2
SP3
SP4
SP5
SP6
SP7
SP8
SP9
SP10
SP11
SP12
SP13
SP14
SP15
SP16
SP17
C.
siCTL
Mean PSI= 0,8815
SP2
counts(inclusion) = 205 (n1), 433 (n2)
counts(skipping) = 12 (n1), 33 (n2)
siDDX5/17
Mean PSI= 0,655
SP2
counts(inclusion) = 63 (n1), 183 (n2)
counts(skipping) = 15 (n1), 53 (n2)
ΔPSI = - 0,226
p = 2,67E-009
siCTL
Mean PSI= 0,974
SP12
counts(inclusion) = 106 (n1), 22 (n2)
counts(skipping) = 0 (n1), 6 (n2)
siDDX5/17
Mean PSI= 0,571
SP12
counts(inclusion) = 63 (n1), 174 (n2)
counts(skipping) = 15 (n1), 53 (n2)
ΔPSI = - 0,403
p = 0
SP3
siCTL
Mean PSI= 0,843
SP16
counts(inclusion) = 135 (n1), 339 (n2)
counts(skipping) = 12 (n1), 33 (n2)
siDDX5/17
Mean PSI= 0,467
SP16
counts(inclusion) = 26 (n1), 94 (n2)
counts(skipping) = 15 (n1), 53 (n2)
ΔPSI = - 0,376
p = 1,17E-013
siCTL
Mean PSI= 0,975
SP10
counts(inclusion) = 106 (n1), 226 (n2)
counts(skipping) = 0 (n1), 6 (n2)
siDDX5/17
Mean PSI= 0,5765
SP10
counts(inclusion) = 97 (n1), 202 (n2)
counts(skipping) = 40 (n1), 66 (n2)
ΔPSI = - 0,398
p = 0
SP17
siCTL
Mean PSI= 0,7915
SP15
counts(inclusion) = 205 (n1), 427 (n2)
counts(skipping) = 12 (n1), 33 (n2)
siDDX5/17
Mean PSI= 0,649
SP15
counts(inclusion) = 63 (n1), 174 (n2)
counts(skipping) = 15 (n1), 53 (n2)
ΔPSI = - 0,232
p = 1,70E-009

Figure 3.

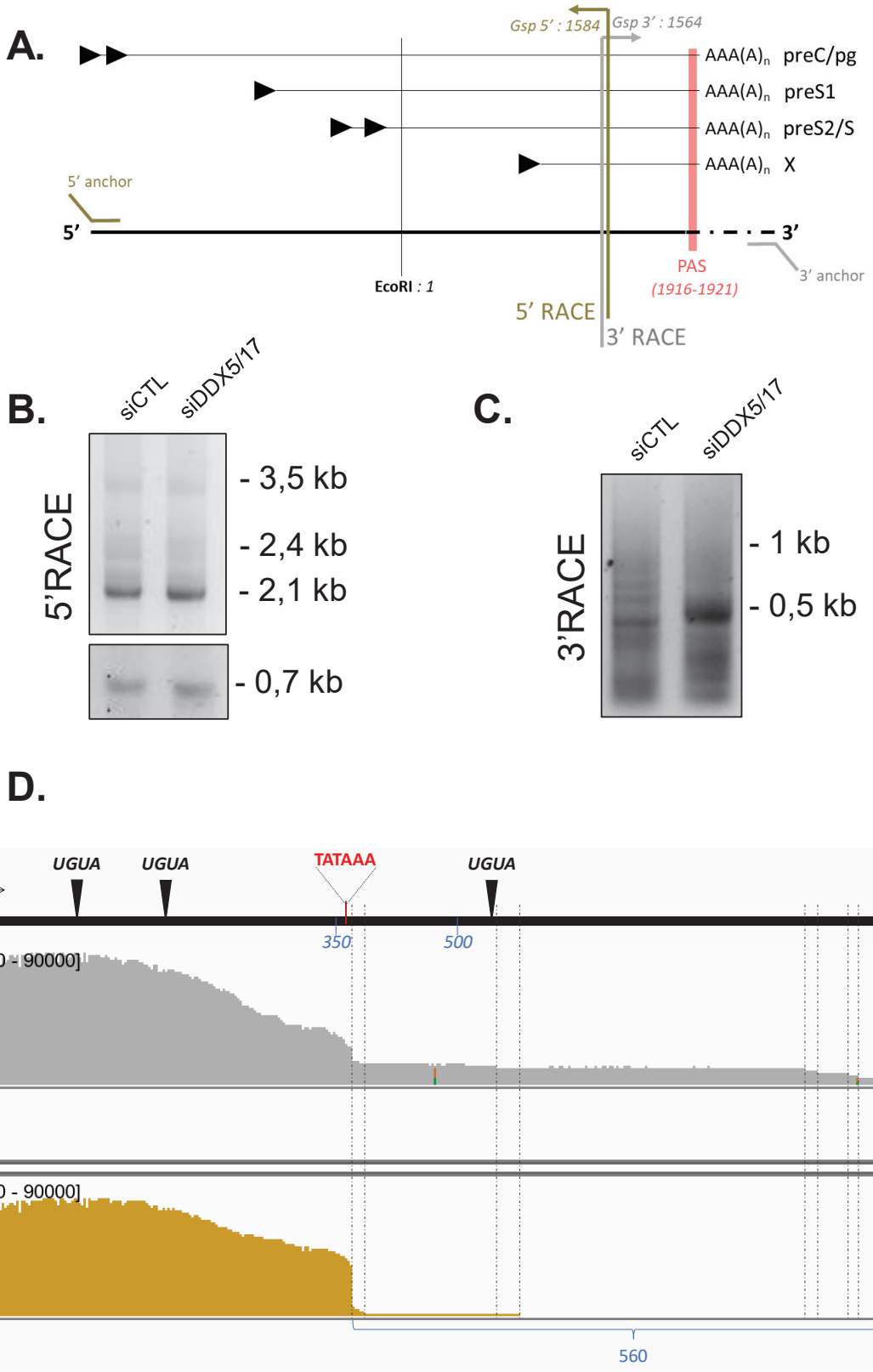

A.
Gsp 5' : 1584
Gsp 3' : 1564
AAA(A)n preC/pg
AAA(A)n preS1
AAA(A)n preS2/S
AAA(A)n X
5' anchor
5'
3'
EcoRI : 1
PAS
(1916-1921)
3' anchor
5' RACE
3' RACE
B.
siCTL
siDDX5/17
5'RACE
- 3,5 kb
- 2,4 kb
- 2,1 kb
- 0,7 kb
C.
siCTL
siDDX5/17
3'RACE
- 1 kb
- 0,5 kb
D.
UGUA
UGUA
TATAAA
UGUA
UGUA
1
350
500
930
[0 - 90000]
siCTL
[0 - 90000]
siDDX5/17
560

Figure 4.

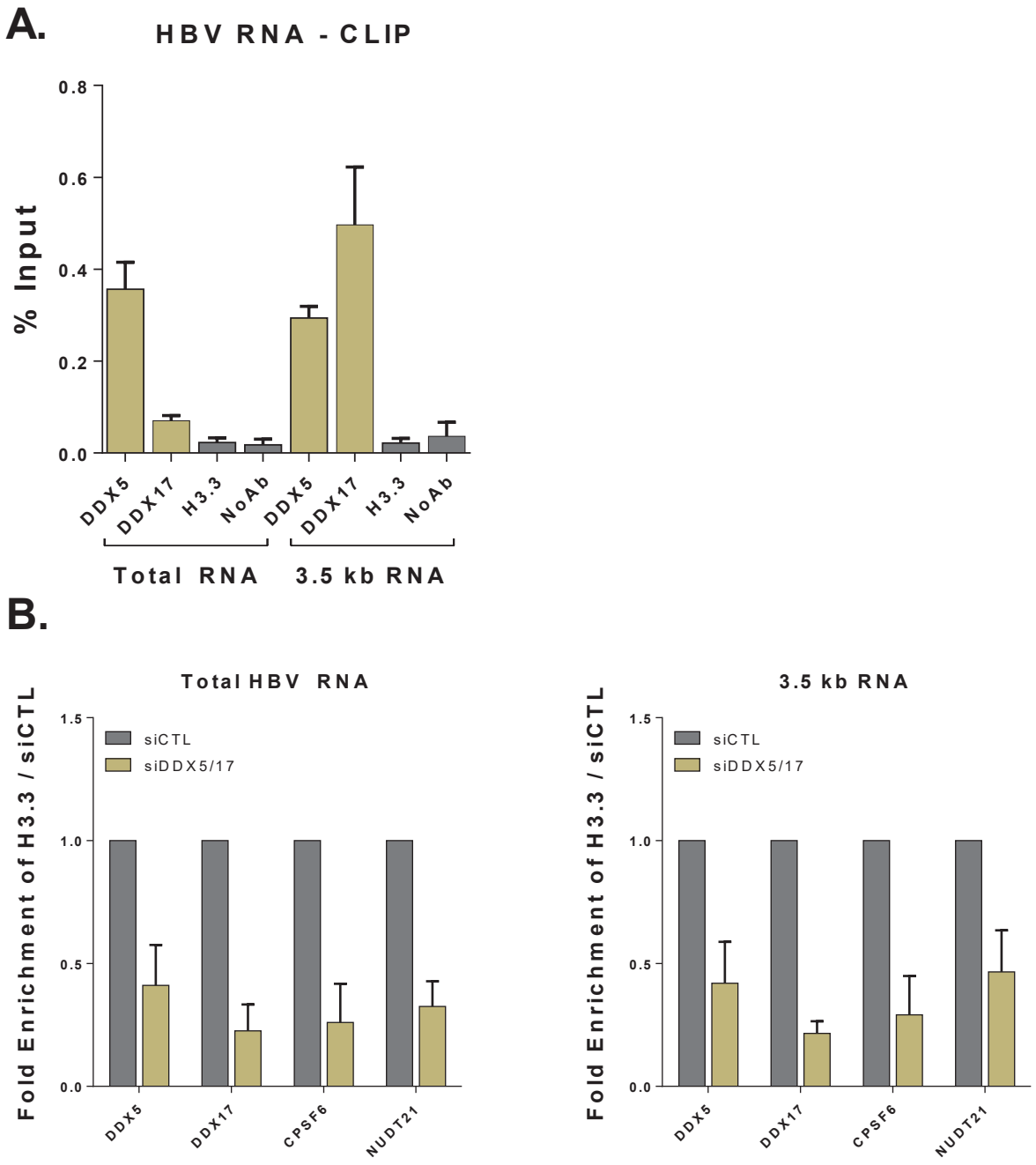

A.
HBV RNA - CLIP
% Input
0.8
0.6
0.4
0.2
0.0
DDX5
DDX17
H3.3
NoAb
DDX5
DDX17
H3.3
NoAb
Total RNA
3.5 kb RNA
B.
Total HBV RNA
Fold Enrichment of H3.3 / siCTL
siCTL
siDDX5/17
1.5
1.0
0.5
0.0
DDX5
DDX17
CPSF6
NUDT21
3.5 kb RNA
Fold Enrichment of H3.3 / siCTL
siCTL
siDDX5/17
1.5
1.0
0.5
0.0
DDX5
DDX17
CPSF6
NUDT21

Figure 5.

SUPPLEMENTARY FIGURES & LEGENDS

Supplementary figure 1: DDX5/17 depletion slightly increases cccDNA transcriptional activity

HepG2-NTCP were infected with HBV at MOI 250 and transfected with siRNA targeting both DDX5 and DDX17 (siDDX5/17) or with non-targeting siRNA as negative control (siCTL) 4dpi. 4 days post-transfection (i.e. 8 days post-infection), a 2 hour-pulse with ethynil uridine was performed and cells were harvested and RNA was extracted.

Neosynthesized RNAs were precipitated and quantified by RT-qPCR. siDDX5/17 (yellow bars) values are normalized to siCTL (grey bars) values. Bars represent the average of 3 independent experiments. The error bars represent the standard error of the mean. Mann-Whitney test was use to compare siCTL and siDDX5/17 conditions; α-threshold = 0.05, *p-value<0.05

Supplementary figure 2: DDX5 and DDX17 are required for correct HBV RNA splicing

A. Characterization of independent replicates #1 and #2 used for RNA-seq. Extinction was validated at mRNA levels for both DDX5 and DDX17. Northern-blot experiments with HBV-specific DIG-labelled probes allowed to confirm the size shifting. RNA quality was monitored with Bioanalyzer for each condition, the RIN number corresponds to the quality score.

B. Principal component analysis of RNAseq samples indicate the proximity between the replicates of each condition (pas sur qu'on puisse parler de "qualité" avec le PCA, et je ne sais pas non plus si "proximity" est le terme adéquat). siCTL duplicates are represented as grey circles, siDDX5/17 duplicates are represented as yellow circles.

C. Left panel: RFC5 exon 2 and SLC26A11 exon 4 were used as validation targets for RNA-seq analysis. RT-PCR was performed with flanking primers in both siCTL and siDDX5/17 conditions. Schematic representation of exon skipping and inclusion mRNA isoforms are represented on the right of the gel picture. Right panel: RT-qPCR validation of 2 upregulated mRNAs upon DDX5/17 depletion. Bars represent the average of 4 independent experiments. Values in the siDDX5/17 condition were normalized to siCTL values. The error bars represent the standard error mean.

Supplementary figure 3: DDX5 and DDX17 are required for HBV 3' end maintenance

A. 3' RACE PCR products were sequenced with MinIONTM Nanopore Technology and aligned to HBV reference genome. The coverage corresponds to HBV alone (black) and siCTL (grey)

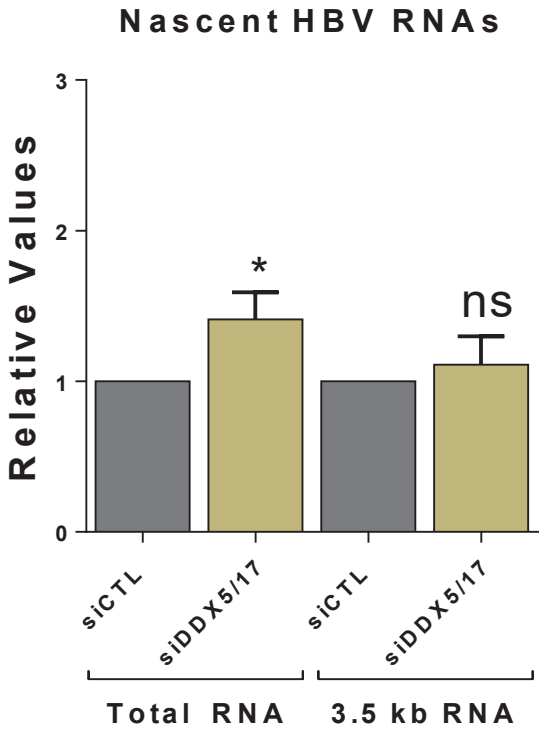
aligned reads visualized on IGV. Coverage scale is 90 000 reads. +1 position correspond to 3' Gsp primer. UGUA motifs are indicated with black arrow heads. 5'-TATAAA-3' indicated in bold red corresponds to HBV canonical PAS. Positions 1, 350, 500 and 900 are indicated in blue. The 3' readthrough position and length are represented with blue annotations.

B. Sequence alignment of PAS among the four major HBV genotypes (A-D). Values from HBVdb.

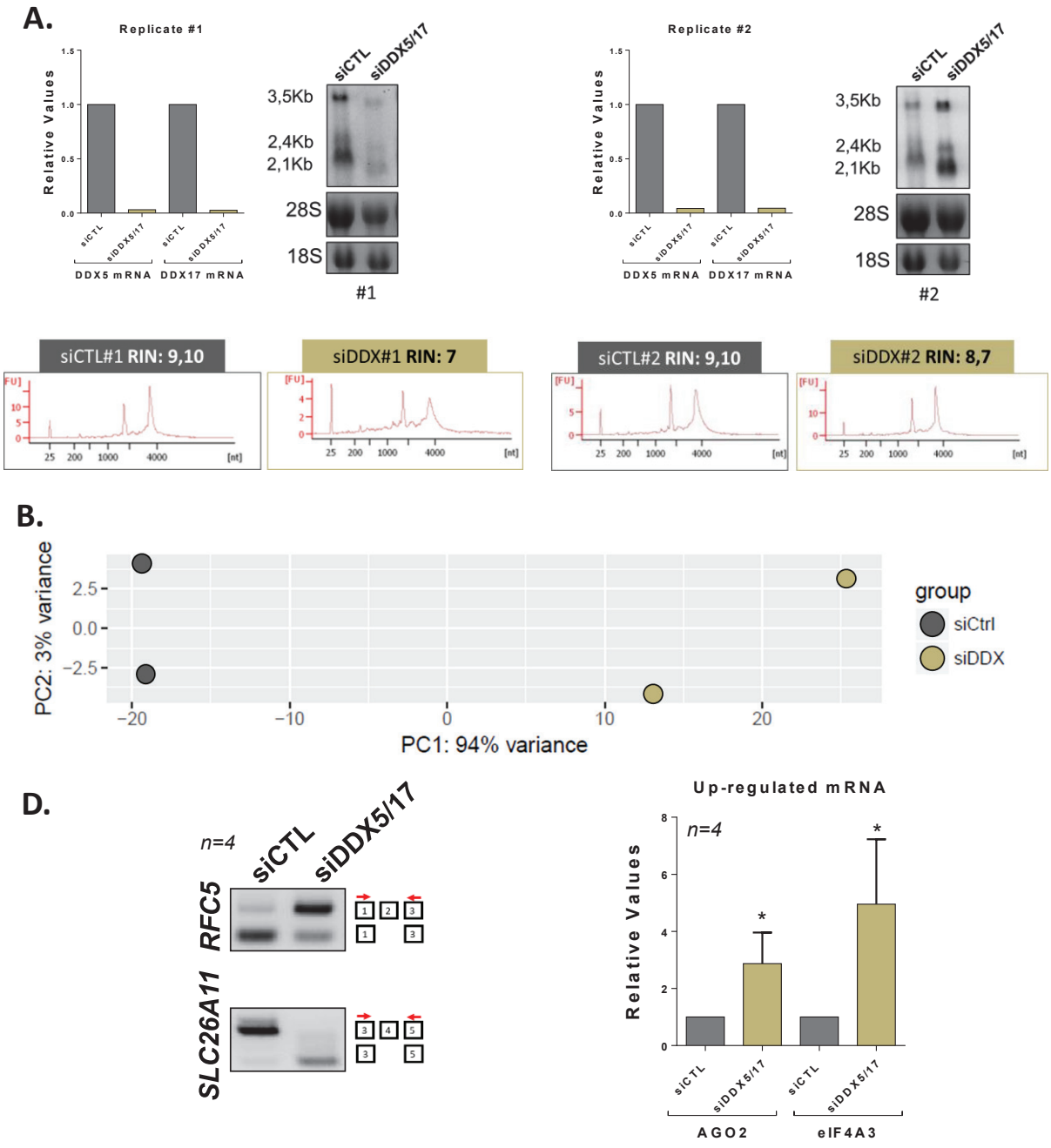
C. Table of the major putative alternative PAS sequences and frequencies in human and indicated presence in HBV 3' end readthrough. Prediction tool: PolyApred.

Supplementary figure 4: DDX5 and DDX17 are associated to cccDNA inPHH and HepG2-NTCP

HepG2-NTCP cells and PHH were infected with HBV at a MOI of 250. 7 days post-infection, ChIP experiments were performed using the anti-DDX5 or anti-DDX17 antibodies (yellow bars) or without antibody (grey bars) followed by qPCR amplificationof cccDNA. The bars represent the average of 3 independent experiments. The error bars represent the standard error the mean. Mann-Whitney test was use to compare NoAb and DDX5/17 antibodies; α-threshold = 0.05, *p-value<0.05

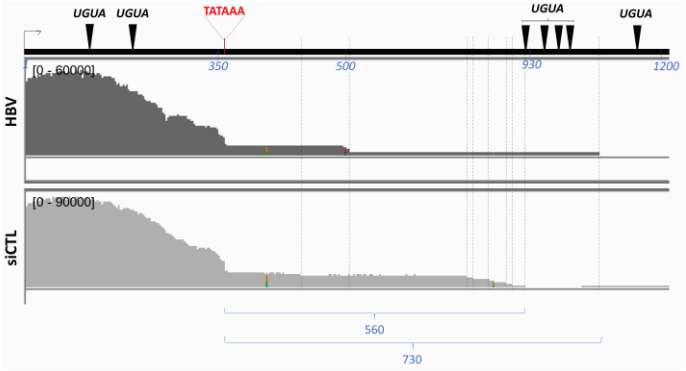


Supplementary Figure 1.



Supplementary Figure 2.

A.



B.

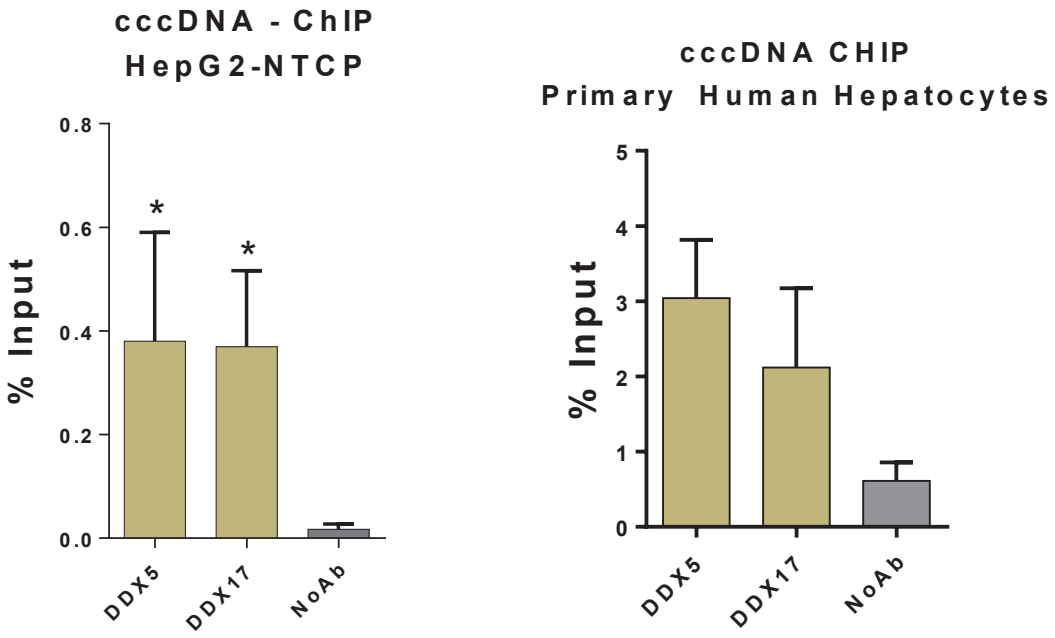
```
                                    * * * *
855 HBV CG : Genotype A : ----------TATAAA--------- >99.8% / >99.9%
                                     **
1749 HBV CG : Genotype B : ----------TATAAA--------- >99.5% / >99.8%
                                     **
2179 HBV CG : Genotype C : ----------TATAAA--------- >99.7% / >99.9%
                                     ***
954 HBV CG : Genotype D : ----------TATAAA--------- >99.7% / >99.5%
```

*** 100% conserved**

C.

PAS	Frequency in Human	Presence in HBV 3'end readthrough
AAUAAA	53%	No
AUUAAA	17%	No
UAUAAA AGUAAA AAGAAA AAUAUA AAUACA CAUAAA GAUAAA AAUGAA UUUAAA ACUAAA AAUAGA	23%	No

Supplementary Figure 3.



Supplementary Figure 4.

2.Supplementary results

In this study, we demonstrated that DDX5 and DDX17 are involved in HBV RNA transcription termination, notably by recruiting CPSF6 and NUDT21, two factors important for the choice of PAS. The depletion of DDX5 and DDX17 led to a strong decrease in CPSF6 and NUTD21 recruitment on viral transcripts, concomitantly with a preferential use of canonical PAS and, thus, to a shortening of HBV RNA 3'-UTR. Our results uncovered for the first time the presence of massive 3' read-through during HBV transcription in the context of HBV infection in HepG2-NTCP cells. So far, the biological significance of this read-through remains to be investigated, but we began to analyze the effect of DDX5/17 downregulation and, thus, the disappearance of the 3'UTR on HBV lifecycle.

DDX5 and DDX17 silencing amplifies HBV particles production

To test whether DDX5/17 silencing affects the subsequent HBV life cycle, we investigated what was happening at the cytoplasmic level of siCTL- and siDDX5/17-transfected cells. We quantified viral particles isolated from the cytoplasm of siCTL- and siDDX5/17-transfected HBV-infected HepG2-NTCP cells. Native agarose gel electrophoresis revealed a significant increase in HBc- and HBs-containing particles in the cytoplasm of HepG2-NTCP cells depleted for DDX5 and DDX17 (Supp Fig1A). By transmission electron microscopy, we were able to observe an increase volume and quantity of cytoplasmic vesicle induced by HBV infection upon DDX5/17 depletion (Supp Fig1B). However, HBsAg secretion decreases when DDX5 and DDX17 are repressed (Fig2A). This observation suggests a cytoplasmic accumulation of the viral particles in the cytoplasm of DDX5/17 depleted cells.

Western Blot analysis of intracellular HBs proteins would help in determining if an increase in protein translation is associated to less secretion.

HBV-induced cytoplasmic vesicles contain viral particles

Cytoplasmic vacuoles only exist in HBV-infected cells (Patient et al., 2007; Rat et al., 2019). However, to the best of our knowledge, no evidence for the presence of intra-vesicular viral particles was demonstrated in HBV infection model. To confirm the presence of viral particles in cytoplasmic vacuoles, we stained HBc and HBs with golden beads, in DDX5/17 depleted cells as these cells appear to contain higher levels of HBV-induced cytoplasmic compartments. The presence of golden beads in cellular vesicles confirmed the presence of HBc and HBs in such

compartments (Supp Fig1C – Upper and Middle panels). Moreover, the concomitant presence of these two viral proteins was confirmed by co-immunostaining with beads with different diameter (5 nm and 15 nm respectively) (Supp Fig1C – Lower pannel). Noteworthy, we can notice a weak number of golden beads within HBV-induced vacuoles. Ultra-thin section preparation physically cuts potential epitopes, explaining why HBc- and HBs-antibody staining does not reflect the expected abundance of particles. Collectively, these data uncovered an accumulation of viral particles in HBV-induced cytoplasmic vesicles which are not efficiently secreted. Moreover, co-immunogold staining demonstrated for the first time that both HBc and HBs are present in the same HBV-induced cytoplasmic vesicles.

DDX5 and DDX17 silencing does not affect pgRNA retro-transcription

To investigate whether the accumulation of cytoplasmic viral particle upon DDX5/17 depletion was due to a default in pgRNA retro-transcription, we carried out cytoplasmic capsid isolation and quantified encapsidated HBV RNA and DNA. A slight increase of total HBV RNA within capsid-containing particles was observed in siDDX5/17 (Supp Fig2A). Interestingly, an equivalent increase in encapsidated HBV DNA was also observed upon DDX5 and DDX17 depletion (Supp Fig2B). These results suggest no defects in viral retro-transcription despite the absence of DDX5 and DDX17.

Novel circular species of encapsidated HBV DNA are revealed by Nanopore sequencing

We demonstrated that DDX5/17 silencing does not affect pgRNA retro-transcription. We further wanted to investigate whether encapsidated HBV DNA within the cytoplasm was qualitatively affected by DDX5/17 silencing. To answer this question, we performed reverse PCR amplification of encapsidated DNA, isolated from cytoplasmic capsid-containing particles (Supp Fig3A). Reverse PCR allows the amplification of retro-transcribed and circularized species. PCR products were sequenced using Nanopore technology and approximately 14 000 reads were aligned to reference full-length HBV genome in both conditions (Supp Fig3B). Cytoplasmic encapsidated HBV DNA revealed a same coverage profile in siCTL and siDDX5/17 condition, uncovering that DDX5/17 silencing does not affect retro-transcribed DNA products into capsids.

As we demonstrated that HBs-Ag secretion was remaining stable upon DDX5/17 silencing despite an accumulation of cytoplasmic viral particles, we also qualitatively examined encapsidated DNA isolated from secreted capsid-containing particles. Complete genome PCR

amplification and Nanopore sequencing, as previously described, allowed the alignment of approximately 14 000 reads in both siCTL and siDDX5/17 conditions (Supp Fig3C). Coverage revealed no difference in siDDX5/17 depleted condition respect to siCTL condition. Altogether, these data suggest that encapsidated viral DNA integrity is not affected by DDX5/17 depletion.

SUPPLEMENTARY MAT & MET

Extraction of encapsidated HBV RNA and DNA

Cells were lysed in lysis buffer (5 mM PIPES, 85 mM KCl, 0,5% NP-40 , 1 mM PMSF) during 20 minutes on ice. After "tight" douncing 10 times, the nuclei were pelleted and the cytoplasm was isolated. DNase I (Roche) and RNase A (Macherey-Nagel) treatments were performed for 2 hours at 37°C on the cytoplasmic fraction. Viral particles were precipitated with PEG800 and resuspended in 1X PBS. Part of the extracted viral particles were loaded on a 1.2% agarose gel in native condition, transferred on PVDF membrane in 1X TNE buffer overnight and processed as classic Western-blotting. The remaining viral particles were treated with Proteinase K (Epicentre), 1% SDS, 1 mM EDTA, 100 mM NaCl in 10 mM Tris pH 7.5, at 37°C overnight with shaking. Encapsidated nucleic acids were extracted with MasterPure™ Complete DNA Purification Kit (Epicentre). (RT)-qPCR was performed as described above.

Transmission Electron Microscopy

Cells were fixed directly in the culture dish with 2% glutaraldehyde for 15 minutes at 4°C, and then incubated for at least 45 minutes at 4°C with 2% glutaraldehyde and 0.1 M cacodylate pH 7. Three washes with 0.1 M cacodylate pH 7.4 and 0.2 M saccharose were performed before incubation with 1% osmium teroxyde and 0.15 M cacodylate pH 7.4. Samples were then dehydrated in successive ethanol baths with increasing concentration (from 30% to 100%) thus allowing epoxy resin substitution and impregnation. Cells were then included and resin was polymerized during 24 hours at 60°C. Ultramicrotome ultrathin sections were generated and contrasted for TEM imaging using JEM-1400 (Jeol) microscope.

Nanopore

DNA products of Complete genome (CG) amplification performed with Platinum™ *Taq* DNA Polymerase High Fidelity (Invitrogen) (see primer sequences table X) were prepared for Nanopore Sequencing with SQK-PBK004 kit, according to the manufacturer's instructions. Analysis was performed with Epi2me (epi2me.nanoporetech.com), NanoPipe (http://bioinformatics.uni-muenster.de), and aligned to the reference HBV genome corresponding to full-length HBV genome (starting position : 1627).

SUPPLEMENTARY RESULTS: FIGURES & LEGENDS

Figure S1: DDX5 and DDX17 silencing amplifies HBV particles production

HepG2-NTCP were infected with HBV at MOI 250 and transfected with siRNA targeting both DDX5 and DDX17 (siDDX5/17) or with non-targeting siRNA as negative control (siCTL) 4dpi. 4 days post-transfection (i.e. 8 days post-infection), cells were harvested and the supernatant was collected

A. Left panel: Analysis of complete and empty cytoplasmic HBV particles by native agarose gel electrophoresis, detected by western-blotting using a HBc-specific antibody to detect nuclecapsids and HBsAg-specific antibody to detect envelope proteins.

Right panel: Densitometry of 3 independent experiments. Values correspond to siDDX5/17 quantification normalized on siCTL condition. The error bars represent the standard error of the mean. Mann-Whitney test was use to compare siCTL and siDDX5/17 conditions; α-threshold = 0.05, *p-value<0,05

B. Ultrastructure of non-infected HepG2-NTCP and HepG2-NTCP cells infected with HBV at MOI 250 and transfected with either siCTL or siDDX5/17 4 dpi was investigated by Transmission Electron Microscopy. Scale: 1 um. Nuc : nucleus, Cy : cytoplasm; MBV : vacuoles.

C. Ultra-thin sections of siDDX5/17 cells were incubated with anti-HBc (upper panel), anti-HBs (middle panel) and both (lower panel) antibodies and revealed with specific secondary antibodies labeled with golden beads of different diameter (5nm diameter for HBc; upper panel, 5 nm for HBs; middle panel, 5 nm for HBc and 15 nm for HBs lower panel).

Figure S2: DDX5 and DDX17 silencing does not affect pgRNA retro-transcription

Capsids were isolated from the cytoplasm of HBV-infected cells, transfected with siCTL or siDDX5/17.

RNA (**A.**) and DNA (**B.**) were extracted from cytoplasmic capsids. Total HBV RNA, 3.5 kb RNAs or total HBV DNA were quantified by Specific (RT)-qPCR. Values are represented as siDDX5/17 (yellow bars) normalized to siCTL values (grey bars).

The bars represent the average of 3 independent experiments. The error bars represent the standard error the mean. Mann-Whitney test was use to compare siCTL and siDDX5/17 conditions; α-threshold = 0.05, *p-value<0.05

Figure S3: Novel circular species of encapsidated HBV DNA are revealed by Nanopore sequencing

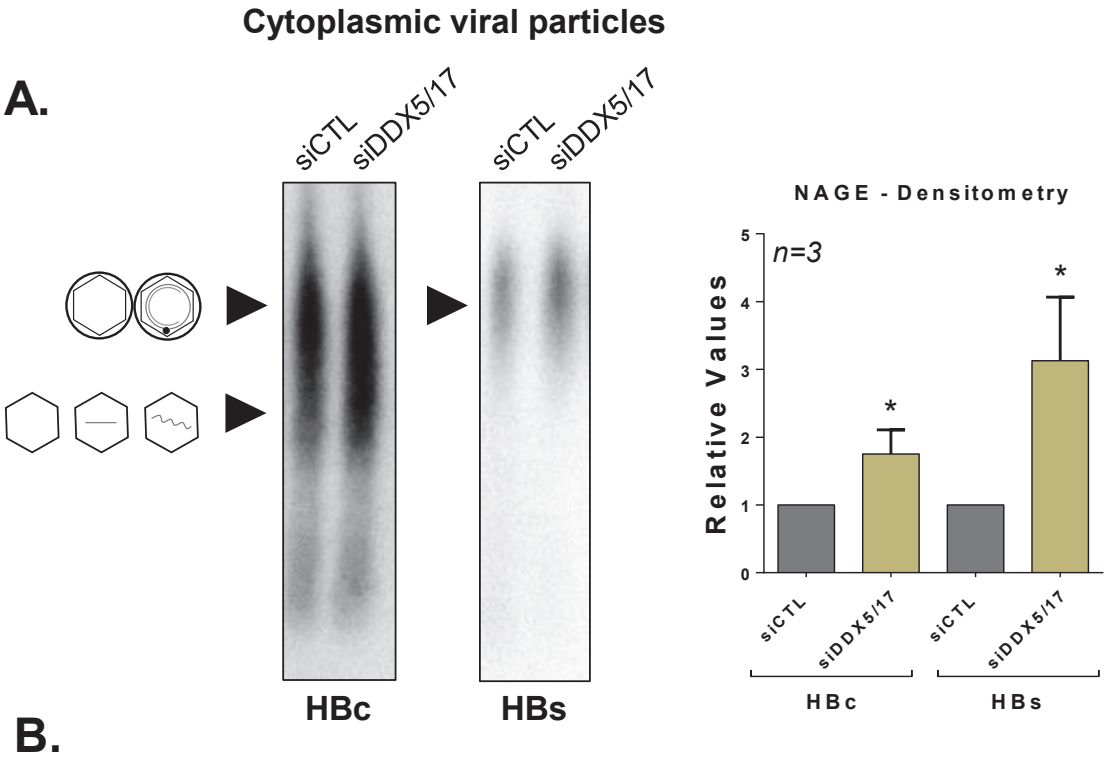
A. Scematic representation of HBV genome and ORFs. Numbering starts at EcoRI site = +1. Forward and reverse primers used for reverse PCR amplification are indicated in red, with their positions in brackets.

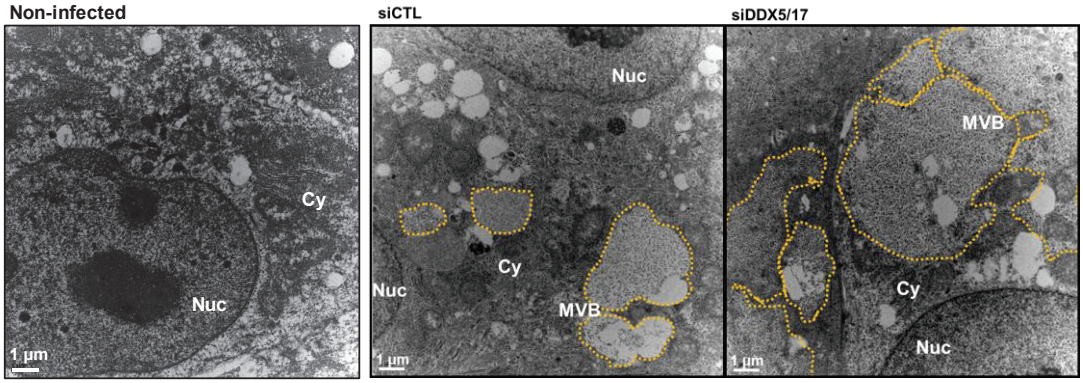
B. Encapsidated HBV DNA was extracted from cytoplasmic viral particles and analyze by MinION™ Nanopore sequencing. Products of reverse PCR were aligned to HBV genome, starting at the location of the first nucleotide of the forward primer. In grey is represented the siCTL condition, in yellow is represented the siDDX5/17 condition.

HBV transcripts are reported in the lower panel together with genome positions. The coverage is visualized using IGV. The coverage is representative of two independent experiments.

C. Encapsidated HBV DNA was extracted from cytoplasmic viral particles and analyze by MinION™ Nanopore sequencing. Products of reverse PCR were aligned to HBV genome, starting at the location of the first nucleotide of the forward primer. In grey is represented the siCTL condition, in yellow is represented the siDDX5/17 condition.

HBV transcripts are reported in the lower panel together with genome positions. The coverage is visualized using IGV. The coverage is representative of two independent experiments.

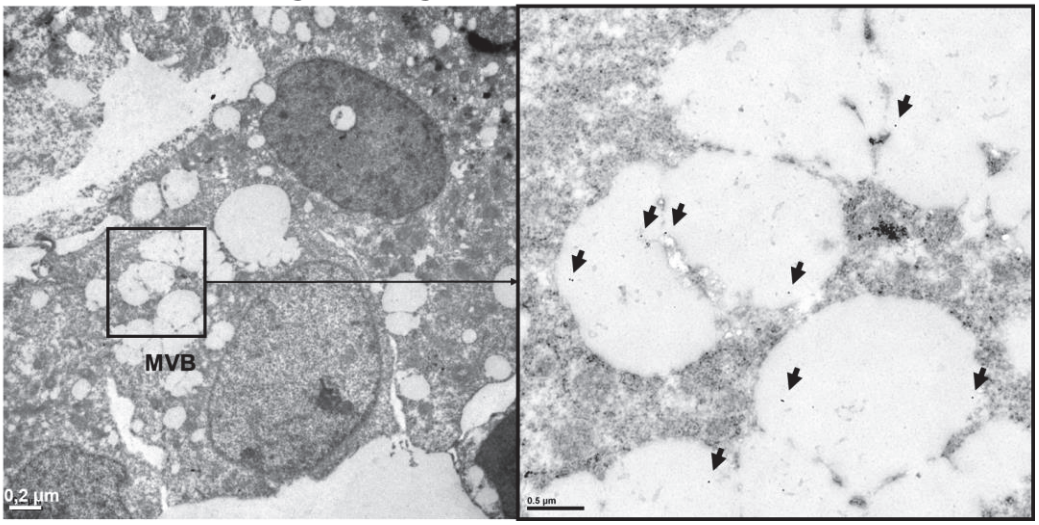




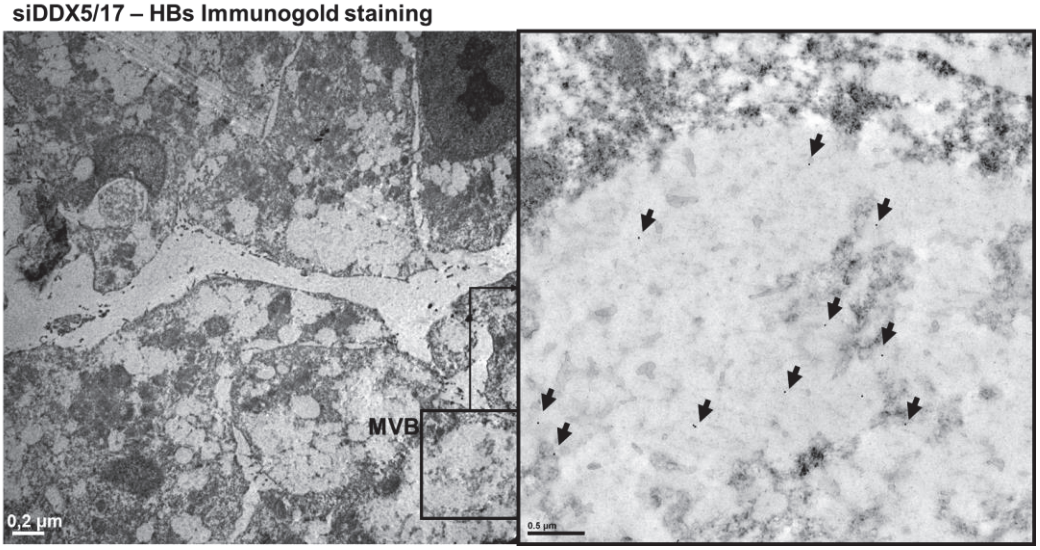
Supplementary Results : Figure S1. (1/2)

C.

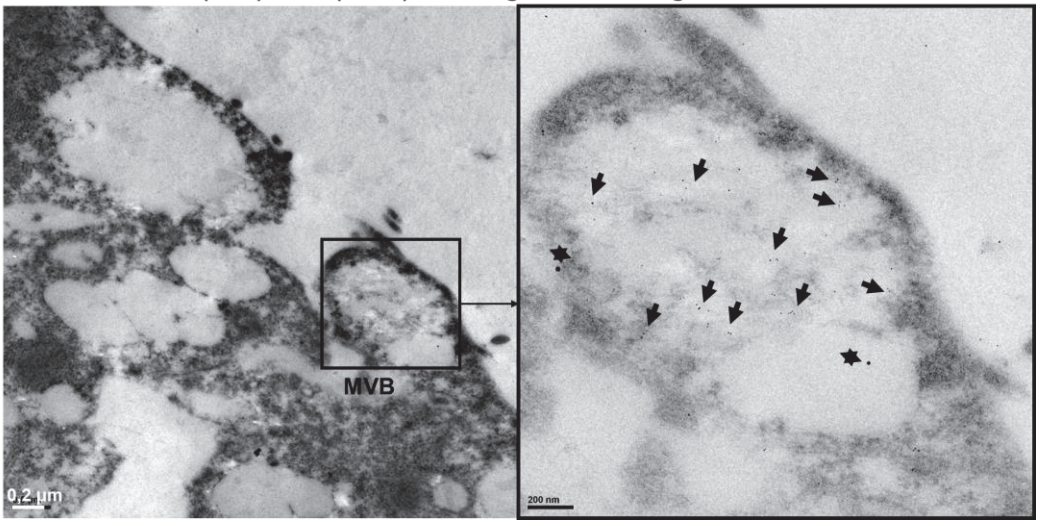
siDDX5/17 – HBc Immunogold staining



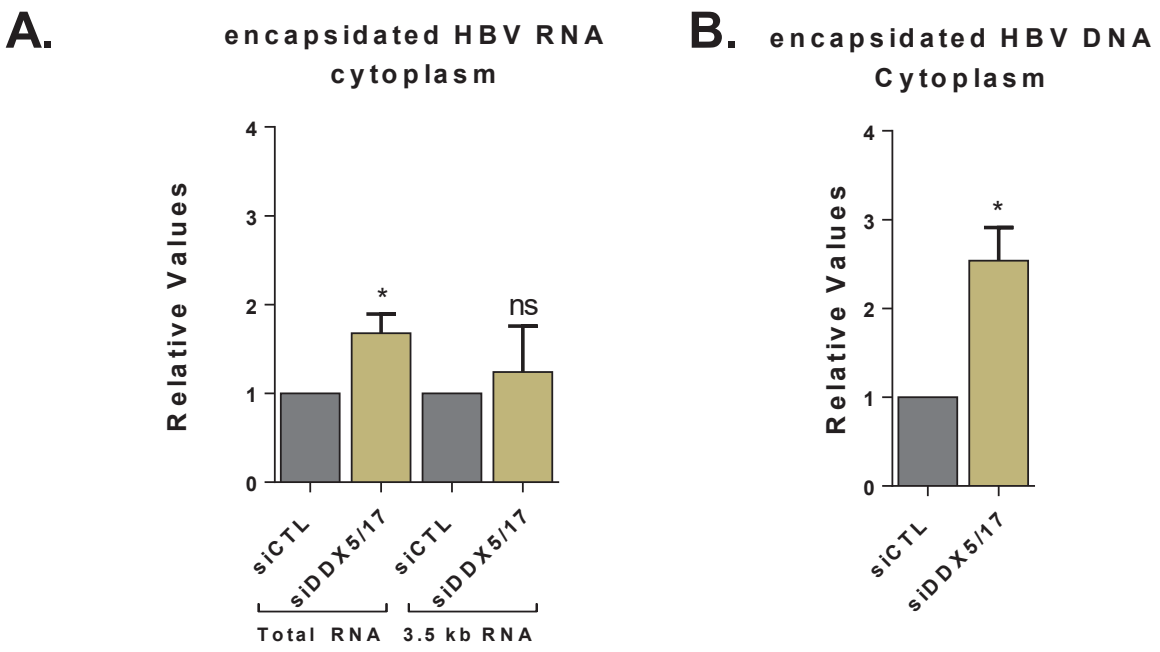
siDDX5/17 – HBs Immunogold staining



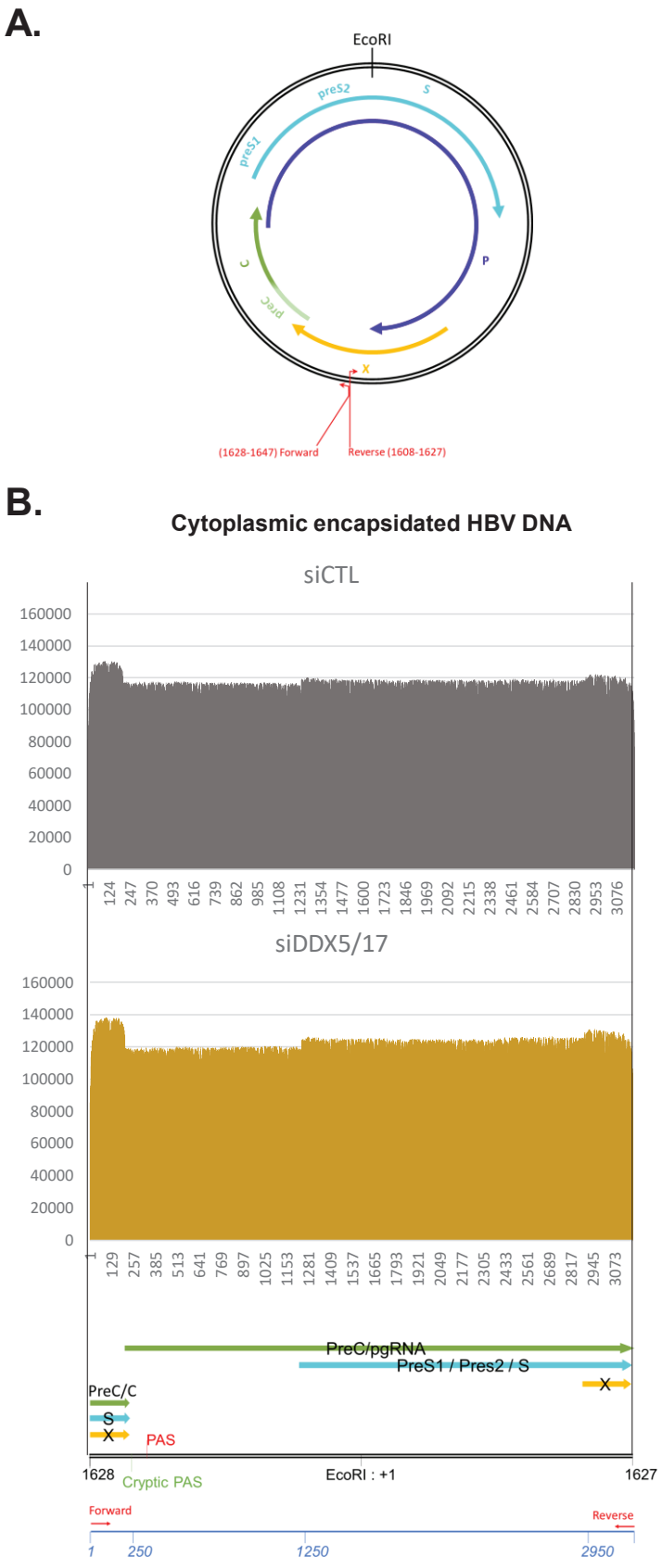
siDDX5/17 – HBc (5nm) / HBs (15nm) Immunogold Co-staining



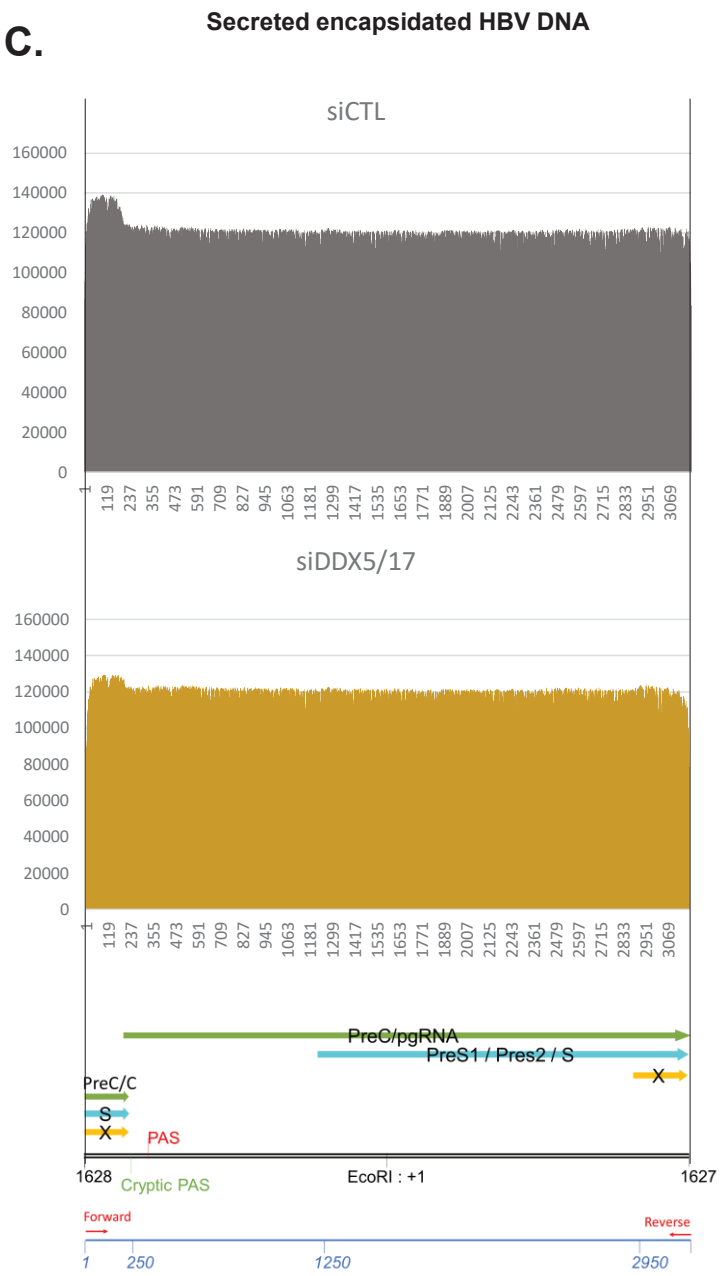
Supplementary Results : Figure S1. (2/2)



Supplementary Results : Figure S2.



Supplementary Results : Figure S3. (1/2)



Supplementary Results : Figure S3. (2/2)

Research Project 2

Article in preparation

Supplementary Results

III. Role of CTCF in cccDNA biology

1. Article in preparation

CTCF regulates HBV mRNA metabolism

Guillaume Giraud[*1, 2], Fleur Chapus[* 1,2], Judith Fresquet[1], Clémentine Sarda[1, 2], Christophe Combet[1, 2], Caroline Charre[1, 2, 3], Fabien Zoulim[1, 2, 3] and Barbara Testoni[1, 2]

[1]Cancer Research Center of Lyon, INSERM U1052, CNRS UMR-5286, 69008, Lyon, France, [2]University of Lyon, Université Claude Bernard Lyon A (UCBL1), 69008, Lyon, France, [3]Hospices Civiles des Lyon (HCL), 9002, Lyon, France

ABSTRACT

Despite the existence of efficient treatments, which maintain the infection under control, HBV is not completely eradicated due to the persistence of the viral covalently closed circular DNA (cccDNA) molecule in the infected hepatocytes. cccDNA is a chromatinized DNA structure and is the template for the transcription of the 6 HBV mRNAs that encode 7 HBV proteins necessary for HBV replication. The transcriptional activity of cccDNA is directly regulated by host factors, among which we previously identified the DEAD-box helicases DDX5 and DDX17, whose depletion triggers post-transcriptional defects of HBV RNAs (Chapus *et al.,* in preparation). CTCF is a transcription factor involved in the tri-dimensional conformation of the cellular genome and in RNA splicing and a protein partner of both DDX5 and DDX17.

We engineered synthetic cccDNA molecules harbouring mutations for the different CTCF binding sites using the minicircle strategy (mcHBV). After transfection of hepatoma cells with the mutated mcHBV, we assessed viral replicative parameters by Southern and

Northern blotting, qPCR, RT-qPCR and protein association to cccDNA by chromatin immunoprecipitation.

Our results showed that: 1) the mutation of the CTCF binding sites alone or in combination reduced the recruitment of CTCF on cccDNA and 2) while not affecting cccDNA formation, it affected both the levels and the quality of HBV RNAs, which accumulated, but migrated differently in Northern Blotting, suggesting a modified length respect to those produced in wild-type condition. Moreover, a decreased binding of DDX5 and DDX17 was observed on the CTCF-mutated cccDNA-like molecules.

Altogether, our results suggest that CTCF mediates the recruitment of DDX5/17 to cccDNA and that the protein complex CTCF/DDX5/DDX17 plays a pivotal role in the correct processing of HBV RNAs.

INTRODUCTION

Hepatitis B virus (HBV) belongs to the *Hepadnaviridae* family and its genome is a 3.2 kb partially double stranded relaxed circular DNA (rcDNA). HBV chronically infects more than 250 million people worldwide and is the leading cause of hepatocellular carcinoma. Despite the existence of treatments that efficiently maintain the infection under control, HBV cannot be fully eliminated because of the persistence of its replicative intermediate, the so-called covalently closed circular DNA (cccDNA) in the infected liver. New efforts are currently done to develop new therapeutic strategies directly targeting cccDNA by either degrading it or by controlling its transcriptional activity to reach a complete HBV cure(Martinez et al., 2019; Yang and Kao, 2014).

In the nucleus of infected hepatocytes, cccDNA forms a stable chromatinized episome that is the unique template of the transcription of the 6 major viral mRNA: the 0.7 kb X RNA, the 2.1 kb S and PreS2 RNA, the 2.4 kb PreS1 RNA, the 3.5 kb PreCore RNA and the 3.5 kb pregenomic RNA (pgRNA). Once synthesized, the pgRNA is retro-transcribed in rcDNA and encapsidated to form new infectious virions. In addition to these major mRNAs, spliced variants derived from the pgRNA and the PreS1 mRNA have been identified in the serum of patients and in cultured cell lines. Nevertheless, their role in HBV replication remains elusive. The transcription of these mRNAs is finely regulated by different regulatory elements, 4 promoters (X, PreS1, PreS2 and PreC), 2 enhancers (EnhI and EnhII) and a common polyadenylation signal (PAS), which require host cell transcription factors and cofactors for their activity(Dandri, 2020).

The CCCTC-binding factor CTCF is DNA binding protein involved in the regulation of gene expression and in particular transcription. It was originally identified as an insulator protein blocking the action of enhancers on a specific promoter, notably in the imprinted locus *IGF2/H19* (Braccioli and de Wit, 2019). CTCF mediates its effect by promoting chromatin looping between distant regions and organizes the genome as topologically associated domains in mammalian cells(Hyle et al., 2019; Khoury et al., 2020; Llères et al., 2019). Independently from its role in chromatin looping, CTCF can also directly repress or activate gene expression by interacting with the RNA Polymerase 2 (RNA Pol II) machinery(Chernukhin et al., 2007). Apart from transcription, CTCF also regulates co-transcriptional mRNA processing steps such as alternative splicing. For example, CTCF is recruited to the CD45 exon 5 genomic region and promotes its inclusion by inducing RNA Pol II pausing (Shukla et al., 2011). Genome-

wide studies have proposed that CTCF regulates alternative splicing by promoting chromatin loops between the alternative exons and their cognate promoter but these correlative studies have to be more properly demonstrated(Ruiz-Velasco et al., 2017). Interestingly, CTCF directly binds to RNA and requires this activity to regulate the expression of at least some of its target genes and to promote chromatin looping(Hansen et al., 2018; Saldana-Meyer et al., 2014; Saldaña-Meyer et al., 2019). In particular, CTCF forms a complex with the long non-coding RNA SRA (Steroide Receptor Activator)) and the protein DDX5, which are important for its insulator activity(Hongjie Yao et al., 2010). Increasing evidence indicates that CTCF is directly recruited to the genome of several DNA viruses and plays important roles in their replication by promoting chromatin looping driving transcriptional repression or alternative splicing(Pentland and Parish, 2015).

Based on these observations, we hypothesized that CTCF regulates HBV replication. In this study, we demonstrated that CTCF is directly recruited to cccDNA and HBV mRNAs and that CTCF regulates HBV mRNA processing.

MATERIALS AND METHODS

Generation of HBV molecule carrying mutation of CTCF BS

PCR from the pMN-HBV plasmid and using overlapping primers (**Table 1**) carrying the designed mutation in the CTCF BS were performed using the PrimeSTAR Max premix (Takara) according to the manufacturer's instruction. The parental plasmid was then digested with 10 U of DpnI (Promega) restriction enzyme for 2 h at 37 ˚C. TOP10 bacteria were transformed with the resulted plasmid for 45 s at 42 ˚C, plated in LB agar plates supplemented with 35 µg/mL kanamycin and incubated overnight at 37 ˚C. Plasmid DNA was extracted with the Nucleospin Plasmid Mini kit and SANGER-sequenced at GATC to check for the mutation. ZYCY10P3ST2 bacteria (System Bioscience) were then transformed with the plasmid DNA carrying the designed mutation, plated and incubated as above. 1 colony was picked up and cultured in Terrific broth medium overnight at 37 ˚C until reaching a DO between 0.4 and 0.6. 2 volumes of LB medium supplemented with 0.04 N NaOH and 20 µg/mL L-Arabinose were added to the culture, which was incubated for 8 h at 37 ˚C. Plasmid DNA corresponding to the HBV genotype D genome with an attr sequence added at the position 2850-2888 was extracted with the Nucleospin Endotoxin-free Xtra Maxi kit (Macherey-Nagel), digested with NdeI restriction enzyme (New England Biolabs) for 2 h at 37 ˚C and with Plasmid-Safe DNAse (System Bioscience) overnight at 37 ˚C according to the manusfacturer's instruction. Plasmid DNA was further purified with Phenol/Chloroform and precipitated with 2.5 volume of EtOH.

Mutated site		**Primer sequence**
CTCF BS1	F	GTGTTTGCTGACGCAACACCAACTGGATGGGGCTTGGTC
	R	GACCAAGCCCCATCCAGTTGGTGTTGCGTCAGCAAACAC
CTCF BS2	F	GCCGATCCATACTGCGGAATTATTAGCAGCTTGTTTTGCTCGCAGCAGG
	R	CCTGCTGCGAGCAAAACAAGCTGCTAATAATTCCGCAGTATGGATCGGC
CTCF BS3	F	CCTCCAGCTTATCGTCCGCCCAATGCCCCTATCCTATCAACACT
	R	AGTGTTGATAGGATAGGGGCATTGGGCGGACGATAAGCTGGAGG
CTCF BS5	F	CTGGGATTCTTTCGCGGAATTATTAGCAGCTTGTTTTGCTCGCAGCGCCTTCAGAGCA
	R	TGCTCTGAAGGCGCTGCGAGCAAAACAAGCTGCTAATAATTCCGCGAAAGAATCCCAG

Table 1: Sequence of the primers used to mutate CTCF BS in cccDNA

Culture and transfection of HepG2-NTCP cells

HepG2-NTCP cells were cultured in DMEM supplemented with 5 % FCS, 1 mM Sodium Pyruvate, 1 mM L-Glutamax, 100 U/mL Penicillin and 100 µg/mL Streptomycin at 37 ˚C in 5 % CO2.

HepG2-NTCP cells were transfected with 12 µg of plasmids with Mirus 2020 transfection reagent according to manufacturer's instruction. Once transfected, HepG2-NTCP cells were cultured as above in antibiotic-free medium. The day after, medium was replaced by fresh medium containing antibiotics. 3 days and a half after the transfection, cells were collected for further analysis.

Chromatin immunoprecipitation

Cells were washed twice with 1 X PBS and cross-linked with 1 % formaldehyde for 10 minutes at 37 ˚C. After 5 minutes of quenching with 125 mM Glycin at 37 ˚C, cells were washed twice with 1 X PBS, centrifuged 5 minutes at 1500 rpm and lysed with Nuclear Lysis Buffer to isolate nuclei for 30 minutes in ice. The lysate was then dounce 10 times and centrifuge 5 minutes at 3000 rpm at 4 ˚C. Nuclear membrane was then broken by 2 cycles of sonication 30 sec ON, 30 sec OFF on a Bioruptor (Diagenode). Debris were pelleted 10 minutes at 14800 rpm at 4 ˚C. The lysate was diluted 10 times with RIPA buffer supplemented with 1 X Complete Mini EDTA-free (Roche Diagnostics) and 1 mM PMSF and pre-cleared for 2 h at 4 ˚C by adding magnetic Protein G coated dynabeads (Life Technologies). Beads were discarded and the proper amount of antibody raised against the protein of interest was added to the chromatin or no antibody was added for the negative control. After an overnight incubation at 4 ˚C, magnetic Protein G coated dynabeads were added followed by another incubation of 2 h at 4 ˚C. 10 % of the No Antibody control was saved to constitute the Input. Beads were washed 5 times with RIPA buffer, once with TE buffer and resuspended in Elution buffer containing proteinase K. Chromatin was reverse crosslinked by incubating the mixture 2 h at 68 ˚C, purified by adding one volume of Phenol:Chloroform:Isoamyl Alcohol 25:24:1 (Life Technologies) and precipitated. The antibodies used in this study are the following: Rabbit anti-CTCF (Diagenode), Rabbit anti-DDX5 (Bethyl) and Rabbit anti-DDX17 (Proteintech).

ELISA

ELISA was performed as previously described (Stadelmayer et al., 2020b).

Real-time PCR

DNA was extracted as previously described (Stadelmayer et al., 2020b). RNA was extracted using TriReagent (MRC) according to manufacturer's instruction. Real time PCRs were performed as previously described(Stadelmayer et al., 2020b).

Northern blot

Purified RNA were denatured at 50°C for one hour with glyoxal reagent (Life Technologies) and then subjected to electrophoresis through 1X phosphate buffer in a 1.2% agarose gel and transferred to a positively charged nylon membrane (Amersham N+, GE). Membrane-bound RNA were hybridized overnight at 42°C to DIG-labeled HBV-specific probes (Table 2). The

membrane was washed twice in low-stringency wash buffer (1X SSC, 0.1% SDS) for 30 min at room temperature and twice in high-stringency wash buffer (0.1X SSC, 0.1% SDS) for 30 min at 65°C. Detection was performed using anti-DIG alkaline phosphatase (1/20 000) and CDP-Star reagent (Roche) according to the manufacturer's recommendation and imaged using ChemiDoc MP imaging system (BioRad). 18S and 28S rRNA were used as loading and quality control.

	DIG-Labeled Probes	
	Forward (5'-3')	**Reverse (5'-3')**
HBV-1	TAGCGCCTCATTTTGTGGGT	CTTCCTGTCTGGCGATTGGT
HBV-2	TAGGACCCCTTCTCGTGTTA	CCGTCCGAAGGTTTGGTACA
HBV-3	ATGTGGTATTGGGGGCCAAG	GGTTGCGTCAGCAAACACTT
HBV-4	TGGACCTTTTCGGCTCCTC	GGGAGTCCGCGTAAAGAGAG
HBV-5	GTCTGTGCCTTCTCATCTG	AGGAGACTCTAAGGCTTCC
HBV-6	TACTGCACTCAGGCAAGCAA	TGCGAATCCACACTCCGAAA
HBV-8	AGACGAAGGTCTCAATCGCC	ACCCACAAAATGAGGCGCTA

Table 2: DIG-Labeled HBV specific probes

CLIP

RIP experiments were carried out 7 days post-infection. Cells were washed in 1X PBS and UV cross-linked at 200 mJ/cm^2. Total cell lysate was obtained by 20 min RIPA lysis buffer incubation on ice, followed by sonication. Total cell lysate was pre-cleaned in RIPA buffer with Protein G-coupled magnetic beads (Dynabeads, Invitrogen), and then subjected to overnight immunoprecipitation at 4 °C using 1 µg of anti-CTCF antibody (Diagenode). Chromatin incubated with 5 µg anti-H3.3 antibody (Abcam Ab62642) or no antibody was processed as other conditions and was used as negative control. Immune complexes were then incubated 2 hours with Protein G-coupled magnetic beads at 4 °C, washed, and eluted in ChIP elution buffer. The flow-through from the no antibody condition was used as input. Immunoprecipitated RNA were extracted using ExtractAll TRI-Reagent (MRC), precipitated in isopropanol, washed in ethanol and resuspended in RNase-free water. RT-qPCR was performed as described above. Samples were normalized to input RNA using the ΔΔCt method were ΔCt = Ct (input) − Ct (immunoprecipitation) and calculated as percentage of the input.

RESULTS

CTCF is recruited to cccDNA and HBV mRNAs

To first address whether CTCF regulates HBV replication, we analysed the sequence of HBV genotype D genome with the CTCFBSDB 2.0 database to predict for potential CTCF binding sites (BS). This analysis revealed the presence of 8 potential CTCF BS dispersed in this sequence suggesting that CTCF is recruited to cccDNA (**Figure 1a**). To test this hypothesis, we performed chromatin immunoprecipitation (ChIP) experiments from HBV-infected HepG2-NTCP cells using or not the anti-CTCF antibody followed by qPCR to quantify cccDNA. As observed in Figure 1b, the percentage of Input obtained after using the anti-CTCF antibody is reproducibly higher than in the negative control indicating that CTCF is recruited to cccDNA.

To test whether CTCF can potentially be recruited to all the HBV genotypes, we aligned these 8 potential CTCF BS to the HBV genotype A to H to check for conservation. This analysis revealed that the stretch of Cs that are critical for the recruitment of CTCF is conserved in most of the genotype strongly suggesting that the recruitment of CTCF to cccDNA could be a common phenomenon to all the genotypes (**Figure 1c**).

Finally, CTCF can also directly bind to RNA and requires this activity for at least some of its function(Hansen et al., 2018; Saldana-Meyer et al., 2014; Saldaña-Meyer et al., 2019). To test whether CTCF can directly bind HBV RNAs, we performed CLIP experiments using the anti-CTCF or the anti-H3.3 (negative control) antibodies in the same condition as above. These experiments were followed by RT-qPCR to quantify the total HBV RNAs or more specifically the 3.5 kb RNAs (preCore and pgRNA). The data obtained show that the percentage of Input obtained using the anti-CTCF antibody is reproducibly higher compared to the negative control for both total and 3.5 kb RNAs (**Figure 1d**). These data show that CTCF is directly recruited to HBV RNAs.

CTCF recruitment to HBV EnhI regulates the metabolism of HBV mRNAs

To address the role of the recruitment of CTCF to cccDNA on HBV replication, we first used the minicircle strategy(Guo et al., 2016) to engineer HBV molecules (mcHBV) carrying silent mutations in the CTCF BS1 and/or CTCF BS2 that are both located to HBV EnhI region (**Figure 2a**).

To test whether the mutation of CTCF BS1 and CTCF BS2 alone or in combination impairs the recruitment of CTCF, we transfected HepG2-NTCP cells with the wild-type (wt) or the mutant synthetic mcHBV molecule. We then performed ChIP experiments using the anti-CTCF antibody or without antibody followed by qPCR to quantify the cccDNA. As expected, the percentage of Input obtained after using the anti-CTCF antibody is higher compared to the negative control in the cells transfected with the wt mcHBV indicating that CTCF is indeed recruited to this synthetic molecule (**Figure 2b**). Strikingly, this percentage of Input is reproducibly reduced in the cells transfected with the CTCF BS1 and/or CTCF BS2 mutant mcHBV compared to the wt mcHBV. Nevertheless, the percentage of Input obtained with the anti-CTCF is still higher compared to the negative control in the cells transfected with these mutant HBV molecule (**Figure 2b**). These data strongly suggest that the mutation of CTCF BS1 and CTCF BS2 alone or in combination decreases but does not fully abolish the recruitment of CTCF to cccDNA.

We then addressed the impact of the decreased CTCF binding to cccDNA on HBV markers. ELISA and qPCR analysis revealed that the mutation of the CTCF BS1 and/or BS2 does not significantly influence the level of secreted HBV antigens (**Figure 2c**) and the level of intracellular total HBV DNA (**Figure 2d**). The same observation is made for intracellular total HBV RNA (**Figure 2e**). However, RT-qPCR experiments showed a reproducible though not significant increase of the 3.5 kb RNAs after mutation of these CTCF BS compared to the wt condition (**Figure 2e**). These results suggest that the recruitment of CTCF to the CTCF BS1 and BS2 is involved in the regulation of the level of the 3.5 kb RNAs.

As the different HBV mRNAs cannot be discriminated by RT-qPCR, we performed Northern blot experiments using HBV specific probes that allow the detection of all HBV mRNAs except for the HBx RNA. Strikingly, we observe the appearance of a band migrating slightly faster than the 3 main bands corresponding to the 3.5 kb, 2.4 kb and 2.1 kb RNAs in the HepG2-NTCP cells transfected with the CTCF BS1 and/or BS2 mutant mcHBV while these 3 bands are unique in the control cells (**Figure 2f**). These results suggest that the mutation of CTCF BS1 and BS2 might impair the HBV mRNA processing.

CTCF recruitment to Core/POL and PreS1/POL ORFs also regulate the metabolism of HBV mRNAs

CTCF BSs are also present in other cccDNA regions than the EnhI region. That is particularly the case of the CTCF BS3 and BS5 located to Core/POL and PreS1/POL ORFs, respectively. To address the role of these particular CTCF BS, we generated mcHBV carrying mutations in these BS as described above (**Figure 3a**). In the case of CTCF BS3, these mutations are silent while for the CTCF BS5, the designed mutations introduced 1 aminoacid change in the POL ORFs: V199I.

We first addressed the impact of those mutations on the recruitment of CTCF to cccDNA. ChIP-qPCR experiments showed that the recruitment of CTCF to mcHBV carrying mutation in the CTCF BS3 and BS5 is reduced compared to the wt mcHBV. Nevertheless, the recruitment of CTCF to the mutant mcHBV is not completely abolished as the percentage of Input obtained using the anti-CTCF antibody is still higher compared to the negative control performed without antibody (**Figure 3b**).

As above, we checked for the impact of this decreased binding of CTCF on HBV markers. ELISA and qPCR analyses showed that the mutation of CTCF BS3 or BS5 does not influence the level of secreted HBV antigens (**Figure 3c**) and intracellular total HBV DNA (**Figure 3d**). Interestingly, RT-qPCR experiments revealed a reproducible though not significant increase of the level of intracellular total HBV mRNA and 3.5 kb RNAs (**Figure 3e**).

We finally performed Northern blot experiments with RNA extracted from HepG2-NTCP cells transfected with the wt or the CTCF BS3 or BS5 mutant mcHBV. Strikingly, while the 2.4 kb and the 2.1 kb migrate at the same size in all the condition, the 3.5 kb RNAs displayed a different migration pattern in the cells transfected with the mutant mcHBV. Indeed, as observed in Figure 3f, the 3.5 kb RNA migrates slower when the CTCF BS3 is mutated while after the mutation of CTCF BS5, the 3.5 kb RNAs migrates faster (**Figure 3f**). These results suggest then that the recruitment of CTCF to CTCF BS3 and BS5 is involved the processing of the 3.5 kb RNAs.

CTCF recruits DDX5 and DDX17 RNA helicases to cccDNA

All the data mentioned above point to a role of CTCF in the metabolism of HBV mRNAs. CTCF forms protein complex with several proteins involved in this process, in particular the RNA

helicase DDX5, which is important for the CTCF insulator function(Hongjie Yao et al., 2010). We recently demonstrated that DDX5 is recruited to cccDNA (Chapus, in prep). This observation led us to hypothesize that CTCF mediates its role in the metabolism of HBV transcripts at least by recruiting DDX5 to cccDNA. To tackle this question, we performed ChIP-qPCR using the anti-DDX5 antibody in HepG2-NTCP cells transfected with the different mutant mcHBV tested above. As expected, we observed the recruitment of DDX5 to the wt mcHBV (**Figure 4a**). Interestingly, the recruitment of DDX5 to the different mutant mcHBV is reproducibly decreased compared to the wt. These results suggest that CTCF binding is required for DDX5 recruitment to cccDNA.

The RNA helicase DDX17 is a close paralog of DDX5 and shares redundant functions with DDX5(Bourgeois et al., 2016; Giraud et al., 2018). We recently demonstrated that as DDX5, DDX17 is recruited to cccDNA (Chapus, in prep). Therefore, we tested whether CTCF recruits DDX17 to cccDNA by new ChIP-qPCR experiments using the anti-DDX17 antibody. The results shown in Figure 4b indicate that the recruitment of DDX17 to the mutant mcHBV is decreased compared to the wt mcHBV. These results suggest then that CTCF is also necessary for DDX17 recruitment to cccDNA.

DISCUSSION

Targeting cccDNA is critical to achieve a complete HBV cure. Therefore, understanding how HBV gene expression is regulated remains essential to reach this goal. The regulation of cccDNA relies on the combinatorial action of viral and host factors. In this study, we strongly suggest that CTCF binds to cccDNA through 8 potential binding sites and contributes to the metabolism of HBV RNAs.

Two of these potential binding sites are located to the EnhI region. The mutation of these CTCF BS alone or in combination triggers an increased level of the 3.5 kb RNA. At this stage, whether the mutation of these sites increases the transcription and/or the stability of the 3.5 kb RNAs remains elusive. Interestingly, CTCF is bound to the enhancer region of HPV18 and represses the expression of the oncoprotein E6/E7. In this particular case, CTCF creates a repressive chromatin landscape associated with the chromatin looping between the enhancer and the E6/E7 proteins(Pentland et al., 2018). In KHSV, CTCF is also bound to an intergenic region of the viral genome and contributes to the repression of lytic gene expression(Kang et al., 2011). These 2 studies point to a role of CTCF in repressing viral enhancer activity. Then, based on these observations, we hypothesize that the increased expression of the 3.5 kb RNA after mutation of the CTCF BS in the EnhI is the reflect of a transcriptional de-repression when CTCF is not recruited. Run-on experiments will definitely help in tackling this hypothesis. The 3.5 kb RNAs correspond to the PreCore and pgRNA, the former encoding the HBeAg. Surprisingly, the mutation of the 2 CTCF BS in the EnhI region is not associated with a significant increase of HBeAg level. This observation suggests that CTCF bound to the EnhI region could specifically repress the transcription of the pgRNA. 5' RACE experiments (Stadelmayer et al., 2020b) with primers discriminating between PreCore and pgRNA will allow to investigate this hypothesis.

An increased level of HBV mRNAs and, in particular, of the 3.5 kb RNAs is also observed after mutation of two other CTCF BS located in the PreCore/POL (CTCF BS3) and PreS1/POL (CTCF BS5) ORFs suggesting that CTCF bound to cccDNA at these positions also represses the transcription of HBV mRNA. As already mentioned, CTCF mediates its effect on transcription at least partly by promoting long-range interaction between regulatory sequences recruiting this factor(Braccioli and de Wit, 2019). In particular, several studies demonstrated that these CTCF-mediated chromatin loop frequently require 2 CTCF binding sites that are in convergent orientation(de Wit et al., 2015; Rao et al., 2014). Interestingly, the CTCF BS1 is in the opposite

direction of the other CTCF BS. We then think that the recruitment of CTCF throughout the cccDNA creates a chromatin topology that represses the expression of HBV mRNAs. Testing this hypothesis relies either on the use of chromosome conformation capture techniques or in super-resolutive microscopy. Nevertheless, the small size of cccDNA makes it challenging the use of such techniques to assess its topology in the presence or absence of CTCF.

Another striking phenotype observed after mutation of the CTCF BS is the change in size of either all HBV mRNAs or more specifically of 3.5 kb RNAs. Indeed, the mutation of the CTCF BS1, BS2 and BS5 triggers a shortening of the HBV mRNA. Spliced variants derived from either the pgRNA or the PreS1 mRNA have been described in the serum of patients and cultured cell lines (Duriez et al., 2017). Nevertheless, their role in HBV replication is still unclear. Apart from being involved in the regulation of transcription, CTCF is important for co-transcriptional processing such as alternative splicing(Shukla et al., 2011). Interestingly, the CTCF BS5 is located to the acceptor site of already described spliced variants (SP7, SP8 and SP14). This observation suggests then that CTCF represses the production of this particular spliced variant by being recruited to the CTCF BS5. Chromatin looping mediated or not by CTCF regulates as well alternative splicing(Tim R. Mercer et al., 2013; Ruiz-Velasco et al., 2017). The shortening of the HBV mRNAs could be therefore mediated by the production of spliced variants controlled by CTCF-mediated long-range interaction. In contrast, the mutation of CTCF BS3 triggers a lengthening of the 3.5 kb RNAs. 3' end mRNA processing is another co-transcriptional processing controlled by long-range interaction between promoters and terminator sites(Lamas-Maceiras et al., 2016). CTCF controls this process by preventing transcriptional readthrough notably after infection with the influenza virus(Heinz et al., 2018). The lengthening of the 3.5 kb RNAs observed after lost binding of CTCF at the BS3 could be therefore mediated by a transcriptional readthrough. The use of the 3rd generation sequencing using the Nanopore sequencing will definitely allow to test these different hypotheses.

CTCF can also directly bind RNA through some of its Zn finger domain(Hansen et al., 2018; Saldana-Meyer et al., 2014; Saldaña-Meyer et al., 2019). This activity is notably important for its contribution to organize the genome. Nevertheless, it is still not known whether the role of CTCF in co-transcriptional processing requires the recruitment of CTCF to RNA and whether CTCF recognizes the same motif as in DNA is not known. Interestingly, CTCF is directly recruited

to HBV RNAs. This observation let the possibility that the contribution of CTCF to HBV mRNA metabolism can be mediated by its binding to HBV mRNA. Determining whether the mutation of the CTCF BS impairs the recruitment of CTCF to RNA by new CLIP experiments is then critical for a better understanding of the mechanisms by which CTCF contributes to HBV.

One of these mechanisms could involve the RNA helicases DDX5 and DDX17. DDX5 together with the long-noncoding RNA SRA forms a complex with CTCF and mediates the CTCF insulator function(Hongjie Yao et al., 2010). Besides, DDX5 and DDX17 are multitask proteins involved at several stage of gene expression(Bourgeois et al., 2016; Giraud et al., 2018). In this study, we demonstrated that CTCF recruits both DDX5 and DDX17 to cccDNA and these data suggest that the protein complex CTCF/DDX5/DDX17 mediates the contribution of CTCF to HBV RNA metabolism. Rescue experiments using CRISPR based technology will help in determining whether forcing the recruitment of DDX5 or DDX7 to the mutated CTCF BS can compensate for the loss of CTCF binding.

All these data point to a role of CTCF in the metabolism of HBV mRNA. The effect of this contribution on new virion production has not been addressed in this study. Infecting cells with viral particles obtained from HepG2-NTCP cells transfected with the mutant cccDNA molecule will allow to determine whether the phenotype observed are in favour or in disfavour of HBV replicative capacity.

REFERENCES

1. Martinez, M. G., Testoni, B. & Zoulim, F. Biological basis for functional cure of chronic hepatitis B. *J. Viral Hepat.* **26**, 786–794 (2019).

2. Yang, H.-C. & Kao, J.-H. Persistence of hepatitis B virus covalently closed circular DNA in hepatocytes: molecular mechanisms and clinical significance. *Emerg Microbes Infect* **3**, e64 (2014).

3. Dandri, M. Epigenetic modulation in chronic hepatitis B virus infection. *Semin Immunopathol* (2020) doi:10.1007/s00281-020-00780-6.

4. Braccioli, L. & de Wit, E. CTCF: a Swiss-army knife for genome organization and transcription regulation. *Essays Biochem.* **63**, 157–165 (2019).

5. Hyle, J. *et al.* Acute depletion of CTCF directly affects MYC regulation through loss of enhancer-promoter looping. *Nucleic Acids Res.* **47**, 6699–6713 (2019).

6. Khoury, A. *et al.* Constitutively bound CTCF sites maintain 3D chromatin architecture and long-range epigenetically regulated domains. *Nat Commun* **11**, 54 (2020).

7. Llères, D. *et al.* CTCF modulates allele-specific sub-TAD organization and imprinted gene activity at the mouse Dlk1-Dio3 and Igf2-H19 domains. *Genome Biol.* **20**, 272 (2019).

8. Chernukhin, I. *et al.* CTCF Interacts with and Recruits the Largest Subunit of RNA Polymerase II to CTCF Target Sites Genome-Wide. *Mol Cell Biol* **27**, 1631–1648 (2007).

9. Shukla, S. *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79 (2011).

10. Ruiz-Velasco, M. *et al.* CTCF-Mediated Chromatin Loops between Promoter and Gene Body Regulate Alternative Splicing across Individuals. *Cell Syst* **5**, 628-637.e6 (2017).

11. Hansen, A. S. *et al.* An RNA-binding region regulates CTCF clustering and chromatin looping: *bioRxiv* (2018) doi:10.1101/495432.

12. Saldana-Meyer, R. *et al.* CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes & Development* **28**, 723–734 (2014).

13. Saldaña-Meyer, R. *et al.* RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Mol. Cell* **76**, 412-422.e5 (2019).

14. Yao, H. *et al.* Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* **24**, 2543–2555 (2010).

15. Pentland, I. & Parish, J. L. Targeting CTCF to Control Virus Gene Expression: A Common Theme amongst Diverse DNA Viruses. *Viruses* **7**, 3574–3585 (2015).

16. Stadelmayer, B. *et al.* Full-length 5'RACE identifies all major HBV transcripts in HBV-infected hepatocytes and patient serum. *J. Hepatol.* (2020) doi:10.1016/j.jhep.2020.01.028.

17. Guo, X. *et al.* The recombined cccDNA produced using minicircle technology mimicked HBV genome in structure and function closely. *Sci Rep* **6**, 1–10 (2016).

18. Bourgeois, C. F., Mortreux, F. & Auboeuf, D. The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nat. Rev. Mol. Cell Biol.* **17**, 426–438 (2016).

19. Giraud, G., Terrone, S. & Bourgeois, C. F. Functions of DEAD box RNA helicases DDX5 and DDX17 in chromatin organization and transcriptional regulation. *BMB Rep* **51**, 613–622 (2018).

20. Pentland, I. *et al.* Disruption of CTCF-YY1–dependent looping of the human papillomavirus genome activates differentiation-induced viral oncogene transcription. *PLOS Biology* **16**, e2005752 (2018).

21. Kang, H., Wiedmer, A., Yuan, Y., Robertson, E. & Lieberman, P. M. Coordination of KSHV latent and lytic gene control by CTCF-cohesin mediated chromosome conformation. *PLoS Pathog.* **7**, e1002140 (2011).

22. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

23. de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Mol. Cell* **60**, 676–684 (2015).

24. Duriez, M. *et al.* Alternative splicing of hepatitis B virus: A novel virus/host interaction altering liver immunity. *J. Hepatol.* **67**, 687–699 (2017).

25. Mercer, T. R. *et al.* DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. *Nat. Genet.* **45**, 852–859 (2013).

26. Lamas-Maceiras, M., Singh, B. N., Hampsey, M. & Freire-Picos, M. A. Promoter-Terminator Gene Loops Affect Alternative 3'-End Processing in Yeast. *J. Biol. Chem.* **291**, 8960–8968 (2016).

27. Heinz, S. *et al.* Transcription Elongation Can Affect Genome 3D Structure. *Cell* **174**, 1522-1536.e22 (2018).

28. Ziebarth, J. D., Bhattacharya, A. & Cui, Y. CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. *Nucleic Acids Res.* **41**, D188-194 (2013).

29. Hayer, J. *et al.* HBVdb: a knowledge database for Hepatitis B Virus. *Nucleic Acids Res.* **41**, D566-570 (2013).

30. Barrington, C., Finn, R. & Hadjur, S. Cohesin biology meets the loop extrusion model. *Chromosome Res* **25**, 51–60 (2017).

FIGURE LEGEND

Figure 1: CTCF is recruited to cccDNA and HBV RNAs

a Table showing the sequence of the 8 potential CTCF BS identified by CTCFBSDB 2.0(Ziebarth et al., 2013) (left panel) and their position in cccDNA (right panel).

b HepG2-NTCP cells were infected with HBV at a MOI of 250. 7 days post-infection, ChIP experiments were performed using the anti-CTCF antibody (blue bars) or without antibody (black bars) followed by cccDNA-specific qPCR. Bars represent the average of 4 independent experiments.

c Conservation of the 8 CTCF BS in the different HBV genotype. The sequence of the different genotypes have been extracted from HBVdb(Hayer et al., 2013).

d HepG2-NTCP cells were infected with HBV at a MOI of 250. 7 days post-infection, CLIP experiments were performed using the anti-CTCF antibody (blue bars) or the anti-H3.3 antibody (black bars) followed by RT-qPCR amplifying the total HBV RNA or the 3.5 kb RNAs.

Bars represent the average of 2 independent experiments. Error bars represent the standard error of the mean.

Figure 2: CTCF recruitment to EnhI region is involved in the metabolism of HBV mRNAs

a Table showing the wt and the mutated sequence of CTCF BS1 and BS2

b HepG2-NTCP cells were transfected with the wt (black bars), the CTCF BS1 mutant (blue bars), the CTCF BS2 mutant (red bars) or the double CTCF BS1/BS2 mutant (purple bars) mcHBV. 3 ½ days post-transfection, ChIP experiments were performed using the anti-CTCF antibody (dark bars) or without antibody (light bars) followed by qPCR experiments amplifying the cccDNA molecule.

c ELISA experiments quantifying secreted HBeAg (top panel) and HBsAg (bottom panel) were performed from cells treated as in **b**. The values are normalized by the amount of cccDNA of each transfection. Bars represent the average of 6 independent experiments.

d DNA was extracted from cells treated as in **b** and total HBV DNA was quantified by qPCR experiments. The values are normalized by the amount of cccDNA of each transfection.

e RNA was extracted from cells treated as in **b** and HBV total RNA (top panel) or 3.5 kb RNAs (bottom panel) were quantified by RT-qPCR experiments. The values are normalized by the amount of cccDNA of each transfection.

f Northern blot experiments were performed from RNA extracted as in **e** using HBV specific probes (top panel). Bottom panel shows the 28S/18S rRNA. The figure is representative of 2 independent experiments. Bars represent the average of 5 independent experiments. Error bars represent the standard error of the mean.

Figure 3: CTCF recruitment to CTCF BS3 and BS5 is involved in the metabolism of HBV mRNA

a Table showing the wt and the mutated sequence of CTCF BS3 and BS5

b HepG2-NTCP cells were transfected with the wt (black bars), the CTCF BS3 mutant (green bars) or the CTCF BS5 mutant (orange bars) mcHBV. 3 ½ days post-transfection, ChIP experiments were performed using the anti-CTCF antibody (dark bars) or without antibody (light bars) followed by qPCR experiments amplifying the cccDNA molecule.

c ELISA experiments quantifying secreted HBeAg (top panel) and HBsAg (bottom panel) were performed from cells treated as in **b**. The values are normalized by the amount of cccDNA of each transfection.

d DNA was extracted from cells treated as in **b** and HBV total DNA was quantified by qPCR experiments. The values are normalized by the amount of cccDNA of each transfection.

e RNA was extracted from cells treated as in **b** and HBV total RNA (top panel) or 3.5 kb RNAs (bottom panel) were quantified by RT-qPCR experiments. The values are normalized by the amount of cccDNA of each transfection. The bars represent the average of 3 independent experiments.

f Northern blot experiments were performed from RNA extracted as in **e** using HBV specific probes (top panel). Bottom panel shows the 28S/18S rRNA. The figure is representative of 2 independent experiments. Bars represent the average of 5 independent experiments. Error bars represent the standard error of the mean.

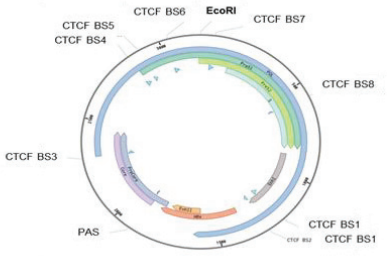
Figure 4: CTCF recruits the RNA helicases DDX5 and DDX17 to cccDNA

a HepG2-NTCP cells were transfected with the wt (black bars), the CTCF BS1 mutant (blue bars), the CTCF BS2 mutant (red bars), the double CTCF BS1/BS2 mutant (purple bars), the CTCF BS3 mutant (green bars) or the CTCF BS5 mutant (orange bars) mcHBV. 3 ½ days post-transfection, ChIP experiments were performed using the anti-DDX5 antibody (dark bars) or without antibody (light bars) followed by qPCR experiments amplifying the cccDNA molecule.

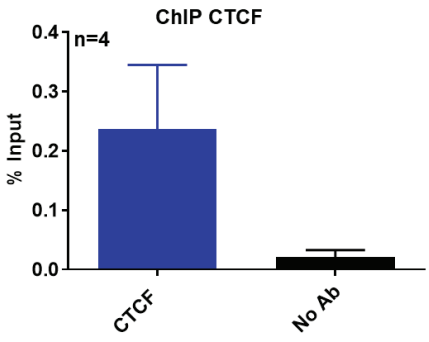
b HepG2-NTCP cells were transfected with the wt (black bars), the CTCF BS1 mutant (blue bars), the CTCF BS2 mutant (red bars), the double CTCF BS1/BS2 mutant (purple bars), the CTCF BS3 mutant (green bars) or the CTCF BS5 mutant (orange bars) mcHBV. 3 ½ days post-transfection, ChIP experiments were performed using the anti-DDX17 antibody (dark bars) or without antibody (light bars) followed by qPCR experiments amplifying the cccDNA molecule. Bars represent the average of 5 independent experiments. Error bars represent the standard error of the mean.

a

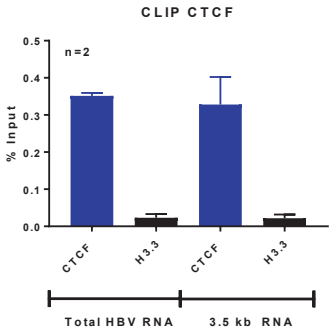
	Motif sequence	Start	End	Score
BS1	CTGACGCAACCCCCACTGGCTGGGGCT	1187	1213	
BS2	GCGGAACTCCTAGCCGC	1273	1289	
BS3	AGACCACCAAATGCCCCTAT	2296	2315	6.38442
BS4	TCTACAGCATGGGGCAGAA	2839	2858	3.59627
BS5	CCGACCACCAGTTGGATCCA	2890	2909	9.96261
BS6	CACGGAGGCCTTTTGGGGTGGAGCCCT	3025	3051	
BS7	TGTATTTCCCTGCTGGTGGCTCCAGTT	50	76	3.29011
BS8	GGTACAGCA	567	575	11.4532



b



d



c

```
CTCF BS1     1 CTGACGCAACCCCCACTGGCTGGGGCT
Genotype A 1187 CTGACGCAACCCCCACTGGCTGGGGCT
Genotype B 1187 CTGACGCAACCCCCACTGGTTGGGGCT
Genotype C 1187 CTGACGCAACCCCCACTGGCTGGGGCT
Genotype D 1187 CTGACGCAACCCCCACTGGCTGGGGCT
Genotype E 1187 CTGATGCAACCCCCACTGGCTGGGGCT
Genotype F 1187 CTGACGCAACCCCCACTGGTTGGGGCT
Genotype G 1187 CTGACGCAACCCCCACTGGTTGGGGCT
Genotype H 1187 CTGACGCAACCCCCACTGGCTGGGGCT

CTCF BS2     1 GCGGAACTCCTAGCCGC
Genotype A 1273 GCGGAACTCCTAGCAGC
Genotype B 1273 GCGGAACTCCTAGCCGC
Genotype C 1273 GCGGAACTCCTAGCCGC
Genotype D 1273 GCGGAACTCCTAGCCGC
Genotype E 1273 GCGGAACTCCTAGCCGC
Genotype F 1273 GCGGAACTCCTAGCCGC
Genotype G 1273 GCGGAACTCCTTGCCAGC
Genotype H 1273 GCGGAACTCCTAGCAGC

CTCF BS3     1 AGACCACCAAATGCCCCTAT
Genotype A 2297 AGACCACCAAATGCCCCTAT
Genotype B 2297 AGACCAGCAAATGCCCCTAT
Genotype C 2297 AGACCACCAAATGCCCCTAT
Genotype D 2297 AGACCACCTAATGCCCCTAT
Genotype E 2297 AGACCACCAAATGCCCCTAT
Genotype F 2297 AGACCACCAAAATGCCCCTAT
Genotype G 2297 AGGCCACCAAATGCCCCTAT
Genotype H 2297 AGACCACCAAATGCCCCTAT

CTCF BS4     1 TCTAC----------AGCATGGGG----------CAGAA
Genotype A 2869 GCTAC----------AGCATGGGAGGTTGGTCAT----------CAAAA
Genotype B 2863 TCTAC----------AGCATGGGA----------GGTTG
Genotype C 2863 GCTAC----------AGCATGGGA----------GGTTG
Genotype D 2863 GCTAC----------AGCATGGGG----------CAGAA
Genotype E 1981 TCCGC----------AGTATGGAT----------CGGCA
Genotype F 2863 GCTAC----------AGCATGGGA------GCACCTCTCTCAACGACAAGA
Genotype G 2863 TCAACGGCGAGAAGGGGCATGGGA----------CAGAA
Genotype H 2863 TCAACGGCGAGAAGGGGCATGGGA----------CAGAA

CTCF BS5     1 CCGACCACCAGTTGGATCCA
Genotype A 2923 CAGACCACCAGTTGGATCCA
Genotype B 2923 CCGATCATCAGTTGGACCCT
Genotype C 2930 CCGATCATCAGTTGGACCCT
Genotype D 2905 CCGACCACCAGTTGGATCCA
Genotype E 2923 CCGACCACCAGTTGGATCCA
Genotype F 2923 CCGATCATCAGTTGGACCCT
Genotype G 2923 CAGACCATCAGCTGGATCCG
Genotype H 2923 CAGACCACCAGTTGGATCCA

CTCF BS6     1 CACGGAGGCCTTTTGGGGTGGAGCCCT
Genotype A 3058 CACGGTGGGCTTCTGGGGTGGAGCCCT
Genotype B 3058 CATGGAGGACTGTTGGGGTGGAGCCCT
Genotype C 3064 CACGGCGGTATTTTGGGGTGGAGCCCT
Genotype D 3025 CACGGAGGCCTTTTGGGGTGGAGCCCT
Genotype E 3055 CACGGAGGCCTTTTGGGGTGGAGCCCT
Genotype F 3058 CATGGGGGCCTGTTGGGGTGGAGCCCT
Genotype G 3058 CACGGTGGCCTGCTGGGGTGGAGCCCT
Genotype H 3058 CACGGTGGGCTTCTGGGGTGGAGCCCT

CTCF BS7     1 TGTATTTCCCTGCTGGTGGCTCCAGTT
Genotype A   47 TGTATTTTCCTGCTGGTGGCTCCAGTT
Genotype B   47 TGTACTTCCTGCTGGTGGCTCCAGTT
Genotype C   47 TGTATCTTCCTGCTGGTGGCTCCAGTT
Genotype D   47 TGTATTTCCCTGCTGGTGGCTCCAGTT
Genotype E   47 TGTATTTTCCTGCTGGTGGCTCCAGTT
Genotype F   47 TGTAGTTTCCTGCTGGTGGCTCCAGTT
Genotype G   47 TGTATTTTCCTGCTGGTGGCTCCAGTT
Genotype H   47 TGTATTTTCCTGCTGGTGGCTCCAGTT

CTCF BS8     1 GGTACAGCA
Genotype A 768 TGTGCAGCA
Genotype B 768 TGTACAACA
Genotype C 768 TGTACAGCA
Genotype D 768 GGTACAGCA
Genotype E 768 TGTACAACA
Genotype F 768 TGTACAACA
Genotype G 768 TGTGCAGCA
Genotype H 768 TGTGCAGCA
```

Figure 1

a

CTCF BS	WT	Mutant	Region
1	CTGACGCAACCCCCAC TGGCTGGGGCT	CTGACGCAACACCAACTGG ATGGGGCT	EnhI
2	GCGGAACTACTAGCCG CTTGTTTTGCTCGCGGC	GCGGAATTATTAGCAGCTTG TTTTGCTCGCAGC	EnhI

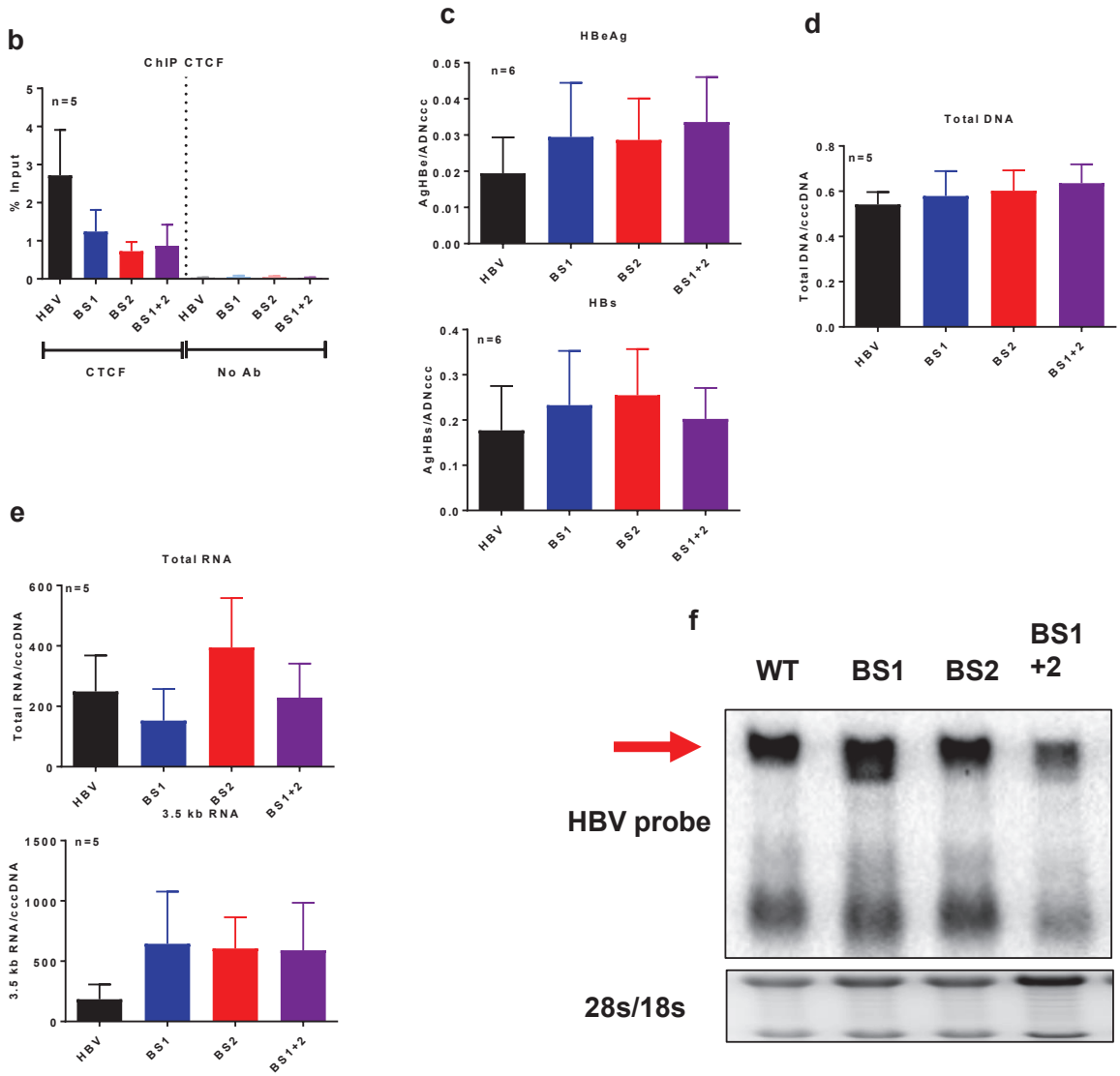


Figure 2

a

CTCF BS	WT	Mutant	Region
3	AGACCACCAAATGCCCCTAT	CGTCCGCCCAATGCCCCTAT	Core/POL ORF
5	GCGGAACTACTAGCCGCTTG TTTTGCTCGCGGC	GCGGAATTATTAGCAGCTTG TTTTGCTCGCAGC	PreS1/POL ORF

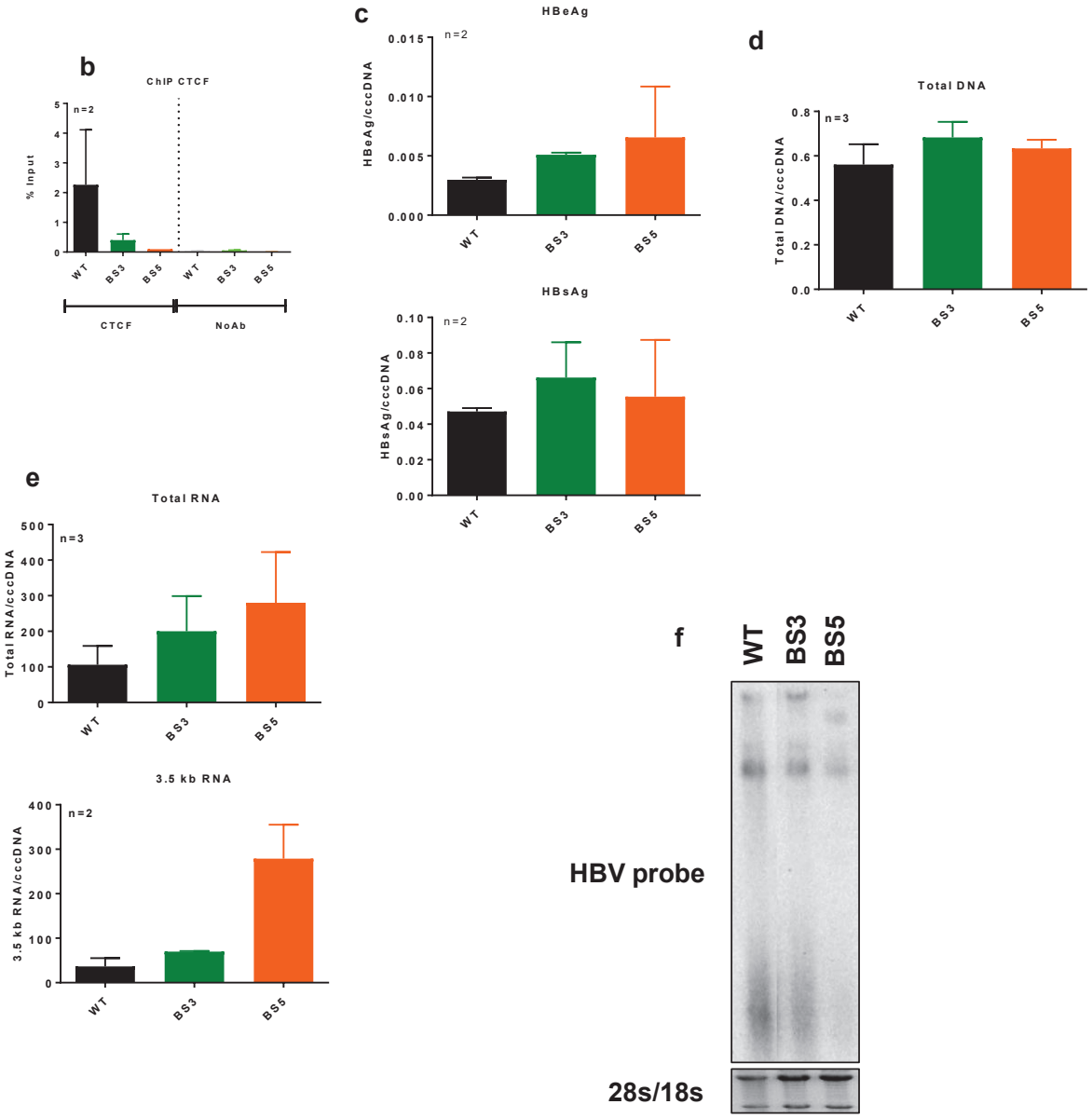


Figure 3

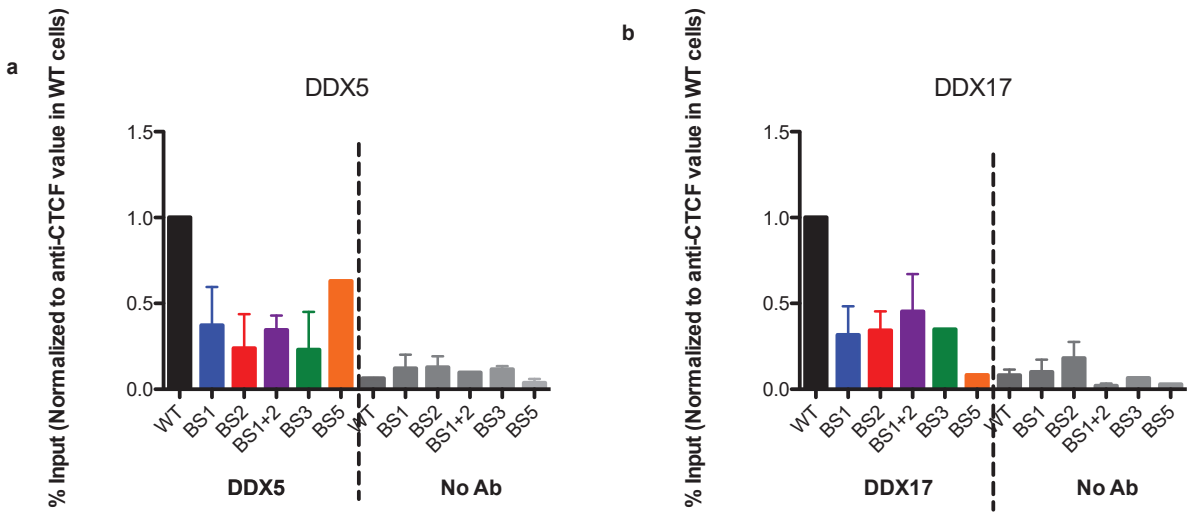

a
% Input (Normalized to anti-CTCF value in WT cells)
DDX5
1.5
1.0
0.5
0.0
WT
BS1
BS2
BS1+2
BS3
BS5
DDX5
No Ab
b
% Input (Normalized to anti-CTCF value in WT cells)
DDX17
DDX17
No Ab

Figure 4

2.Supplementary results

In this study, we demonstrated that CTCF can be recruited at different binding sites located along the HBV genome and conserved across the major HBV genotypes. Among these sites, four were mutated, either alone or in combination. Mutation of CTCF binding sites, achieved by the generation of WT or mutated cccDNA-like molecules by minicircle strategy (mcHBV), impairs CTCF recruitment to cccDNA. Moreover, mutation of CTCF binding sites affects HBV RNA processing, and results in longer or shorter viral transcripts in Northern-Blot analysis. Finally, we demonstrated that the loss of recruitment of CTCF onto cccDNA triggers a loss of recruitment of DDX5 and DDX17, previously demonstrated to regulate HBV RNA processing (Project 1).

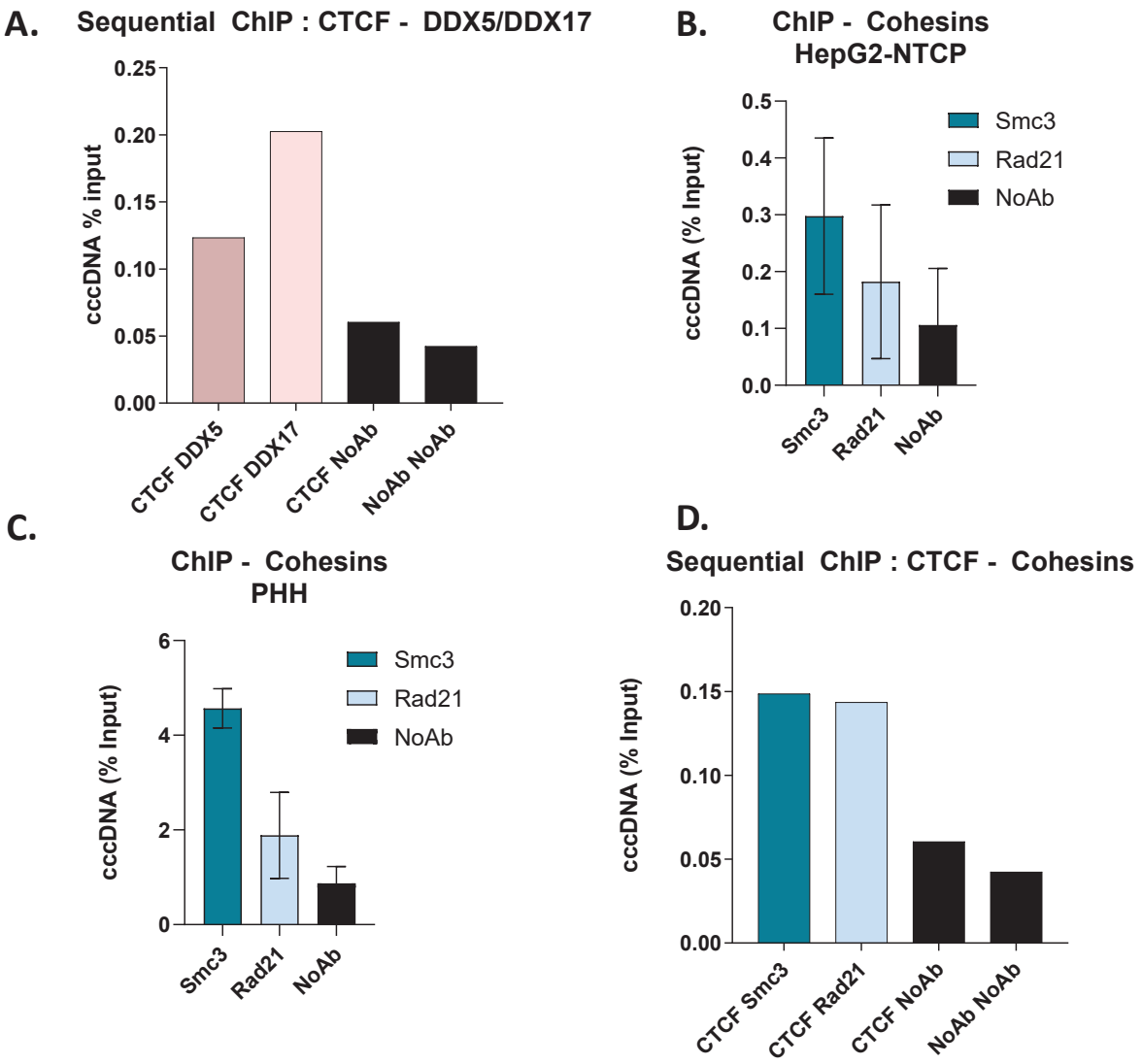
CTCF, DDX5, DDX17 and Cohesins are concomitantly present on a same cccDNA molecule.

CTCF is an insulator protein which mediates its functions, notably in IGF2/H19 imprinting, thanks to DDX5 and probably DDX17 (Yao 2010). We demonstrated that mutations in CTCF binding sites result in loss of CTCF recruitment to cccDNA concomitantly to a loss of DDX5 and DDX17 association. These results suggest that CTCF and DDX5/DDX17 are concomitantly bound to the same cccDNA molecule and that, probably, CTCF mediates DDX5/17 resruitment. To test this hypothesis, we performed sequential ChIP-qPCR. Sequential ChIP-qPCR experiments consist in a first immunoprecipitation of CTCF-associated DNA, followed by a second immunoprecipitation of the eluate with antibodies specific for other target proteins. These experiments allow the detection of chromatin concomitantly associated to different proteins of interest. In our study, we investigated whether CTCF and the DDX5/17 were present on a same cccDNA molecule. We thus performed sequential ChIP-qPCR with CTCF, followed by either DDX5 or DDX17. CTCF-NoAb and NoAb-NoAb are the negative controls for each condition. We observed a higher signal for CTCF-DDX5 and for CTCF-DDX17 conditions than for the negative control conditions, witnessing the concomitant presence of CTCF and DDX5, and CTCF and DDX17 in a same cccDNA molecule (Supp Fig2A).

CTCF is also known to be a structural protein involved in chromatin looping notably via its interaction with the cohesin complex (Barrington et al., 2017). Cohesin complex is

composed of three subunits, namely SMC1, SMC3 and RAD21 ***(Barrington et al., 2017)***. Cohesins are essential for the insulator function of CTCF (Yao 2010). As we demonstrated that CTCF is associated to cccDNA, we therefore asked whether cohesins could also be present on this minichromosome. In HBV-infected HepG2-NTCP cells, ChIP-qPCR experiments performed seven days post-infection revealed the presence of Smc3 and Rad21 on HBV cccDNA (Supp Fig2B). These results were confirmed in PHH (Supp Fig2C). Moreover, sequential ChIP-qPCR experiments with CTCF immunoprecipitation followed by either SMC3 or RAD21 immunoprecipitation, revealed that CTCF and SMC3, and CTCF and RAD21 are concomitantly bound to a same cccDNA molecule (Supp Fig2D).

Collectively, these data suggest that CTCF recruitment onto cccDNA is concomitant with the presence of DDX5/17 and the cohesin complex.



Supplementary Figure

Supplementary Figure

A. HepG2-NTCP cells were infected with HBV at a MOI of 250. 7 days post-infection, sequential ChIP experiments were performed using the anti-CTCF antibody followed by a second immunoprecipitation with an anti-DDX5 antibody (dark pink bar), an anti-DDX17 antibody (light pink bar) or without antibody (right black bar) that serves as a first negative control. A no-antibody immunoprecipitation followed by a second no-antibody immunoprecipitation is a second negative control (left black bar). qPCR amplification was

used to quantify immunoprecipitated cccDNA molecule. The bars represent a unique experiment. The error bars represent the standard error mean.

B. HepG2-NTCP and **C**. PHH cells were infected ith HBV at a MOI of 250 during 7 days. ChIP experiments were performed using anti-SMC3 (dark blue bars) and anti-RAD21 (ligh blue bars) antibodies. No-antibody condition was used as negative control (black bars). The bars represent the average of 3 independent experiments in HepG2-NTCP and 2 independent experiments in PHH. The error bars represent the standard error mean.

D. HepG2-NTCP cells were infected with HBV at a MOI of 250. 7 days post-infection, sequential ChIP experiments were performed using the anti-CTCF antibody followed by a second immunoprecipitation with an anti-SMC3 antibody (dark blue bar), an anti-RAD21 antibody (light blue bar) or without antibody (right black bar) that serves as a first negative control. A no-antibody immunoprecipitation followed by a second no-antibody immunoprecipitation is a second negative control (left black bar). qPCR amplification was used to quantify immunoprecipitated cccDNA molecule. The bars represent a unique experiment. The error bars represent the standard error mean.

Discussion

Discussion

HBV chronically infects more than 250 million people in the world and is a major health burden by being the leading cause of HCC, causing more than 800 000 deaths each year (Nassal, 2015). Despite the existence of prophylactic vaccine and efficient antiviral treatments, HBV persists in hepatocytes via its transcriptionally active replicative intermediate, the so-called cccDNA. This viral minichromosome is the template for the viral transcription and is not affected by current treatments, leading to a perpetual transcription and subsequent secretion of viral antigens. HBV transcription has to be finely regulated to achieve correct HBV transcript synthesis. The mechanisms at the basis of this complex regulation are poorly understood and under-investigated. Regarding the complexity of HBV gene expression, many questions remain unresolved regarding:

- the host factors involved in the correct cccDNA transcriptional activity, in particular transcription termination

- the host factors involved in HBV pre-mRNA splicing events and the role of HBV spliced variants in HBV biology

- a putative role of cccDNA tri-dimensional organization in transcriptional regulation and HBV RNA co-transcriptional processing

One of the main interests of the laboratory is to investigate the host factors regulating cccDNA activity and persistence *in vivo*. A project coupling cccDNA-chromatin immunoprecipitation and Mass-Spectrometry analysis identified a number of host factors associated to transcriptionally active cccDNA in infected primary human hepatocytes (Testoni et al. *See below* ***Appendice 1***).

During my thesis internship, I pursued the aim of better understanding the role of selected cccDNA-associated factors in cccDNA transcription and viral RNA metabolism. To do so, I investigated the role of two cellular RNA helicases belonging to the DEAD-box family well described to participate to several processes of gene expression regulation (Bourgeois et al., 2016). Notably, two prototypic members of this family, DDX5 and its paralog DDX17, are

implicated in splicing regulation (Dardenne et al., 2014b) and have been associated to the modulation of the lifecycle of different viruses (Wenyu Cheng et al., 2018). Differently from many studies interrogating the cellular and viral roles of DDX5 and DDX17 independently, we decided to concomitantly investigate their function in HBV life cycle, and particularly in the viral RNA processing, given the fact that DDX5 and DDX17 operate as a heterodimer and have redundant roles. Moreover, DDX5, and probably DDX17, are part of an insulator complex and participate to the insulator function of CTCF (H. Yao et al., 2010). Genome-wide analysis revealed that DDX5 binding on chromatin extensively overlaps CTCF binding sites (H. Yao et al., 2010). Interestingly, CTCF also regulates co-transcriptional processes, such as splicing (Tim R Mercer et al., 2013; Shukla et al., 2011). In a second time, we therefore investigated whether CTCF could be associated to cccDNA and display regulatory functions in cccDNA biology and whether DDX5 and DDX17 could exist as a cccDNA-associated complex together with CTCF.

Characteristics of experimental models

We silenced DDX5 and DDX17 concomitantly by siRNA transfection in order to avoid redundancy between these two paralogs (Giraud et al., 2018). Most of the experiments were conducted in HepG2-NTCP cells, a hepatoma cell line broadly used to study the complete viral life cycle thanks to the stable expression of the HBV receptor NTCP. Indeed, HepG2-NTCP cells are a good surrogate for PHH, the natural host of HBV, as they present robust level of viral infection and a high number of cccDNA copies (E. Verrier et al., 2016). Moreover, HepG2-NTCP cells are easily available and simple to culture and infect. While the HepG2-NTCP model has been widely used to screen antiviral drugs targeting cell entry, such as Myrcludex B, and for the discovery of host factors involved in HBV replication, HepG2-NTCP are transformed cells different from PHH in their metabolic properties (Luangsay et al., 2015). This model nevertheless appears relevant for our studies as most of our results upon DDX5/17 depletion were further confirmed in PHH, thus allowing us to trustfully conduct mechanistic experiments in HepG2-NTCP cells.

Confirming the role of DDX5 and DDX17 in HBV RNA processing in vivo in a murine model would reinforce our demonstrations. Moreover, validating in patient samples the 3' readthrough observed in wild-type HBV transcripts in cultured HepG2-NTCP and PHH would highlight viral RNA properties ignored so far and help the development of novel biomarkers for cccDNA transcriptional activity.

Importance of viral splicing

As mentioned in the introduction, HBV RNAs, and notably pgRNA, can be spliced. Several spliced isoforms have been identified in the different HBV genotypes and vary in length and nature depending on the donor and acceptor splice sites used. To date, HBV splicing was investigated by two main strategies. The first one relies on PCR amplification of all the spliced isoforms by using flanking primers located upstream of the main donor site (position 2067) and downstream of the main acceptor site (position 489). Discrimination between the different variants is achieved by differential migration on agarose gel according to their length (Bayliss et al., 2013; Duriez et al., 2017). Getting rid of full-length transcript contamination depends on elongation time. However, PCR amplification efficiency depends on the length of the amplicon and thus induces a bias in complete variant detection. The second technique relies on complete genome amplification by reverse PCR on encapsidated DNA extracted from CHB patient sera (Chen et al., 2015; Soussan et al., 2008). Nevertheless, this technique identifies DNA molecules and assumes to reflect pgRNA splicing, without directly identifying what happens at RNA level.

We used a non-biased approached based on RNA sequencing to characterize HBV transcripts. Concerning HBV transcriptome, RNA-seq allowed the detection of viral transcripts, and notably X transcript, in PHH after four hours of infection and in CHB patient samples (Niu et al., 2017). We used for the first time RNAseq data to identify splicing junctions and uncovered five junctions that were mostly skipped upon DDX5/17 depletion. To date, a little is known about HBV RNA splicing, and what is the role of spliced variants and which are the mechanisms regulating HBV pre-mRNA splicing remain an unresolved question.

HBV is not the only virus generating spliced isoforms. Indeed, alternative splicing is a crucial mechanism for several viruses that allows the expression of multiple proteins from a small and compacted genome. Depending on the virus, spliced variants do not have the same functions. For HIV-1, that encodes 40 mRNAs among which unspliced, partially spliced and completely spliced transcripts are generated, alternative splicing is essential for viral replication. Indeed, several spliced-encoded proteins act at different steps of HIV-1 life cycle and mutating splicing regulatory elements in HIV-1 genome inhibits viral replication (Martin Stoltzfus, 2009). Noteworthy, DDX17 has been described in different studies to regulate HIV-1 splicing processes (Lorgeoux et al., 2013; Sithole et al., 2018). Moreover, Herpes Simplex Virus 1 (HSV-

1) harbors six well-described spliced variants and several donor and acceptor splicing sites have recently been identified by long-read sequencing, using MinION™ Nanopore Technology (Tang et al., 2019; Tombácz et al., 2019). One of these spliced variants encodes a secreted glycoprotein that is suggested to serve as HSV-1 virulence factor (Sedlackova et al., 2008).

Viral alternative splicing is also responsible for multiple virus-mediated oncogenicity. The Papilloma Virus (HPV) is an episomal virus that encodes 20 mRNAs among which some are spliced and code viral proteins that have been identified in patient samples. HPV spliced variants encode oncoproteins implicated in HPV-mediated cancer progression (Graham and Faizo, 2017). Some Polyomaviruses, such Simian Virus 40 (SV40) generates four mRNAs from splicing of a unique mRNA, among which two spliced-deriving proteins are the oncogenic large and small Tumor antigens (Ajiro and Zheng, 2014). Epstein-Barr Virus (EBV), another episomal virus, encodes the LMP1 oncoprotein coming from a doubly-spliced transcript. Interestingly, the single splicing of the same pre-mRNA generates another protein isoform that acts as negative regulator of LMP1, but which is rarely translated in EBV infections (Ajiro and Zheng, 2014). Finally, the oncogenic property of Human T-lymphotropic Virus (HTLV-1) hinges on Tax oncoprotein encoded by a double-spliced mRNA (Ajiro and Zheng, 2014). It is important to note that besides the uncovering of spliced variants for all these viruses, and the elucidation of their functions for some of them, the viral and cellular factors regulating alternative splicing are not fully described and remain to be better investigated.

In HBV infection, at least sixteen spliced variants deriving from pgRNA and four spliced variants deriving from preS2 have been identified (Candotti and Allain, 2016; Chen et al., 2015). Spliced variants are associated with HCC development. Indeed, spliced HBV RNAs are found in 80% of patients serum samples and their level increase of 0.1% each year prior HCC development (Bayliss et al., 2013).

Only three spliced variants are known to encode proteins. The major spliced isoform SP1, measuring approximately 2 kb, is well-described to encode the so-called HBSP. Anti-HBSP antibodies are present in the serum of chronic carriers and their detection is correlated with HBV DNA replication. Moreover, anti-HBSP antibodies strongly correlate with liver fibrosis and TNF-α secretion (Duriez et al., 2017; Soussan et al., 2003). HBSP has also been linked to cell migration and invasiveness, notably by interacting with Cathepsin B. HBSP overexpression in hepatoma cell line promotes cell motility and migration (Chen et al., 2012).

A doubly-spliced pgRNA-deriving 2.2 kb-long variant corresponds to the SP7 and is translated into the doubly-spliced protein HBDSP. HBDSP has been reported to activate some regulatory elements of HBV genome, notably CURS, Enhancer I and the PreS1 and PreS2 promoters, but not BCP and X promoter (Chen et al., 2010). Importantly, these results were generated by luciferase reporter assay under HBDSP overexpression and would deserve confirmation in infectious models. Finally, a fusion protein containing the N-ter of Pol and the C-ter of S antigen, called Polymerase-Surface fusion protein (P-S FP) is encoded by the singly spliced isoform SP14. This 43-kDa protein is glycosylated and can substitute L-HBsAg in secreted Dane and 22 nm particles (Huang et al., 2000). Moreover, P-S FP appears to colocalize with the endoplasmic reticulum, the microtubules, the intermediate filaments, and also partially with NPC, and seems to inhibit viral replication by a yet-to-be described mechanism (Park et al., 2008). This protein is similar to an L protein encoded by a spliced isoform of DHBV pgRNA (Obert et al., n.d.). Interestingly, HBV P-S FP does not contain the myristoylation site of PreS1 domain that is known to be crucial for viral entry (Yan et al., 2012b).

Besides SP1, SP14 and SP7, functions of the other spliced variants deriving from HBV pgRNA have not been investigated.

Our RNAseq data revealed that the exonic sequences at position 2237 to 2447, 2350 to 2447 and 2350 to 2471 are more skipped when cells are depleted for DDX5 and DDX17. While short read sequencing, such as RNAseq, cannot permit to say to which transcripts these exons belong, description of HBV spliced isoforms in the literature (Candotti and Allain, 2016; Chen et al., 2015) allowed us to assess that these exons are part of HBV spliced isoforms SP2, SP10, SP12, SP15 and SP16. Interestingly, skipping the exon 2237-2447 from SP12, exon 2350-2447 from SP2 and exon 2350-2471 from SP16 leads to the switch to the singly spliced isoform called SP3 (donor site: 2067, acceptor site: 489). SP3 is one of the three major products of pgRNA splicing, together with SP1 and SP2, representing 70% of pgRNA-deriving spliced variants in patient sera (Chen et al., 2015). SP3 has not been yet described to encode any proteins, and its implication in HBV replication cycle remains obscure. Moreover, skipping exon 2337-2447 from SP15 and exon 2350-2447 from SP10 leads to a switch to the SP17 variant (donor site: 2067, acceptor site: 282). Interestingly, the first group to describe this seventeenth spliced isoform was Chen and colleagues, that identified and named for the first time SP15, SP16 and SP17 spliced isoforms after identification in sera of CHB patients, either

treated or not with IFN (Chen et al., 2015). In this study, spliced isoform level in the serum of CHB patients was inversely correlated to IFN responsiveness (Chen et al., 2015). However, this correlation was not attributed to particular isoforms. Particularly, despite that SP15, SP16 and SP17 were identified for the first time, their implication, and implication of SP2, SP3, SP12 and SP10 in HBV replication were not investigated.

Elucidating the role of all HBV spliced variants, and the precise mechanisms of the already described viral proteins encoded by spliced variants would supply knowledge on their role in HBV viral replication and associated pathogenesis. Our results propose DDX5 and DDX17 as regulators of HBV RNA alternative splicing. Nevertheless, it remains unclear how DDX5 and DDX17 regulate HBV RNA splicing. Furthermore, we could wonder what is the biological significance of favoring SP13 and SP17 isoforms in spite of SP2, SP10, SP12, SP15 and SP16 upon DDX5/17 silencing. We demonstrated that DDX5 and DDX17 are associated to cccDNA and to viral RNA, however, precise binding sites cannot be investigated so far due to technical limitations (*see below*). Moreover, spliced isoforms have been linked to HCC development (Bayliss et al., 2013) and HBSP is described to provide proliferative properties to cells (Chen et al., 2012). Direct implication of splicing in HBV-mediated HCC is yet to be confirmed. Investigating the role of the alternatively spliced variants observed in wild-type and DDX5/17 depleted cells still needs to be performed and would provide information on the consequences of DDX5/17 silencing in HBV-mediated pathogenesis. DDX5 was demonstrated to be downregulated in a subset of HCC tumors (Zhang et al., 2016). It would be interesting to investigate whether the isoforms identified in DDX5/17 depleted cells are correlated with HCC development in patient samples.

Importance of 3' end processing

Polyadenylation of cellular mRNA is required for RNA fate regulation, notably stability and translation. Indeed, the polyA tail is associated with PABP that connect m7G cap to circularize the mRNA thereby promoting translation efficiency. Interestingly, canonical histones mRNAs are not polyadenylated but rather contain a conserved stem-loop structure that mimicks polyA tail functions, notably during translation (Marzluff et al., 2008). Genome-wide analysis uncovered than dozens of mRNA, notably encoding Zinc-finger proteins, but also lncRNA and other RNA species, do not contain the canonical polyadenylated tail (Yang et al., 2011). Rather, such non-polyadenylated transcripts, like MALAT1 and MENβ, harbor an RNA secondary triple

helical structures that compensates the lack of polyA tail. This secondary structure insures RNA stability, together with nuclear export and translation regulation (Wilusz, 2016). Strikingly, CPSF6 and NUDT21, in competition with hnRNPK are involved in the 3' end formation regulation of MENβ. Indeed, MEN lncRNA exists as two different isoforms depending on their 3' end. MENε is shorter and polyadenylated while MENβ is longer and ends with a cloverleaf secondary structure similar to MALAT1. On the contrary, hnRNPK recruitment to MENε PAS blocks CFI recruitment and favors the synthesis of the non-polyadenylated MENβ isoform (Wilusz, 2016). hnRNPK is one of the identified targets in the ChROP project (**Appendice 1**) and we further validated its association to cccDNA upon HBV infection in HepG2-NTCP and PHH. Preliminary analysis suggests that the viral RNA harboring a longer 3' correspond to non-polyadenylated HBV transcripts. Interestingly, some non-polyadenylated cellular transcripts contain 3' UTR that extend beyond the PAS along several hundreds of nucleotides (Yang et al., 2011). Moreover, we correlated the use of the proximal PAS to decreased recruitment of CPSF6 and NUDT21 upon DDX5/17 depletion. As CPSF6 and NUDT21 are involved in distal PAS choice (Hardy and Norbury, 2016) our results suggest a role of CPSF6 and NUDT21 in the choice of HBV distal PAS. Moreover, DDX5 and DDX17 are involved in transcriptional termination (Katahira et al., 2019; Mersaoui et al., 2019; Padmanabhan et al., 2012) and strikingly, DDX5, DDX17 and CPSF6 belong to the same termination and export complex THO interactome (Katahira et al., 2019). However, whether CPSF6 and NUDT21 directly regulates HBV PAS choice remains to be elucidated. Developing specific knockdown of these two factors would bring to light their role in HBV RNA 3' processing. Moreover, it would be interesting to investigate whether long HBV transcripts possess secondary RNA structures that compensate the lack of polyA tail.

Moreover, the 3' UTR length, which represents the region between the STOP codon and the end of the transcript, is important for RNA regulation as it is targeted by several miRNA. The length of 3' UTR is associated to the proliferative state of cells. Under proliferative conditions, the 3' UTR is generally shorter. That is the case for B and T cells that have to be rapidly activated to properly insure immune response (Sandberg et al., 2008). This observation is also true for transformed and cancer cells that exhibit shorter mRNA isoforms to acquire a higher stability and to produce more proteins. Importantly, in proliferative cells, the proximal PAS is

preferentially used to generate shorter 3' UTR (Curinha et al., 2014). On the contrary, during development and differentiation, transcripts harbor longer 3' UTR (Curinha et al., 2014).

Viruses also possess polyA tail and can differentially use distal or proximal PAS. During HPV replication cycle, the distal PAS is inhibited in early infection in order to prevent the expression of viral proteins involved in late stage of infection. In later infection, hnRNPC binds the two cleavage and polyadenylation factors CPSF30 and Fip1 to inhibit early PAS, thereby causing readthrough in viral transcription (Nilsson et al., 2018).

As abovementioned, 3' RACE PCR product sequencing by MinION^{TM} Nanopore Technology surprisingly revealed an extensive 3' readthrough in HBV transcripts which is lost upon DDX5/17 depletion. The 3' readthrough observed in HBV transcripts is thus not exclusive to HBV, but rather frequently exists in cellular transcripts, and can be of benefit in other viral infections. Moreover, the loss of 3' readthrough observed upon DDX5 and DDX17 depletion could be connected to the slight increase in viral transcripts levels and to the accumulation of viral particles in the cytoplasm of depleted cells. Indeed, presence of longer 3'UTR increases the probability of miRNA target sequences presence to further fate regulation of the transcript. Moreover, miRNA association to 3' UTR of mRNA induces translation downregulation, either by blocking this process, of by mediating transcript degradation. We thus investigated potential miRNA target sequences in the longer part of HBV RNA 3'UTR, absent from HBV transcripts upon DDX5/17 depletion.

With MiRBase in silico prediction tool, we identified that longer 3' UTR harbors binding sites for six miRNAs, namely miR169, miR72, miR727, miR500, miR222 and miR548. Among them, miR500, miR222 and miR548 were reported to be modulated in HCC patients. Firstly, miR500 increases in HCC patients by the presence of epigenetic active histone PTM on its promoter. MiR500 is involved in cell proliferation by targeting and mediating degradation of the pro-apoptotic factor BID (Bao et al., 2018). On the contrary, miR222 is downregulated in HBV-related HCC, notably by HBx. MiR222 targets p27 (Kip1), whose downregulation correlates with HCC patient poor prognosis (Bandopadhyay et al., 2014). MiR548 is also downregulated in HBV-related HCC and seems downregulated by HBx. MiR548 targets HBx-Interacting protein (HBXIP) directly in its mRNA 3'UTR (Hu et al., 2016). HBXIP is a negative regulator of HBx but is also described to be an oncoprotein promoting tumorigenesis and progression (Hu et al., 2016). In this study, it is not clear whether upregulated HBXIP subsequently impacts HBx, or

favors tumorigenesis. Collectively, these predictions suggest that shorter HBV transcripts induced by DDX5 and DDX17 depletion lose these miRNA target sequences and would harbor a less downregulated translation. Investigating whether these three miRNA candidates regulate HBV transcripts would help understanding how the viral transcripts are post-transcriptionally regulated in hepatocytes. Notably, does the silencing of these candidates, or other potential miRNAs, increase HBV RNA levels and viral proteins translation, as observed upon DDX5/17 silencing?

Moreover, one could speculate that transcription termination site could vary along HBV infection natural course, as it has been reported for HPV life cycle. Investigating the length of HBV RNA 3' UTR in HBV infection kinetics would provide information in the biological significance of such viral RNA 3' readthrough. Indeed, is HBV RNA 3' UTR shorter during the early stage of infection, leading to an efficient infection establishment by optimal viral protein synthesis? On the contrary, is the 3' UTR lengthening upon chronic stages of infection necessary to maintain discrete expression of HBV proteins and maintain low antiviral cellular response?

Model of HBV RNA transcription termination by DDX5/17

During my PhD project, I also demonstrated that DDX5 and DDX17 regulate HBV RNA transcription termination. Indeed, in wild-type condition, we observed a so-far ignored extensive 3' readthrough of approximately 50% of HBV transcripts that completely disappears upon DDX5/17 depletion. We correlated the 3' UTR shortening of HBV transcripts in siDDX5/17 to a loss of recruitment of DDX5 and DDX17 onto HBV RNA together with the loss of recruitment of CPSF6 and NUDT21, two subunits of CFI complex involved in PAS choice (Hardy and Norbury, 2016). Moreover, we demonstrated that DDX5/17 silencing triggers an increase in cytoplasmic viral particles.

Collectively, our data suggest the following working model in **Figure 33**:

In wild-type condition, RNA Pol II stops at HBV PAS (position 1916-1921) approximately half of the time. RNA Pol II pausing at PAS leads to transcription termination and polyadenylation of neosynthesized HBV RNA. However, RNA Pol II pausing at PAS is not systematic and transcription continues in around 50% of time until encountering two TGTA motifs present in HBV DNA template. Recruitment of DDX5/17 near these motifs favors the recruitment of

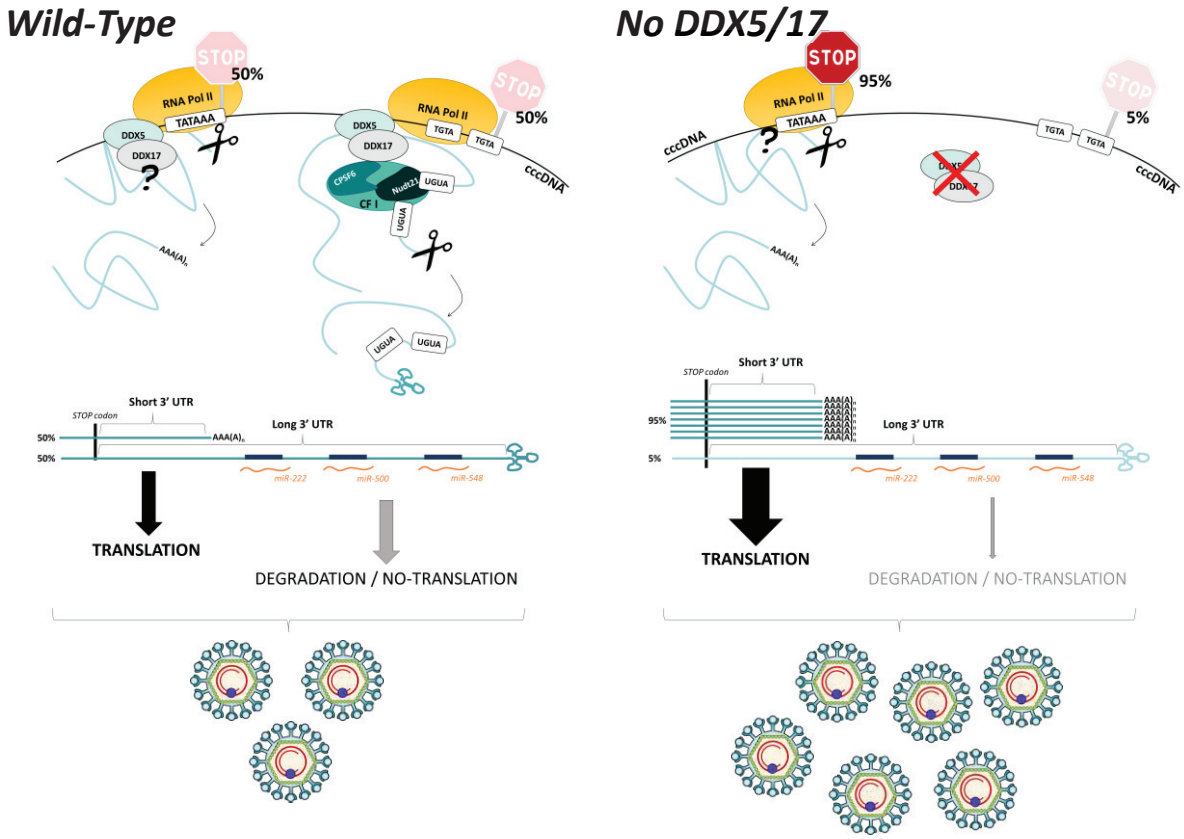


Figure 33 : Working model of DDX5/17-mediated mechanism involved in HBV RNA 3' end processing

Wild-type: RNA Pol II can pause at the PAS in 50% of cases. This mechanism seems to require a yet-to-be described role of DDX5/17. Pol II pausing at the PAS leads to canonical 3' end processing (represented by the scissor) and generation of polyadenylated viral RNA with a short 3' UTR. However, in 50% of cases, RNA Pol II does not stop at the PAS and continue until TGTA motifs. DDX5 and DDX17 are associated to cccDNA and/or HBV RNA near these motifs and recruit CPSF6 and NUDT21 to their binding motifs UGUA onto viral RNA. CPSF6 and NUDT21 trigger 3' end processing at this location, yet, the absence of PAS in this region triggers a non-canonical 3' end processing with folding in secondary RNA structure. Longer HBV RNA transcripts harbor longer 3' UTR and additional target sequence to miRNA (target sequences indicated with dark blue squares and miRNA annotated in orange). A subsequent balance between translation and degradation induces a certain number of novel viral particles.
When DDX5/17 are depleted, only the canonical 3' end processing occurs and leads to increased proportion of short polyadenylated viral transcripts, and subsequent enhanced translation. In this context, more viral particles are produced.

CPSF6 and NUDT21 to their binding motif UGUA onto neosynthesized HBV RNA. Presence of CFI complex at UGUA motifs triggers 3' end formation, yet, no canonical PAS is present in this region. We thus propose that transcription ends without recruiting the polyadenylation machinery. Instead, HBV longer transcripts end with a specific secondary structure, as described for non-polyadenylated cellular transcripts (Wilusz, 2016; Yang et al., 2011). 3' end secondary structure stabilizes non-polyadenylated HBV transcripts. These longer transcripts therefore harbor a longer 3' UTR that contains target sequences to three microRNAs, namely mir22, mir500 and miR548. In this context, we assume that translation mostly occurs from polyadenylated transcripts, while non-polyadenylated HBV RNA mostly undergo degradation or translation blocking. The translation rate insures synthesis of a certain number of viral particles.

When DDX5 and DDX17 are depleted, CPSF6 and NUDT21 are not recruited to UGUA motifs, leading to the loss of longer HBV transcripts. Instead, almost all of the transcripts end at the PAS and are polyadenylated. These shorter transcripts harbor a short 3' UTR that does not contain additional miRNA target sequences and that allow higher production of viral proteins and subsequent viral particles assembly.

Noteworthy, the direct implication of DDX5 and DDX17 in the recognition of the PAS remains obscure so far. DDX5 and DDX17 are required for RNA secondary structure resolution, such as G-quadruplex. G-quadruplex are highly dynamic structures modulated by cellular factors. G-quadruplex resolution by DDX5/17 is essential for proper alternative splicing (Dardenne et al., 2014). Importantly, another project in our team investigates G4 structures along HBV genome and HBV transcripts. Interestingly, a putative G4 structure is located few bp upstream of the PAS and has been validated to also be present on viral RNA. We could therefore suggest that DDX5/17 presence in WT condition leads to G4 resolution on some HBV RNA molecules. Subsequently, when the G4 is stable, RNA Pol II pauses near PAS and triggers transcription termination and following 3' end processing. RNA with stable G4 structure are thus shorter

and polyadenylated. On the contrary, when the G4 is resolved by DDX5/17 helicase activity, RNA Pol II can pass along this unwound structure and pauses when it encounters a proper motif, namely UGUA. In this latter condition, HBV RNAs are longer and do not harbor a polyadenylated tail. Being dynamic, this process could lead to more or less stable G4 structures along HBV RNA and consequent presence of short and long transcripts. When DDX5

and DDX17 are knocked down, G4 resolution does not occur anymore, leading to almost all HBV RNA ending at the PAS.

Pursuing investigation of G4 structure and dynamic in HBV genome and viral RNA, and implication of DDX5/17 in their maintenance, would supply further information on the mechanism leading to HBV RNA 3' readthrough.

Technical limitations

We used next generation sequencing (RNAseq) in combination with third generation sequencing (MinION™ Nanopore Technology) to characterize HBV transcripts in WT and DDX5/17 depleted conditions. Our RNAseq data demonstrated that DDX5/17 is involved in HBV RNA splicing. Nevertheless, short-read sequencing does not permit to assess to which transcript belongs each splicing event. Precisely discriminating the different HBV mRNA species would be possible using single-molecule sequencing technologies.

To date, few studies took advantage of Nanopore technology in HBV biology and investigated only HBV genome (McNaughton et al., 2019; Sauvage et al., 2018). For the first time, we were able to apply third generation sequencing using Nanopore technology to HBV transcript 3' termination characterization and accurately located the termination site of most of the viral transcripts 13 +/-1 nt downstream of the PAS. Strikingly we also uncovered an extensive 3' readthrough in viral transcripts. Importantly, to sufficiently enrich for viral transcripts, we performed PCR amplification of 3' RACE products. PCR amplification may induce bias during elongation step.

Indeed, to date, getting rid of PCR amplification, and subsequent technical bias, remains a challenging issue in HBV RNA sequencing. Specific enrichment for the viral transcripts and direct single-RNA molecule sequencing would finally permit a complete characterization of HBV RNAs. Capture of viral transcripts, with HBV-specific biotinylated probe precipitation, would contribute achieving such characterization. Enrichment of viral transcripts would also allow deeper sequencing by RNA-seq and better description of splicing events and following statistical analysis.

We also demonstrated for the first time that DDX5 and DDX17 recruit the two cleavage and polyadenylation factors CPSF6 and NUDT21 that are involved in the choice of distal PAS.

Besides demonstrating the binding of DDX5, DDX17, CPSF6 and NUDT21 on HBV RNA and cccDNA, CLIP and ChIP results did not allow the precise localization of these factors.

CLIP experiments consist in UV cross-linking that only reveals direct binding of a protein to RNA. Identifying the precise binding sites of DDX5, DDX17, CPSF6 and NUDT21 would confirm our data and reinforce their described role in HBV RNA splicing and 3' processing. For example, it would be important to demonstrate that CPSF6 and NUTD21 do associate to the 5'-UGUA-3' motifs identified in 3' readthrough and may allow to detect a potential preference of binding to one or another motif along the 3' readthrough thereby explaining the choice in HBV RNA termination site. Photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) is a method that relies on photoactivable nucleoside incorporation in nascent RNA. Upon UV cross-linking, photoactivable nucleosides located in RBP binding sites undergo characteristic mutations and allow identification of RBP binding sites with a single-nucleotide resolution. This method enables transcriptome-scale investigation when combined to NGS (Danan et al., 2016). Implementing such experiments for HBV RNA would contribute to the generation of a precise mapping of RBP on viral transcripts.

Furthermore, we observed that RNA Pol II is similarly associated to HBV cccDNA in control and DDX5/17 knockdown conditions (data not shown), yet the specific phosphorylation of Pol II CTD has not been investigated, neither the precise localization of RNA Pol II. Knowing that the phosphorylation status of RNA Pol II-CTD is a key regulator in co-transcriptional processes and modulates specific factor recruitment, investigating whether some changes occur upon DDX5/17 depletion would bring further information in HBV RNA processing and potential recruited factors.

Co-transcriptional regulation and processing of pre-mRNA is also modulated at the template DNA level by recruitment of specific factors to the chromatin. Investigating the precise binding site of DDX5 and DDX17 would bring further evidences on their role in HBV RNA processing. Nevertheless, efficient sonication of cccDNA is rarely achievable due to the highly compact structure of the viral minichromosome. Enzymatic digestion, for example using MNase, as described by Tropberger in 2015, has allowed experiments such as ChIP-seq and identification of nucleosome positioning and histone enrichment on cccDNA (Tropberger et al., 2015). However, in this study, no proof of an efficient digestion of cccDNA was demonstrated, and unfortunately, digesting cccDNA with MNase treatment was technically not feasible in our

hands. Implementation of cccDNA enzymatic digestion or sonication would bring powerful tools to better understand cccDNA biology and implications of viral and cellular factors in its regulation.

HBV episome, a 3D-structured genome?

We demonstrated that CTCF is associated to cccDNA and mutations in CTCF binding sites impaired the recruitment of CTCF on the viral minichromosome. However, where CTCF is exactly recruited, notably, whether the mutation in one CTCF binding site affects only the binding at this site or also impairs CTCF recruitment on the other sites remain unresolved points. Notably, we identified convergent CTCF binding sites, feature commonly involved in chromatin looping (de Wit et al., 2015). CTCF is extensively implicated in intergenic long-range interactions and intragenic DNA-looping (Cavalli and Misteli, 2013). The presence of an antisense CTCF binding site (BS1) raises the question of possible DNA loops along HBV genome, by association with one of the seven (BS2-8) available positive strand-located binding sites. Moreover, DNA looping is achieved when two convergent CTCF binding sites recruit CTCF proteins in an oriented manner. Subsequently, cohesin complex forms a “handcuff” around DNA loop and stabilizes it by interacting with CTCF C-ter domain (Barrington et al., 2017). Interestingly, we also demonstrated the presence of Smc3 and Rad21, two subunits of the cohesin ring, on cccDNA, concomitantly with the association of CTCF. Besides cohesin complex, DDX5, and probably DDX17, are important for CTCF insulator functions (H. Yao et al., 2010). Moreover, genome-wide analysis reported that CTCF binding sites are frequently occupied by DDX5 (H. Yao et al., 2010). Strikingly, mutations in CTCF binding site on HBV genome also affect DDX5 and DDX17 recruitment onto cccDNA, but in which region both of these factors are recruited cannot be identified so far.

The presence of CTCF and identification of convergent CTCF binding sites on HBV genome, together with the demonstrated association of cohesin complex subunits Smc3 and Rad21, and DDX5 and DDX17 on cccDNA attractively suggest that HBV genome might form chromatin loops. Although average of chromatin loop length identified along human genome is approximately 100 kb, the question of chromatin looping within viral genome, which are much smaller than human genome, has already been addressed and revealed the participation of CTCF in viral episome 3D structure. CTCF has been suggested to participate to HSV-1 genome (152 kb-long) insulator functions where CTCF binding sites are independently regulated

(Washington et al., 2018). Moreover, CTCF promotes the viral transcription by facilitating RNA Pol II elongation and by preventing repressed chromatin state of HSV-1 (Washington et al., 2018). Notably, CTCF binding sites are located between neighboring genes (Lang et al., 2017) that further suggests insulator functions. However, CTCF functions in HSV-1 genome organization have been partially demonstrated, notably with luciferase reporter assay, and remain to be fully determined. On the contrary, the structural organization of HPV-18 mediated by CTCF has been well described. Chromosome conformation capture (3C) experiments revealed that CTCF directs the formation of chromatin loops within HPV genome, measuring 7.9 kb, and connected CTCF-mediated HPV genome looping to the presence of YY1 transcription factor (Pentland et al., 2018). In this study, CTCF-YY1-mediated chromatin looping regulates the transcriptional repression of early gene expression by PRC2 recruitment and subsequent H3K27me3 marks (Pentland et al., 2018). Loss of YY1 upon cellular differentiation leads to the resolution of HPV chromatin loops and to the following loss of PCR2 and H3K27me3 repressive marks, thereby enabling viral oncoprotein transcription (Pentland et al., 2018). Interestingly, CTCF association with HPV genome has been associated to Smc1 recruitment (Mehta et al., 2015), but implication of Smc1 in HPV genome organization has not been addressed. CTCF-mediated chromatin looping has also been reported for EBV genome organization, notably to contact specific promoter and enhancer. EBV genome is 170 kb long and mutations in CTCF binding sites disrupted loop formation and subsequently altered transcriptional repression of the promoter-regulated viral gene (Tempera et al., 2011). Collectively, these studies bring evidences that episomal genomes do harbor DNA loops and that CTCF insulator functions described for cellular genes also take place in viral genomes. As HPV and EBV harbor longer genomes than average length of human chromatin loop it could be easily expected that such genomes can be spatially structured. However, it is important to note that HSV-1 also forms chromatin loops within an 8 kb-long genome, unravelling that cellular processes can be adapted to viral context.

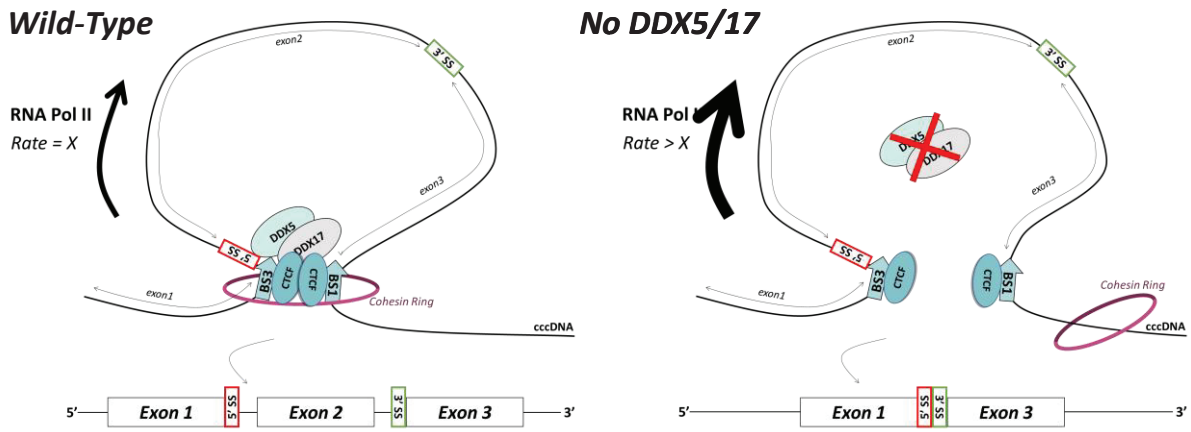


Figure 34: Working model of DDX5/17-mediated mechanism in HBV RNA splicing

Wild-type: CTCF recruitment to BS1 and BS3 (blue convergent arrows) , together with cohesin loading *via* DDX5/17 induces a certain rate (=X) of elongating RNA Pol II. Subsequently, the splicing donor site (5'SS in red square) and acceptor site (3'SS in green square) are not used as splicing sites and neosynthesized HBV RNA contain exon1, exon2 and exon3.
In DDX5/17 silenced condition, cohesin ring is re-located elsewhere along cccDNA molecule, leading to the resolution of the chromatin loop and subsequent increased elongation rate (>X) of RNA Pol II. Enhancement of elongation rate leads to the skipping of exon2. In this condition, neosynthesized transcripts contain exon1 and exon3.

It remains unclear whether DDX5 and DDX17 are part of CTCF complex on HBV cccDNA or whether they act independently, and whether CTCF and DDX5/17 display insulator functions on HBV genome. Achieving efficient cccDNA digestion would answer whether CTCF/DDX5/DDX17 exist as a complex on cccDNA. Moreover, specifically recruiting CTCF or DDX5/DDX17 by rescue experiments using CRISPR based technology will help in determining whether forcing the recruitment of DDX5 or DDX7 to the mutated CTCF BS can compensate for the lost binding of CTCF. Finally, adapting C-technologies, such as 3C or 4C technics, to small episomes would bring the last layer of information required to answer HBV genome conformational questions.

Model of HBV RNA splicing modulated by DDX5/17

During my thesis internship, I demonstrated in a first project that DDX5 and DDX17 are involved in HBV pgRNA splicing regulation. The role of DDX5/17 may be mediated by their direct interaction with HBV RNA and/or cccDNA. DDX5, and possibly DDX17, are required for CTCF insulator activity, together with cohesin (H. Yao et al., 2010). We also showed in a second project that CTCF, SMC3 and RAD21 are associated to cccDNA, and that CTCF, SMC3, RAD21, DDX5 and DDX17 are concomitantly associated to a same cccDNA molecule. Moreover, CTCF is involved in alternative splicing regulation by modulating RNA Pol II pausing and subsequent exon inclusion (Shukla et al., 2011). HBV genome contains eight potential CTCF binding sites. Among them, BS1 is convergent to the seven others. These observations are opening the possibility of chromatin loop formation along cccDNA and potential consequences in splicing regulation. We therefore propose a functional model of HBV RNA splicing in presence or in absence of DDX5 and DDX17 in **Figure 34**.

First, in wild-type condition, CTCF binding sites BS1 and BS3 recruit CTCF in a convergent orientation. In parallel, DDX5/17 stabilize the cohesin subunits onto cccDNA. Convergence of CTCF allows cohesin ring to generate a chromatin loop between BS1 and BS3, according to the well-described loop extrusion model (Barrington et al., 2017). Interestingly, BS3 (2296-2315) is located around the major HBV donor sites of splicing (positions 2067 and 2087) and the acceptor sites located at positions 2337 and 2350. Chromatin loop between BS1 and BS3 trigger a certain processivity to RNA Pol II elongation. We called this elongation rate “X”. During transcription in this 3D-context, RNA Pol II elongating at a rate = X transcribes transcripts containing exon1, exon2 and exon3.

While DDX5 and DDX17 depletion does not impair CTCF recruitment to CTCF binding sites, it may re-localize cohesin ring to other regions on the chromatin (H. Yao et al., 2010). Resolution of the chromatin loop between BS1 and BS3 in a context of DDX5/17 knockdown changes chromatin conformation around splicing donor and acceptor sites. We suggest in this context that RNA Pol II processivity increases to an elongation rate higher than "X". Increased elongation rate therefore does not allow RNA Pol II to transcribe exon 2 and neosynthesized transcripts are constituted of exon 1 and exon 3.

Why do viral particles accumulate in the cytoplasm of DDX5/17 depleted cells?

As shortening of viral RNAs upon DDX5/17 depletion is due to the major use of HBV PAS and the shortening of 3' UTR, and knowing that shorter 3' UTR is associated to increased RNA stability and protein translation, we hypothesized that HBV RNAs should be more rapidly processed in silenced condition. To assess the effect of DDX5 and DDX17 knockdown in the viral replication cycle, we isolated cytoplasmic particles. We observed a global increase in the quantity of these particles, together with an increase of encapsidated viral RNA and DNA, when the cells were suppressed for DDX5/17 expression. The increase of encapsidated DNA concomitantly with the increase of encapsidated HBV RNA witnesses a correct retro-transcription of pgRNA in both of control and silenced conditions. We biochemically investigated the nucleic acids contained in accumulated viral particles to investigate a potential impact of DDX5/17 depletion in pgRNA retro-transcription. We observed an increase in HBV total RNA and DNA in HBc-containing particles within the cytoplasm of DDX5/17 depleted cells. To fully complete the study of HBV replication upon DDX5/17 depletion, we characterized encapsidated HBV DNA in cytoplasmic and secreted capsid-containing particles by Nanopore sequencing. We performed Nanopore sequencing after complete genome reverse PCR amplification that allows the amplification of fully circularized DNA. Such reverse PCR was already performed to amplify complete HBV genome in several studies (Günther et al., 1995; Sauvage et al., 2018), using primers located the nucleotide upstream of precore ATG (position 1813). However, we decided to design primers located upstream of the already-described primers, diverging at the position 1628, approximately 200 bp upstream of the commonly used primers. We designed this new primer couple in a position that does not prevent amplification of potential HBV DNA deriving from truncated or differentially ended transcripts. In cytoplasmic capsid-containing particles, we observed a similar coverage in

siDDX5/17 with respect to siCTL condition, meaning that DDX5/17 silencing does not affect encapsidated HBV DNA. We confirmed these results in secreted particles as no changes was observed in viral DNA in both control and depleted conditions.

Morphological characterization of cells in siCTL and siDDX5/17 conditions was assessed by TEM. Ultrastructure imaging confirmed the increase in cytoplasmic particles, as we were able to distinctly observe more cytoplasmic vacuoles with a greater cytoplasmic surface occupation upon DDX5/17 repression. Noteworthy, TEM technique is not suitable for precise quantification. Accurate quantification of vacuoles should be done with electron tomograph.

Several publications used TEM to investigate cellular morphology upon HBV infection. Some studies showed cytoplasmic vesicles and assumed viral particles to be contained in such vesicles. However, in some studies, the spherical structures present in cytoplasmic vacuoles had indeed a diameter larger than expected for HBV structures: described to be 42-nm diameter for Dane particles, 27-nm diameter for naked core particles and 22-nm diameter for spheres and filaments (Gerber et al., 1974; Hu and Liu, 2017; Huang et al., 2012; Meuleman et al., 2006). Contrariwise, TEM data from different publications show similar features and suggest that such vesicles are indeed induced by HBV infection (Jiang et al., 2016; Roingeard et al., 1990; Xu et al., 1997). Validation of the presence of viral proteins in so-far described vesicles has been achieved by immunogold staining that consists in immunostaining with anti-HBs antibodies and hybridization with a secondary antibody coupled to nanogold beads, which are dense to electrons and appear clearly in TEM. Using this technique, different studies confirmed the presence of viral particles in such cytoplasmic compartments, as nanogold beads coupled to secondary antibodies against anti-HBc or anti-HBs proteins were visible in vesicles structures only in stained conditions (Patient et al., 2007; Rat et al., 2019). Moreover, such viral particle-containing vesicles were also found in duck primary hepatocytes upon DHBV infection (Mhamdi et al., 2007).

However, most studies conducted about the morphology of cytoplasmic vesicles and contained viral particles were performed upon plasmid transfection, encoding either surface proteins or complete HBV genome or in cell lines not known to be host for HBV, such as BHK (Patient et al., 2007). To date, no study uncovered the morphology associated to HBV infection, and the presence of HBc- and HBs-containing viral particles in described cytoplasmic compartments. For the first time, we provided evidences that HBV infection

in a relevant cell line as HepG2 NTCP, induces the presence of large membrane-surrounded vacuoles that contain HBc and HBs viral proteins as it was previously reported in HBV DNA transfection experiments. Our data suggests that DDX5/17 depletion leads to accumulation of viral and subviral particles within such vacuoles, and these vacuoles have an increased size compared to non-depleted controls. More biochemical studies would be necessary to determine the nature and properties of such compartments. Previous studies suggested that HBV particles would be contained in membrane-surrounded vesicles consistent with multivesicular bodies and autophagosomes (Inoue et al., 2015). Further studies using immunogold multi-staining to investigate the nature of the vesicles we observed would be very informational. Moreover, silencing structural proteins of these compartments would provide more information about precise pathway of HBV particle secretion.

We also observed that accumulated viral particle-containing vesicles within the cytoplasm of infected cells upon DDX5/17 depletion seem not to be efficiently secreted. Rab7 was demonstrated to be involved in the negative regulation of HBV particle secretion (Inoue et al., 2015). Thus, we could hypothesize that DDX5/17 silencing could upregulate Rab7 expression. To answer this question, we looked for RNA-seq data and observed no modulation of Rab7 expression upon DDX5/17 depletion. However, depletion of DDX5/17 triggers a more than 6-fold upregulation of Rab27b, a protein demonstrated to limit HBV capsid secretion as its knockdown by siRNA leads to an increase of secreted HBV particles (Döring and Prange, 2015). HBV egress is also dependent on the human endosomal sorting complex required for transport (ESCRT)-III complex and notably to its subunits γ2-adaptin and Nedd4 (Lambert et al., 2007), which are not affected by DDX5/17 silencing. Nevertheless, DDX5 and DDX17 knockdown modulate the expression of other ESCRT-III components, such as CHMP4A (2,5-fold upregulated). The ESCRT-I subunits VSP28 and VSP37D are 60% and 80% downregulated, respectively, upon DDX5/17 depletion. Strikingly both of these factors are required for HIV-1 budding (Stuchell et al., 2004; Weiss and Göttlinger, 2011), and we could therefore suggest that they share this role with HBV budding. Further investigating the functions of these factors would probably help explaining such retention of non-secreted HBV particles. Besides, functional analysis of

these putative candidates could highlight HBV secretory pathways poorly understood so far.

Strikingly, our RNAseq data also revealed a 4-fold upregulation of HSP90, demonstrated to participate to the viral polymerase switch to its active confirmation, required for pgRNA retro-transcription initiation, by supplying the viral enzyme with ATP (Hu et al., 2004b). This observation is in line with the increase in HBV RNA and DNA within capsid shells upon DDX5/17 depleted conditions.

Finally, we can interrogate about the lack of increased secretion of viral particles that would be expected after particle accumulation in the cytoplasm of depleted cells. It is important to note that silencing experiments cannot be conducted for more than four days after the beginning of siRNA treatments, due to a rebound of DDX5 and DDX17 expression afterwards. Besides potential impairment of secretory pathways upon DDX5/17 silencing, we could imagine that the accumulation could be ascribed to a delay in secretion kinetic that would be different after longer silencing. Moreover, one could also imagine that a lack of increase in factors involved in secretory pathway would not allow to properly secrete accumulated cytoplasmic particles due to a saturated system. Pursuing investigations on accumulated particles would be permitted by longer DDX5 and DDX17 silencing that could be obtained by transient shRNA transfection or inducible CRISPR/Cas9 systems. Noteworthy, completely knocking down DDX5 and DDX17 is lethal for cells.

In conclusion, my PhD project results provide new insights in the co-transcriptional regulation of HBV RNA and uncover striking information about HBV RNA 3' ending that are crucial for future investigations on HBV RNA transcription and processing. Understanding the cccDNA biology, notably its transcriptional regulation, and developing non-invasive biomarkers reflecting cccDNA amount and transcriptional activity in infected patients are essential keys toward the development of relevant biomarkers and curative treatments.

References

References

Ajiro, M., Zheng, Z.-M., 2014. Oncogenes and RNA splicing of human tumor viruses. Emerging Microbes & Infections 3, 1–16. https://doi.org/10.1038/emi.2014.62

Almassalha, L.M., Tiwari, A., Ruhoff, P.T., Stypula-Cyrus, Y., Cherkezyan, L., Matsuda, H., Dela Cruz, M.A., Chandler, J.E., White, C., Maneval, C., Subramanian, H., Szleifer, I., Roy, H.K., Backman, V., 2017. The Global Relationship between Chromatin Physical Topology, Fractal Structure, and Gene Expression. Scientific Reports 7. https://doi.org/10.1038/srep41061

Altinel, K., Hashimoto, K., Wei, Y., Neuveut, C., Gupta, I., Suzuki, A.M., Dos Santos, A., Moreau, P., Xia, T., Kojima, S., Kato, S., Takikawa, Y., Hidaka, I., Shimizu, M., Matsuura, T., Tsubota, A., Ikeda, H., Nagoshi, S., Suzuki, H., Michel, M.-L., Samuel, D., Buendia, M.A., Faivre, J., Carninci, P., 2016. Single-Nucleotide Resolution Mapping of Hepatitis B Virus Promoters in Infected Human Livers and Hepatocellular Carcinoma. Journal of Virology 90, 10811–10822. https://doi.org/10.1128/JVI.01625-16

Anders, S., Pyl, P.T., Huber, W., 2015. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. https://doi.org/10.1093/bioinformatics/btu638

Bai, L., Zhang, X., Kozlowski, M., Li, W., Wu, M., Liu, J., Chen, L., Zhang, J., Huang, Y., Yuan, Z., 2018. Extracellular Hepatitis B Virus RNAs Are Heterogeneous in Length and Circulate as Capsid-Antibody Complexes in Addition to Virions in Chronic Hepatitis B Patients. J Virol 92, e00798-18, /jvi/92/24/e00798-18.atom. https://doi.org/10.1128/JVI.00798-18

Bandopadhyay, M., Banerjee, A., Sarkar, N., Panigrahi, R., Datta, S., Pal, A., Singh, S.P., Biswas, A., Chakrabarti, S., Chakravarty, R., 2014. Tumor suppressor micro RNA miR-145 and onco micro RNAs miR-21 and miR-222 expressions are differentially modulated by Hepatitis B virus X protein in malignant hepatocytes. BMC Cancer 14, 721. https://doi.org/10.1186/1471-2407-14-721

Bao, L., Zhang, M., Han, S., Zhan, Y., Guo, W., Teng, F., Liu, F., Guo, M., Zhang, L., Ding, G., Wang, Q., 2018. MicroRNA-500a Promotes the Progression of Hepatocellular Carcinoma by Post-Transcriptionally Targeting BID. Cell Physiol Biochem 47, 2046–2055. https://doi.org/10.1159/000491472

Baralle, M., Baralle, F.E., 2018. The splicing code. Biosystems 164, 39–48. https://doi.org/10.1016/j.biosystems.2017.11.002

Barrington, C., Finn, R., Hadjur, S., 2017. Cohesin biology meets the loop extrusion model. Chromosome Res 25, 51–60. https://doi.org/10.1007/s10577-017-9550-3

Bayliss, J., Lim, L., Thompson, A.J.V., Desmond, P., Angus, P., Locarnini, S., Revill, P.A., 2013a. Hepatitis B virus splicing is enhanced prior to development of hepatocellular carcinoma. Journal of Hepatology 59, 1022–1028. https://doi.org/10.1016/j.jhep.2013.06.018

Beck, J., Nassal, M., 2007. Hepatitis B virus replication. World J Gastroenterol 13, 48–64. https://doi.org/10.3748/wjg.v13.i1.48

Belloni, L., Allweiss, L., Guerrieri, F., Pediconi, N., Volz, T., Pollicino, T., Petersen, J., Raimondo, G., Dandri, M., Levrero, M., 2012. IFN-α inhibits HBV transcription and replication in

cell culture and in humanized mice by targeting the epigenetic regulation of the nuclear cccDNA minichromosome. Journal of Clinical Investigation 122, 529–537. https://doi.org/10.1172/JCI58847

Belloni, L., Pollicino, T., De Nicola, F., Guerrieri, F., Raffa, G., Fanciulli, M., Raimondo, G., Levrero, M., 2009a. Nuclear HBx binds the HBV minichromosome and modifies the epigenetic regulation of cccDNA function. Proceedings of the National Academy of Sciences 106, 19975–19979. https://doi.org/10.1073/pnas.0908365106

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

Benoit-Pilven, C., Marchet, C., Chautard, E., Lima, L., Lambert, M.-P., Sacomoto, G., Rey, A., Cologne, A., Terrone, S., Dulaurier, L., Claude, J.-B., Bourgeois, C.F., Auboeuf, D., Lacroix, V., 2018. Complementarity of assembly-first and mapping-first approaches for alternative splicing annotation and differential analysis from RNAseq data. Sci Rep 8, 4307. https://doi.org/10.1038/s41598-018-21770-7

Bhanja Chowdhury, J., Roy, D., Ghosh, S., 2011. Identification of a unique splicing regulatory cluster in hepatitis B virus pregenomic RNA. FEBS Letters 585, 3348–3353. https://doi.org/10.1016/j.febslet.2011.09.026

Bock, C.-T., Malek, N.P., Tillmann, H.L., Manns, M.P., Trautwein, C., 2000. The Enhancer I Core Region Contributes to the Replication Level of Hepatitis B Virus In Vivo and In Vitro. Journal of Virology 74, 2193–2202. https://doi.org/10.1128/JVI.74.5.2193-2202.2000

Bock, C.-T., Schranz, P., Schröder, C.H., Zentgraf, H., 1994. Hepatitis B virus genome is organized into nucleosomes in the nucleus of the infected cell. Virus Genes 8, 215–229. https://doi.org/10.1007/BF01703079

Bock, C. Thomas, Schwinn, S., Locarnini, S., Fyfe, J., Manns, M.P., Trautwein, C., Zentgraf, H., 2001. Structural organization of the hepatitis B virus minichromosome1. Journal of Molecular Biology 307, 183–196. https://doi.org/10.1006/jmbi.2000.4481

Bonilla Guerrero, R., Roberts, L.R., 2005. The role of hepatitis B virus integrations in the pathogenesis of human hepatocellular carcinoma. Journal of Hepatology 42, 760–777. https://doi.org/10.1016/j.jhep.2005.02.005

Borgia, G., Gentile, I., 2014. Vertical transmission of hepatitis B virus: challenges and solutions. International Journal of Women's Health 605. https://doi.org/10.2147/IJWH.S51138

Böttcher, B., Vogel, M., Ploss, M., Nassal, M., 2006. High Plasticity of the Hepatitis B Virus Capsid Revealed by Conformational Stress. Journal of Molecular Biology 356, 812–822. https://doi.org/10.1016/j.jmb.2005.11.053

Bourgeois, C.F., Mortreux, F., Auboeuf, D., 2016a. The multiple functions of RNA helicases as drivers and regulators of gene expression. Nature Reviews Molecular Cell Biology 17, 426–438. https://doi.org/10.1038/nrm.2016.50

Boyd, A., Lacombe, K., Lavocat, F., Maylin, S., Miailhes, P., Lascoux-Combe, C., Delaugerre, C., Girard, P.-M., Zoulim, F., 2016. Decay of ccc-DNA marks persistence of intrahepatic viral DNA synthesis under tenofovir in HIV-HBV co-infected patients. Journal of Hepatology 65, 683–691. https://doi.org/10.1016/j.jhep.2016.05.014

Braccioli, L., de Wit, E., 2019. CTCF: a Swiss-army knife for genome organization and transcription regulation. Essays Biochem. 63, 157–165. https://doi.org/10.1042/EBC20180069

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34, 525–527. https://doi.org/10.1038/nbt.3519

Brumbaugh, J., Di Stefano, B., Wang, X., Borkent, M., Forouzmand, E., Clowers, K.J., Ji, F., Schwarz, B.A., Kalocsay, M., Elledge, S.J., Chen, Y., Sadreyev, R.I., Gygi, S.P., Hu, G., Shi, Y., Hochedlinger, K., 2018. Nudt21 Controls Cell Fate by Connecting Alternative Polyadenylation to Chromatin Signaling. Cell 172, 106-120.e21. https://doi.org/10.1016/j.cell.2017.11.023

Bruss, V., 2004. Envelopment of the hepatitis B virus nucleocapsid. Virus Research 106, 199–209. https://doi.org/10.1016/j.virusres.2004.08.016

Bruss, V., Ganem, D., 1991. The role of envelope proteins in hepatitis B virus assembly. Proceedings of the National Academy of Sciences 88, 1059–1063. https://doi.org/10.1073/pnas.88.3.1059

Burwitz, B.J., Wettengel, J.M., Mück-Häusl, M.A., Ringelhan, M., Ko, C., Festag, M.M., Hammond, K.B., Northrup, M., Bimber, B.N., Jacob, T., Reed, J.S., Norris, R., Park, B., Moller-Tank, S., Esser, K., Greene, J.M., Wu, H.L., Abdulhaqq, S., Webb, G., Sutton, W.F., Klug, A., Swanson, T., Legasse, A.W., Vu, T.Q., Asokan, A., Haigwood, N.L., Protzer, U., Sacha, J.B., 2017. Hepatocytic expression of human sodium-taurocholate cotransporting polypeptide enables hepatitis B virus infection of macaques. Nat Commun 8, 2146. https://doi.org/10.1038/s41467-017-01953-y

Candotti, D., Allain, J.-P., 2016a. Biological and clinical significance of hepatitis B virus RNA splicing: an update. Annals of Blood 2, 6–6. https://doi.org/10.21037/aob.2017.05.01

Cavalli, G., Misteli, T., 2013a. Functional implications of genome topology. Nature Structural & Molecular Biology 20, 290–299. https://doi.org/10.1038/nsmb.2474

Chang, M.S., Nguyen, M.H., 2017. Epidemiology of hepatitis B and the role of vaccination. Best Practice & Research Clinical Gastroenterology 31, 239–247. https://doi.org/10.1016/j.bpg.2017.05.008

Charre, C., Levrero, M., Zoulim, F., Scholtès, C., 2019. Non-invasive biomarkers for chronic hepatitis B virus infection management. Antiviral Research 169, 104553. https://doi.org/10.1016/j.antiviral.2019.104553

Chen, J., Wu, M., Wang, F., Zhang, W., Wang, W., Zhang, X., Zhang, J., Liu, Yinghui, Liu, Yi, Feng, Y., Zheng, Y., Hu, Y., Yuan, Z., 2015. Hepatitis B virus spliced variants are associated with an impaired response to interferon therapy. Sci Rep 5, 16459. https://doi.org/10.1038/srep16459

Chen, P., Gan, Y., Han, N., Fang, W., Li, J., Zhao, F., Hu, K., Rayner, S., 2013. Computational Evolutionary Analysis of the Overlapped Surface (S) and Polymerase (P) Region in Hepatitis B Virus Indicates the Spacer Domain in P Is Crucial for Survival. PLoS ONE 8, e60098. https://doi.org/10.1371/journal.pone.0060098

Chen, W.-N., Chen, J.-Y., Jiao, B.-Y., Lin, W.-S., Wu, Y.-L., Liu, L.-L., Lin, X., 2012. Interaction of the Hepatitis B Spliced Protein with Cathepsin B Promotes Hepatoma Cell Migration and Invasion. Journal of Virology 86, 13533–13541. https://doi.org/10.1128/JVI.02095-12

Chen, W.-N., Chen, J.-Y., Lin, W.-S., Lin, J.-Y., Lin, X., 2010. Hepatitis B doubly spliced protein, generated by a 2.2 kb doubly spliced hepatitis B virus RNA, is a pleiotropic activator protein mediating its effects via activator protein-1- and CCAAT/enhancer-binding protein-binding sites. Journal of General Virology 91, 2592–2600. https://doi.org/10.1099/vir.0.022517-0

Chernukhin, I., Shamsuddin, S., Kang, S.Y., Bergström, R., Kwon, Y.-W., Yu, W., Whitehead, J., Mukhopadhyay, R., Docquier, F., Farrar, D., Morrison, I., Vigneron, M., Wu, S.-Y., Chiang, C.-M., Loukinov, D., Lobanenkov, V., Ohlsson, R., Klenova, E., 2007. CTCF

Interacts with and Recruits the Largest Subunit of RNA Polymerase II to CTCF Target Sites Genome-Wide. Mol Cell Biol 27, 1631–1648. https://doi.org/10.1128/MCB.01993-06

Chi, B., Wang, K., Du, Y., Gui, B., Chang, X., Wang, L., Fan, J., Chen, S., Wu, X., Li, G., Cheng, H., 2014. A Sub-Element in PRE enhances nuclear export of intronless mRNAs by recruiting the TREX complex via ZC3H18. Nucleic Acids Research 42, 7305–7318. https://doi.org/10.1093/nar/gku350

Cho, E., Kim, H., Park, C., So, H., Park, R.K., Kim, H.C., 2013. Impact of Nucleotide Mutations at the HNF3- and HNF4-Binding Sites in Enhancer 1 on Viral Replication in Patients with Chronic Hepatitis B Virus Infection. Gut and Liver 7, 569–575. https://doi.org/10.5009/gnl.2013.7.5.569

Cougot, D., Allemand, E., Riviere, L., Benhenda, S., Duroure, K., Levillayer, F., Muchardt, C., Buendia, M.-A., Neuveut, C., 2012. Inhibition of PP1 Phosphatase Activity by HBx: A Mechanism for the Activation of Hepatitis B Virus Transcription. Science Signaling 5, ra1–ra1. https://doi.org/10.1126/scisignal.2001906

Covelo-Molares, H., Bartosovic, M., Vanacova, S., 2018. RNA methylation in nuclear pre-mRNA processing. Wiley Interdisciplinary Reviews: RNA 9, e1489. https://doi.org/10.1002/wrna.1489

Curinha, A., Oliveira Braz, S., Pereira-Castro, I., Cruz, A., Moreira, A., 2014a. Implications of polyadenylation in health and disease. Nucleus 5, 508–519. https://doi.org/10.4161/nucl.36360

Danan, C., Manickavel, S., Hafner, M., 2016. PAR-CLIP: A Method for Transcriptome-Wide Identification of RNA Binding Protein Interaction Sites, in: Dassi, E. (Ed.), Post-Transcriptional Gene Regulation. Springer New York, New York, NY, pp. 153–173. https://doi.org/10.1007/978-1-4939-3067-8_10

Dandri, M., 2020. Epigenetic modulation in chronic hepatitis B virus infection. Semin Immunopathol. https://doi.org/10.1007/s00281-020-00780-6

Dandri, M., Locarnini, S., 2012. New insight in the pathobiology of hepatitis B virus infection. Gut 61, i6–i17. https://doi.org/10.1136/gutjnl-2012-302056

Dardenne, E., Pierredon, S., Driouch, K., Gratadou, L., Lacroix-Triki, M., Espinoza, M.P., Zonta, E., Germann, S., Mortada, H., Villemin, J.-P., Dutertre, M., Lidereau, R., Vagner, S., Auboeuf, D., 2012. Splicing switch of an epigenetic regulator by RNA helicases promotes tumor-cell invasiveness. Nature Structural & Molecular Biology 19, 1139–1146. https://doi.org/10.1038/nsmb.2390

Dardenne, E., Polay Espinoza, M., Fattet, L., Germann, S., Lambert, M.-P., Neil, H., Zonta, E., Mortada, H., Gratadou, L., Deygas, M., Chakrama, F.Z., Samaan, S., Desmet, F.-O., Tranchevent, L.-C., Dutertre, M., Rimokh, R., Bourgeois, C.F., Auboeuf, D., 2014a. RNA Helicases DDX5 and DDX17 Dynamically Orchestrate Transcription, miRNA, and Splicing Programs in Cell Differentiation. Cell Reports 7, 1900–1913. https://doi.org/10.1016/j.celrep.2014.05.010

De Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., de Laat, W., 2015. CTCF Binding Polarity Determines Chromatin Looping. Mol. Cell 60, 676–684. https://doi.org/10.1016/j.molcel.2015.09.023

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

de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., de Laat, W., 2015. CTCF Binding Polarity Determines Chromatin Looping. Molecular Cell 60, 676–684. https://doi.org/10.1016/j.molcel.2015.09.023

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635

Döring, T., Prange, R., 2015. Rab33B and its autophagic Atg5/12/16L1 effector assist in hepatitis B virus naked capsid formation and release: HBV capsid biogenesis depends on Rab33B and Atg5/12/16L1. Cell Microbiol 17, 747–764. https://doi.org/10.1111/cmi.12398

Dupinay, T., Gheit, T., Roques, P., Cova, L., Chevallier-Queyron, P., Tasahsu, S., Le Grand, R., Simon, F., Cordier, G., Wakrim, L., Benjelloun, S., Trépo, C., Chemin, I., 2013. Discovery of naturally occurring transmissible chronic hepatitis B virus infection among *Macaca fascicularis* from mauritius island: Hepatology. Hepatology 58, 1610–1620. https://doi.org/10.1002/hep.26428

Duriez, M., Mandouri, Y., Lekbaby, B., Wang, H., Schnuriger, A., Redelsperger, F., Guerrera, 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., 2017a. Alternative splicing of hepatitis B virus: A novel virus/host interaction altering liver immunity. Journal of Hepatology 67, 687–699. https://doi.org/10.1016/j.jhep.2017.05.025

Dvinge, H., 2018. Regulation of alternative mRNA splicing: old players and new perspectives. FEBS Letters 592, 2987–3006. https://doi.org/10.1002/1873-3468.13119

Eaton, J.D., Francis, L., Davidson, L., West, S., 2020. A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. Genes & Development 34, 132–145. https://doi.org/10.1101/gad.332833.119

Elmore, L.W., Hancock, A.R., Chang, S.-F., Wang, X.W., Chang, S., Callahan, C.P., Geller, D.A., Will, H., Harris, C.C., 1997. Hepatitis B virus X protein and p53 tumor suppressor interactions in the modulation of apoptosis. Proceedings of the National Academy of Sciences 94, 14707–14712. https://doi.org/10.1073/pnas.94.26.14707

Fuda, N.J., Ardehali, M.B., Lis, J.T., 2009. Defining mechanisms that regulate RNA polymerase II transcription in vivo. Nature 461, 186–192. https://doi.org/10.1038/nature08449

Fuller-Pace, F.V., 2006. DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. Nucleic Acids Research 34, 4206–4215. https://doi.org/10.1093/nar/gkl460

Fuller-Pace, F.V., Moore, H.C., 2011. RNA helicases p68 and p72: multifunctional proteins with important implications for cancer development. Future Oncology 7, 239–251. https://doi.org/10.2217/fon.11.1

Gallucci, L., Kann, M., 2017. Nuclear Import of Hepatitis B Virus Capsids and Genome. Viruses 9, 21. https://doi.org/10.3390/v9010021

Ganem, D., 2004. Hepatitis B Virus Infection — Natural History and Clinical Consequences. The New England Journal of Medicine 12.

Gao, W., Hu, J., 2007. Formation of Hepatitis B Virus Covalently Closed Circular DNA: Removal of Genome-Linked Protein. Journal of Virology 81, 6164–6174. https://doi.org/10.1128/JVI.02721-06

Geißler, V., Altmeyer, S., Stein, B., Uhlmann-Schiffler, H., Stahl, H., 2013. The RNA helicase Ddx5/p68 binds to hUpf3 and enhances NMD of Ddx17/p72 and Smg5 mRNA. Nucleic Acids Research 41, 7875–7888. https://doi.org/10.1093/nar/gkt538

Gerber, M.A., Hadziyannis, S., Vissoulis, C., Schaffner, F., Paronetto, F., Popper, H., 1974. Electron Microscopy and Immunoelectronmicroscopy of Cytoplasmic Hepatitis B Antigen in Hepatocytes 75, 14.

Gerelsaikhan, T., Tavis, J.E., Bruss, V., 1996. Hepatitis B Virus Nucleocapsid Envelopment Does Not Occur without Genomic DNA Synthesis. J. VIROL. 70, 6.

Ghosh, A., Lima, C.D., 2010. Enzymology of RNA cap synthesis. Wiley Interdisciplinary Reviews: RNA 1, 152–172. https://doi.org/10.1002/wrna.19

Giraud, G., Terrone, S., Bourgeois, C.F., 2018a. Functions of DEAD box RNA helicases DDX5 and DDX17 in chromatin organization and transcriptional regulation. BMB Reports 51, 613–622. https://doi.org/10.5483/BMBRep.2018.51.12.234

Gómez-Moreno, A., Garaigorta, U., 2017. Hepatitis B Virus and DNA Damage Response: Interactions and Consequences for the Infection. Viruses 9, 304. https://doi.org/10.3390/v9100304

Graham, S.V., Faizo, A.A.A., 2017. Control of human papillomavirus gene expression by alternative splicing. Virus Research 231, 83–95. https://doi.org/10.1016/j.virusres.2016.11.016

Gripon, P., Le Seyec, J., Rumin, S., Guguen-Guillouzo, C., 1995. Myristylation of the Hepatitis B Virus Large Surface Protein Is Essential for Viral Infectivity. Virology 213, 292–299. https://doi.org/10.1006/viro.1995.0002

Guerrieri, F., Belloni, L., D'Andrea, D., Pediconi, N., Le Pera, L., Testoni, B., Scisciani, C., Floriot, O., Zoulim, F., Tramontano, A., Levrero, M., 2017. Genome-wide identification of direct HBx genomic targets. BMC Genomics 18. https://doi.org/10.1186/s12864-017-3561-5

Günther, S., Li, B.C., Miska, S., Krüger, D.H., Meisel, H., Will, H., 1995. A novel method for efficient amplification of whole hepatitis B virus genomes permits rapid functional analysis and reveals deletion mutants in immunosuppressed patients. Journal of virology 69, 5437–5444. https://doi.org/10.1128/JVI.69.9.5437-5444.1995

Guo, H., Xu, C., Zhou, T., Block, T.M., Guo, J.-T., 2012. Characterization of the Host Factors Required for Hepadnavirus Covalently Closed Circular (ccc) DNA Formation. PLoS ONE 7, e43270. https://doi.org/10.1371/journal.pone.0043270

Guo, X., Chen, P., Hou, X., Xu, W., Wang, D., Wang, T., Zhang, L., Zheng, G., Gao, Z., He, C.-Y., Zhou, B., Chen, Z.-Y., 2016. The recombined cccDNA produced using minicircle technology mimicked HBV genome in structure and function closely. Sci Rep 6, 1–10. https://doi.org/10.1038/srep25552

Guo, Y.-H., Li, Y.-N., Zhao, J.-R., Zhang, J., Yan, Z., 2011. HBc binds to the CpG islands of HBV cccDNA and promotes an epigenetic permissive state. Epigenetics 6, 720–726. https://doi.org/10.4161/epi.6.6.15815

Habig, J.W., Loeb, D.D., 2006. Sequence Identity of the Direct Repeats, DR1 and DR2, Contributes to the Discrimination between Primer Translocation and in Situ Priming During Replication of the Duck Hepatitis B Virus. Journal of Molecular Biology 364, 32–43. https://doi.org/10.1016/j.jmb.2006.08.095

Hansen, A.S., Hsieh, T.-H.S., Cattoglio, C., Pustova, I., Darzacq, X., Tjian, R., 2018. An RNA-binding region regulates CTCF clustering and chromatin looping: bioRxiv. https://doi.org/10.1101/495432

Hardy, J.G., Norbury, C.J., 2016a. Cleavage factor Im (CFIm) as a regulator of alternative polyadenylation. Biochemical Society Transactions 44, 1051–1057. https://doi.org/10.1042/BST20160078

Hayer, J., Jadeau, F., Deléage, G., Kay, A., Zoulim, F., Combet, C., 2013. HBVdb: a knowledge database for Hepatitis B Virus. Nucleic Acids Res. 41, D566-570. https://doi.org/10.1093/nar/gks1022

He, F., Chen, E.-Q., Liu, L., Zhou, T.-Y., Liu, C., Cheng, X., Liu, F.-J., Tang, H., 2012. Inhibition of hepatitis B Virus replication by hepatocyte nuclear factor 4-alpha specific short hairpin RNA. Liver International 32, 742–751. https://doi.org/10.1111/j.1478-3231.2011.02748.x

He, Y., Yuan, C., Chen, L., Lei, M., Zellmer, L., Huang, H., Liao, D., 2018. Transcriptional-Readthrough RNAs Reflect the Phenomenon of "A Gene Contains Gene(s)" or "Gene(s) within a Gene" in the Human Genome, and Thus Are Not Chimeric RNAs. Genes 9, 40. https://doi.org/10.3390/genes9010040

Heinz, S., Texari, L., Hayes, M.G.B., Urbanowski, M., Chang, M.W., Givarkes, N., Rialdi, A., White, K.M., Albrecht, R.A., Pache, L., Marazzi, I., García-Sastre, A., Shaw, M.L., Benner, C., 2018. Transcription Elongation Can Affect Genome 3D Structure. Cell 174, 1522-1536.e22. https://doi.org/10.1016/j.cell.2018.07.047

Heise, T., 2006. The hepatitis B virus PRE contains a splicing regulatory element. Nucleic Acids Research 34, 353–363. https://doi.org/10.1093/nar/gkj440

Hensel, K.O., Cantner, F., Bangert, F., Wirth, S., Postberg, J., 2018. Episomal HBV persistence within transcribed host nuclear chromatin compartments involves HBx. Epigenetics & Chromatin 11. https://doi.org/10.1186/s13072-018-0204-2

Hu, J., 2009. RNA-protein interactions in hepadnavirus reverse transcription. Frontiers in Bioscience Volume, 1606. https://doi.org/10.2741/3328

Hu, J., Flores, D., Toft, D., Wang, X., Nguyen, D., 2004a. Requirement of Heat Shock Protein 90 for Human Hepatitis B Virus Reverse Transcriptase Function. Journal of Virology 78, 13122–13131. https://doi.org/10.1128/JVI.78.23.13122-13131.2004

Hu, J., Lin, Y.-Y., Chen, P.-J., Watashi, K., Wakita, T., 2019. Cell and Animal Models for Studying Hepatitis B Virus Infection and Drug Development. Gastroenterology 156, 338–354. https://doi.org/10.1053/j.gastro.2018.06.093

Hu, J., Liu, K., 2017a. Complete and Incomplete Hepatitis B Virus Particles: Formation, Function, and Application. Viruses 9, 56. https://doi.org/10.3390/v9030056

Hu, X.-M., Yan, X.-H., Hu, Y.-W., Huang, J.-L., Cao, S.-W., Ren, T.-Y., Tang, Y.-T., Lin, L., Zheng, L., Wang, Q., 2016. miRNA-548p suppresses hepatitis B virus X protein associated hepatocellular carcinoma by downregulating oncoprotein hepatitis B x-interacting protein: miR-548p suppresses HBx-HCC by inhibiting HBXIP. Hepatol Res 46, 804–815. https://doi.org/10.1111/hepr.12618

Huang, H.-C., Chen, C.-C., Chang, W.-C., Tao, M.-H., Huang, C., 2012a. Entry of Hepatitis B Virus into Immortalized Human Primary Hepatocytes by Clathrin-Dependent Endocytosis. Journal of Virology 86, 9443–9453. https://doi.org/10.1128/JVI.00873-12

Huang, H.-L., Jeng, K.-S., Hu, C.-P., Tsai, C.-H., Lo, S.J., Chang, C., 2000. Identification and Characterization of a Structural Protein of Hepatitis B Virus: A Polymerase and Surface

Fusion Protein Encoded by a Spliced RNA. Virology 275, 398–410. https://doi.org/10.1006/viro.2000.0478

Hwang, J.-R., Park, S.-G., 2018. Mouse models for hepatitis B virus research. Laboratory Animal Research 34, 85. https://doi.org/10.5625/lar.2018.34.3.85

Hyle, J., Zhang, Y., Wright, S., Xu, B., Shao, Y., Easton, J., Tian, L., Feng, R., Xu, P., Li, C., 2019. Acute depletion of CTCF directly affects MYC regulation through loss of enhancer-promoter looping. Nucleic Acids Res. 47, 6699–6713. https://doi.org/10.1093/nar/gkz462

Imam, H., Khan, M., Gokhale, N.S., McIntyre, A.B.R., Kim, G.-W., Jang, J.Y., Kim, S.-J., Mason, C.E., Horner, S.M., Siddiqui, A., 2018. *N6* -methyladenosine modification of hepatitis B virus RNA differentially regulates the viral life cycle. Proceedings of the National Academy of Sciences 115, 8829–8834. https://doi.org/10.1073/pnas.1808319115

Inoue, J., Krueger, E.W., Chen, J., Cao, H., Ninomiya, M., McNiven, M.A., 2015a. HBV secretion is regulated through the activation of endocytic and autophagic compartments mediated by Rab7 stimulation. Journal of Cell Science 128, 1696–1706. https://doi.org/10.1242/jcs.158097

Jain, S., Chang, T.-T., Chen, S., Boldbaatar, B., Clemens, A., Lin, S.Y., Yan, R., Hu, C.-T., Guo, H., Block, T.M., Song, W., Su, Y.-H., 2015. Comprehensive DNA methylation analysis of hepatitis B virus genome in infected liver tissues. Scientific Reports 5. https://doi.org/10.1038/srep10478

Jeong, J.-K., Yoon, G.-S., Ryu, W.-S., 2000. Evidence that the 5J-End Cap Structure Is Essential for Encapsidation of Hepatitis B Virus Pregenomic RNA. J. VIROL. 74, 7.

Jiang, B., Himmelsbach, K., Ren, H., Boller, K., Hildt, E., 2016. Subviral Hepatitis B Virus Filaments, like Infectious Viral Particles, Are Released via Multivesicular Bodies. J. Virol. 90, 3330–3341. https://doi.org/10.1128/JVI.03109-15

Jonkers, I., Lis, J.T., 2015. Getting up to speed with transcription elongation by RNA polymerase II. Nature Reviews Molecular Cell Biology 16, 167–177. https://doi.org/10.1038/nrm3953

Julithe, R., Abou-Jaoude, G., Sureau, C., 2014. Modification of the Hepatitis B Virus Envelope Protein Glycosylation Pattern Interferes with Secretion of Viral Particles, Infectivity, and Susceptibility to Neutralizing Antibodies. Journal of Virology 88, 9049–9059. https://doi.org/10.1128/JVI.01161-14

Juven-Gershon, T., Hsu, J.-Y., Theisen, J.W., Kadonaga, J.T., 2008. The RNA polymerase II core promoter — the gateway to transcription. Current Opinion in Cell Biology 20, 253–259. https://doi.org/10.1016/j.ceb.2008.03.003

Kairat, A., Beerheide, W., Zhou, G., Tang, Z.-Y., Edler, L., Schröder, C.H., 1999. Truncated Hepatitis B Virus RNA in Human Hepatocellular Carcinoma: Its Representation in Patients with Advancing Age. Intervirology 42, 228–237. https://doi.org/10.1159/000024982

Kang, H., Wiedmer, A., Yuan, Y., Robertson, E., Lieberman, P.M., 2011. Coordination of KSHV latent and lytic gene control by CTCF-cohesin mediated chromosome conformation. PLoS Pathog. 7, e1002140. https://doi.org/10.1371/journal.ppat.1002140

Kann, M., Gerlich, W.H., 1994. Effect of core protein phosphorylation by protein kinase C on encapsidation of RNA within core particles of hepatitis B virus. Journal of Virology 68, 7993–8000. https://doi.org/10.1128/JVI.68.12.7993-8000.1994

Katahira, J., Ishikawa, H., Tsujimura, K., Kurono, S., Hieda, M., 2019a. Human THO coordinates transcription termination and subsequent transcript release from the *HSP70* locus. Genes to Cells 24, 272–283. https://doi.org/10.1111/gtc.12672

Keasler, V.V., Hodgson, A.J., Madden, C.R., Slagle, B.L., 2007. Enhancement of Hepatitis B Virus Replication by the Regulatory X Protein In Vitro and In Vivo. Journal of Virology 81, 2656–2662. https://doi.org/10.1128/JVI.02020-06

Khoury, A., Achinger-Kawecka, J., Bert, S.A., Smith, G.C., French, H.J., Luu, P.-L., Peters, T.J., Du, Q., Parry, A.J., Valdes-Mora, F., Taberlay, P.C., Stirzaker, C., Statham, A.L., Clark, S.J., 2020. Constitutively bound CTCF sites maintain 3D chromatin architecture and long-range epigenetically regulated domains. Nat Commun 11, 54. https://doi.org/10.1038/s41467-019-13753-7

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., Salzberg, S.L., 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36. https://doi.org/10.1186/gb-2013-14-4-r36

Kim, D.H., Kang, H.S., Kim, K.-H., 2016. Roles of hepatocyte nuclear factors in hepatitis B virus infection. World Journal of Gastroenterology 22, 7017. https://doi.org/10.3748/wjg.v22.i31.7017

Kitamura, K., Que, L., Shimadu, M., Koura, M., Ishihara, Y., Wakae, K., Nakamura, T., Watashi, K., Wakita, T., Muramatsu, M., 2018. Flap endonuclease 1 is involved in cccDNA formation in the hepatitis B virus. PLOS Pathogens 14, e1007124. https://doi.org/10.1371/journal.ppat.1007124

Köck, J., Rösler, C., Zhang, J.-J., Blum, H.E., Nassal, M., Thoma, C., 2010. Generation of Covalently Closed Circular DNA of Hepatitis B Viruses via Intracellular Recycling Is Regulated in a Virus Specific Manner. PLoS Pathogens 6, e1001082. https://doi.org/10.1371/journal.ppat.1001082

Königer, C., Wingert, I., Marsmann, M., Rösler, C., Beck, J., Nassal, M., 2014. Involvement of the host DNA-repair enzyme TDP2 in formation of the covalently closed circular DNA persistence reservoir of hepatitis B viruses. Proc Natl Acad Sci U S A 111, E4244–E4253. https://doi.org/10.1073/pnas.1409986111

Kramvis, A., Kew, M.C., 1999. The core promoter of hepatitis B virus. Journal of Viral Hepatitis 13.

Kremsdorf, D., Strick-Marchand, H., 2017. Modeling hepatitis virus infections and treatment strategies in humanized mice. Current Opinion in Virology 25, 119–125. https://doi.org/10.1016/j.coviro.2017.07.029

Lamas-Maceiras, M., Singh, B.N., Hampsey, M., Freire-Picos, M.A., 2016. Promoter-Terminator Gene Loops Affect Alternative 3'-End Processing in Yeast. J. Biol. Chem. 291, 8960–8968. https://doi.org/10.1074/jbc.M115.687491

Lambert, C., Doring, T., Prange, R., 2007. Hepatitis B Virus Maturation Is Sensitive to Functional Inhibition of ESCRT-III, Vps4, and 2-Adaptin. Journal of Virology 81, 9050–9060. https://doi.org/10.1128/JVI.00479-07

Lambert, M.-P., Terrone, S., Giraud, G., Benoit-Pilven, C., Cluet, D., Combaret, V., Mortreux, F., Auboeuf, D., Bourgeois, C.F., 2018b. The RNA helicase DDX17 controls the transcriptional activity of REST and the expression of proneural microRNAs in neuronal differentiation. Nucleic Acids Research 46, 7686–7700. https://doi.org/10.1093/nar/gky545

Lamm, G., 1996. p72: a human nuclear DEAD box protein highly related to p68. Nucleic Acids Research 24, 3739–3747. https://doi.org/10.1093/nar/24.19.3739

Lan, Y.T., Li, J., Liao, W., Ou, J., 1999. Roles of the Three Major Phosphorylation Sites of Hepatitis B Virus Core Protein in Viral Replication. Virology 259, 342–348. https://doi.org/10.1006/viro.1999.9798

Lanford, R.E., Notvall, L., Lee, H., Beames, B., 1997. Transcomplementation of Nucleotide Priming and Reverse Transcription between Independently Expressed TP and RT Domains of the Hepatitis B Virus Reverse Transcriptase. J. VIROL. 71, 9.

Lang, F., Li, X., Vladimirova, O., Hu, B., Chen, G., Xiao, Y., Singh, V., Lu, D., Li, L., Han, H., Wickramasinghe, J.M.A.S.P., Smith, S.T., Zheng, C., Li, Q., Lieberman, P.M., Fraser, N.W., Zhou, J., 2017. CTCF interacts with the lytic HSV-1 genome to promote viral transcription. Sci Rep 7, 39861. https://doi.org/10.1038/srep39861

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Le Duff, Y., Blanchet, M., Sureau, C., 2009. The Pre-S1 and Antigenic Loop Infectivity Determinants of the Hepatitis B Virus Envelope Proteins Are Functionally Independent. Journal of Virology 83, 12443–12451. https://doi.org/10.1128/JVI.01594-09

LeCluyse, E.L., Alexandre, E., 2010. Isolation and Culture of Primary Hepatocytes from Resected Human Liver Tissue, in: Maurel, P. (Ed.), Hepatocytes. Humana Press, Totowa, NJ, pp. 57–82. https://doi.org/10.1007/978-1-60761-688-7_3

Lee, W.M., 1997. Hepatitis B Virus Infection. The New England Journal of Medicine 13.

Lee, Y., Rio, D.C., 2015. Mechanisms and Regulation of Alternative Pre-mRNA Splicing. Annual Review of Biochemistry 84, 291–323. https://doi.org/10.1146/annurev-biochem-060614-034316

Leistner, C.M., Gruen-Bernhard, S., Glebe, D., 2007. Role of glycosaminoglycans for binding and infection of hepatitis B virus. Cellular Microbiology 0, 070810224957001-??? https://doi.org/10.1111/j.1462-5822.2007.01023.x

Lev Maor, G., Yearim, A., Ast, G., 2015. The alternative role of DNA methylation in splicing regulation. Trends in Genetics 31, 274–280. https://doi.org/10.1016/j.tig.2015.03.002

Levrero, M., Zucman-Rossi, J., 2016. Mechanisms of HBV-induced hepatocellular carcinoma. Journal of Hepatology 64, S84–S101. https://doi.org/10.1016/j.jhep.2016.02.021

Li, T., Robert, E.I., van Breugel, P.C., Strubin, M., Zheng, N., 2010. A promiscuous α-helical motif anchors viral hijackers and substrate receptors to the CUL4–DDB1 ubiquitin ligase machinery. Nature Structural & Molecular Biology 17, 105–111. https://doi.org/10.1038/nsmb.1719

Lin, C.-L., Kao, J.-H., 2015. Hepatitis B Virus Genotypes and Variants. Cold Spring Harbor Perspectives in Medicine 5, a021436–a021436. https://doi.org/10.1101/cshperspect.a021436

Liu, Q., Somiya, M., Kuroda, S., 2016. Elucidation of the early infection machinery of hepatitis B virus by using bio-nanocapsule. World Journal of Gastroenterology 22, 8489. https://doi.org/10.3748/wjg.v22.i38.8489

Llères, D., Moindrot, B., Pathak, R., Piras, V., Matelot, M., Pignard, B., Marchand, A., Poncelet, M., Perrin, A., Tellier, V., Feil, R., Noordermeer, D., 2019. CTCF modulates allele-specific sub-TAD organization and imprinted gene activity at the mouse Dlk1-Dio3 and Igf2-H19 domains. Genome Biol. 20, 272. https://doi.org/10.1186/s13059-019-1896-8

Locarnini, S., Hatzakis, A., Chen, D.-S., Lok, A., 2015. Strategies to control hepatitis B: Public policy, epidemiology, vaccine and drugs. Journal of Hepatology 62, S76–S86. https://doi.org/10.1016/j.jhep.2015.01.018

Lok, A.S., Zoulim, F., Dusheiko, G., Ghany, M.G., 2017. Hepatitis B cure: From discovery to regulatory approval: Lok et al. Hepatology 66, 1296–1313. https://doi.org/10.1002/hep.29323

Lorgeoux, R.-P., Pan, Q., Le Duff, Y., Liang, C., 2013a. DDX17 promotes the production of infectious HIV-1 particles through modulating viral RNA packaging and translation frameshift. Virology 443, 384–392. https://doi.org/10.1016/j.virol.2013.05.026

Lorgeoux, R.-P., Pan, Q., Le Duff, Y., Liang, C., 2013b. DDX17 promotes the production of infectious HIV-1 particles through modulating viral RNA packaging and translation frameshift. Virology 443, 384–392. https://doi.org/10.1016/j.virol.2013.05.026

Luangsay, S., Ait-Goughoulte, M., Michelet, M., Floriot, O., Bonnin, M., Gruffaz, M., Rivoire, M., Fletcher, S., Javanbakht, H., Lucifora, J., Zoulim, F., Durantel, D., 2015. Expression and functionality of Toll- and RIG-like receptors in HepaRG cells. Journal of Hepatology 63, 1077–1085. https://doi.org/10.1016/j.jhep.2015.06.022

Lucifora, J., Arzberger, S., Durantel, D., Belloni, L., Strubin, M., Levrero, M., Zoulim, F., Hantz, O., Protzer, U., 2011a. Hepatitis B virus X protein is essential to initiate and maintain virus replication after infection. Journal of Hepatology 55, 996–1003. https://doi.org/10.1016/j.jhep.2011.02.015

Lucifora, J., Durantel, D., Testoni, B., Hantz, O., Levrero, M., Zoulim, F., 2010. Control of hepatitis B virus replication by innate response of HepaRG cells. Hepatology 51, 63–72. https://doi.org/10.1002/hep.23230

Lucifora, J., Salvetti, A., Marniquet, X., Mailly, L., Testoni, B., Fusil, F., Inchauspé, A., Michelet, M., Michel, M.-L., Levrero, M., Cortez, P., Baumert, T.F., Cosset, F.-L., Challier, C., Zoulim, F., Durantel, D., 2017a. Detection of the hepatitis B virus (HBV) covalently-closed-circular DNA (cccDNA) in mice transduced with a recombinant AAV-HBV vector. Antiviral Research 145, 14–19. https://doi.org/10.1016/j.antiviral.2017.07.006

Macovei, A., Petrareanu, C., Lazar, C., Florian, P., Branza-Nichita, N., 2013. Regulation of Hepatitis B Virus Infection by Rab5, Rab7, and the Endolysosomal Compartment. Journal of Virology 87, 6415–6427. https://doi.org/10.1128/JVI.00393-13

Macovei, A., Radulescu, C., Lazar, C., Petrescu, S., Durantel, D., Dwek, R.A., Zitzmann, N., Nichita, N.B., 2010. Hepatitis B Virus Requires Intact Caveolin-1 Function for Productive Infection in HepaRG Cells. Journal of Virology 84, 243–253. https://doi.org/10.1128/JVI.01207-09

Makokha, G.N., Abe-Chayama, H., Chowdhury, S., Hayes, C.N., Tsuge, M., Yoshima, T., Ishida, Y., Zhang, Y., Uchida, T., Tateno, C., Akiyama, R., Chayama, K., 2019. Regulation of the Hepatitis B virus replication and gene expression by the multi-functional protein TARDBP. Scientific Reports 9. https://doi.org/10.1038/s41598-019-44934-5

Mallinjoud, P., Villemin, J.-P., Mortada, H., Polay Espinoza, M., Desmet, F.-O., Samaan, S., Chautard, E., Tranchevent, L.-C., Auboeuf, D., 2014. Endothelial, epithelial, and fibroblast cells exhibit specific splicing programs independently of their tissue of origin. Genome Research 24, 511–521. https://doi.org/10.1101/gr.162933.113

Marino, M.M., Rega, C., Russo, R., Valletta, M., Gentile, M.T., Esposito, S., Baglivo, I., De Feis, I., Angelini, C., Xiao, T., Felsenfeld, G., Chambery, A., Pedone, P.V., 2019. Interactome mapping defines BRG1, a component of the SWI/SNF chromatin remodeling complex, as a new partner of the transcriptional regulator CTCF. Journal of Biological Chemistry 294, 861–873. https://doi.org/10.1074/jbc.RA118.004882

Martin Stoltzfus, C., 2009. Chapter 1 Regulation of HIV-1 Alternative RNA Splicing and Its Role in Virus Replication, in: Advances in Virus Research. Elsevier, pp. 1–40. https://doi.org/10.1016/S0065-3527(09)74001-1

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

Marzluff, W.F., Wagner, E.J., Duronio, R.J., 2008. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet 9, 843–854. https://doi.org/10.1038/nrg2438

Mayer, A., Landry, H.M., Churchman, L.S., 2017. Pause & go: from the discovery of RNA polymerase pausing to its functional implications. Current Opinion in Cell Biology 46, 72–80. https://doi.org/10.1016/j.ceb.2017.03.002

McAuliffe, V.J., Purcell, R.H., Gerin, J.L., 1980. Type B Hepatitis: A Review of Current Prospects for a Safe and Effective Vaccine. Reviews of Infectious Diseases 2, 470–492. https://doi.org/10.1093/clinids/2.3.470

McNaughton, A.L., Roberts, H.E., Bonsall, D., de Cesare, M., Mokaya, J., Lumley, S.F., Golubchik, T., Piazza, P., Martin, J.B., de Lara, C., Brown, A., Ansari, M.A., Bowden, R., Barnes, E., Matthews, P.C., 2019. Illumina and Nanopore methods for whole genome sequencing of hepatitis B virus (HBV). Sci Rep 9, 7081. https://doi.org/10.1038/s41598-019-43524-9

Mehta, K., Gunasekharan, V., Satsuka, A., Laimins, L.A., 2015. Human Papillomaviruses Activate and Recruit SMC1 Cohesin Proteins for the Differentiation-Dependent Life Cycle through Association with CTCF Insulators. PLoS Pathog 11, e1004763. https://doi.org/10.1371/journal.ppat.1004763

Mercer, Tim R, Edwards, S.L., Clark, M.B., Neph, S.J., Wang, H., Stergachis, A.B., John, S., Sandstrom, R., Li, G., Sandhu, K.S., Ruan, Y., Nielsen, L.K., Mattick, J.S., Stamatoyannopoulos, J.A., 2013a. DNase I–hypersensitive exons colocalize with promoters and distal regulatory elements. Nature Genetics 45, 852–859. https://doi.org/10.1038/ng.2677

Mersaoui, S.Y., Yu, Z., Coulombe, Y., Karam, M., Busatto, F.F., Masson, J., Richard, S., 2019a. Arginine methylation of the DDX 5 helicase RGG / RG motif by PRMT 5 regulates resolution of RNA:DNA hybrids. The EMBO Journal 38. https://doi.org/10.15252/embj.2018100986

Messageot, F., Salhi, S., Eon, P., Rossignol, J.-M., 2003. Proteolytic Processing of the Hepatitis B Virus e Antigen Precursor: CLEAVAGE AT TWO FURIN CONSENSUS SEQUENCES. Journal of Biological Chemistry 278, 891–895. https://doi.org/10.1074/jbc.M207634200

Meuleman, P., Libbrecht, L., Wieland, S., De Vos, R., Habib, N., Kramvis, A., Roskams, T., Leroux-Roels, G., 2006. Immune Suppression Uncovers Endogenous Cytopathic Effects of the Hepatitis B Virus. Journal of Virology 80, 2797–2807. https://doi.org/10.1128/JVI.80.6.2797-2807.2006

Mhamdi, M., Funk, A., Hohenberg, H., Will, H., Sirma, H., 2007. Assembly and budding of a hepatitis B virus is mediated by a novel type of intracellular vesicles. Hepatology 46, 95–106. https://doi.org/10.1002/hep.21666

Mogensen, T.H., 2009. Pathogen Recognition and Inflammatory Signaling in Innate Immune Defenses. Clinical Microbiology Reviews 22, 240–273. https://doi.org/10.1128/CMR.00046-08

Moolla, N., Kew, M., Arbuthnot, P., 2002. Regulatory elements of hepatitis B virus transcription. Journal of Viral Hepatitis 9, 323–331. https://doi.org/10.1046/j.1365-2893.2002.00381.x

Moreau, P., Cournac, A., Palumbo, G.A., Marbouty, M., Mortaza, S., Thierry, A., Cairo, S., Lavigne, M., Koszul, R., Neuveut, C., 2018. Tridimensional infiltration of DNA viruses into the host genome shows preferential contact with active chromatin. Nature Communications 9. https://doi.org/10.1038/s41467-018-06739-4

Mueller, H., Lopez, A., Tropberger, P., Wildum, S., Schmaler, J., Pedersen, L., Han, X., Wang, Y., Ottosen, S., Yang, S., Young, J.A.T., Javanbakht, H., 2019. PAPD5/7 Are Host Factors That Are Required for Hepatitis B Virus RNA Stabilization. Hepatology 69, 1398–1411. https://doi.org/10.1002/hep.30329

Naftelberg, S., Schor, I.E., Ast, G., Kornblihtt, A.R., 2015. Regulation of Alternative Splicing Through Coupling with Transcription and Chromatin Structure. Annual Review of Biochemistry 84, 165–198. https://doi.org/10.1146/annurev-biochem-060614-034242

Nassal, M., 2015a. 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., 1992. The Arginine-Rich Domain of the Hepatitis B Virus Core Protein Is Required for Pregenome Encapsidation and Productive Viral Positive-Strand DNA Synthesis but Not for Virus Assembly. J. VIROL. 66, 10.

Nassal, M., Rieger, A., Steinau, O., 1992. Topological analysis of the hepatitis B virus core particle by cysteine-cysteine cross-linking. Journal of Molecular Biology 225, 1013–1025. https://doi.org/10.1016/0022-2836(92)90101-O

Nechaev, S., Adelman, K., 2011. Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1809, 34–45. https://doi.org/10.1016/j.bbagrm.2010.11.001

Neugebauer, K.M., 2002. On the importance of being co-transcriptional. Journal of Cell Science 115, 3865–3871. https://doi.org/10.1242/jcs.00073

Neve, J., Patel, R., Wang, Z., Louey, A., Furger, A.M., 2017. Cleavage and polyadenylation: Ending the message expands gene regulation. RNA Biology 14, 865–890. https://doi.org/10.1080/15476286.2017.1306171

Newbold, J.E., Xin, H., Tencza, M., Sherman, G., Dean, J., Bowden, S., Locarnini, S., 1995. The Covalently Closed Duplex Form of the Hepadnavirus Genome Exists In Situ as a Heterogeneous Population of Viral Minichromosomes. J. VIROL. 69, 8.

Ni, Y., Lempp, F.A., Mehrle, S., Nkongolo, S., Kaufman, C., Fälth, M., Stindt, J., Königer, C., Nassal, M., Kubitz, R., Sültmann, H., Urban, S., 2014. Hepatitis B and D Viruses Exploit Sodium Taurocholate Co-transporting Polypeptide for Species-Specific Entry into Hepatocytes. Gastroenterology 146, 1070-1083.e6. https://doi.org/10.1053/j.gastro.2013.12.024

Ni, Y., Urban, S., 2017. Hepatitis B Virus Infection of HepaRG Cells, HepaRG-hNTCP Cells, and Primary Human Hepatocytes, in: Guo, H., Cuconati, A. (Eds.), Hepatitis B Virus. Springer New York, New York, NY, pp. 15–25. https://doi.org/10.1007/978-1-4939-6700-1_2

Nilsson, K., Wu, C., Schwartz, S., 2018. Role of the DNA Damage Response in Human Papillomavirus RNA Splicing and Polyadenylation. IJMS 19, 1735. https://doi.org/10.3390/ijms19061735

Ning, X., Nguyen, D., Mentzer, L., Adams, C., Lee, H., Ashley, R., Hafenstein, S., Hu, J., 2011. Secretion of Genome-Free Hepatitis B Virus – Single Strand Blocking Model for Virion Morphogenesis of Para-retrovirus. PLoS Pathog 7, e1002255. https://doi.org/10.1371/journal.ppat.1002255

Nishitsuji, H., Ujino, S., Harada, K., Shimotohno, K., 2018a. TIP60 Complex Inhibits Hepatitis B Virus Transcription. J Virol 92, e01788-17, /jvi/92/6/e01788-17.atom. https://doi.org/10.1128/JVI.01788-17

Nishitsuji, H., Ujino, S., Harada, K., Shimotohno, K., 2018b. TIP60 Complex Inhibits Hepatitis B Virus Transcription. Journal of Virology 92. https://doi.org/10.1128/JVI.01788-17

Niu, C., Livingston, C.M., Li, L., Beran, R.K., Daffis, S., Ramakrishnan, D., Burdette, D., Peiser, L., Salas, E., Ramos, H., Yu, M., Cheng, G., Strubin, M., Delaney IV, W.E., Fletcher, S.P., 2017. The Smc5/6 Complex Restricts HBV when Localized to ND10 without Inducing an Innate Immune Response and Is Counteracted by the HBV X Protein Shortly after Infection. PLoS ONE 12, e0169648. https://doi.org/10.1371/journal.pone.0169648

Obert, S., Zachmann-Brand, B., Deindl, E., Tucker, W., Bartenschlager, R., Schaller, H., n.d. A spliced hepadnavirus RNA that is essential for virus replication 10.

Ogawa, T., Enomoto, M., Fujii, H., Sekiya, Y., Yoshizato, K., Ikeda, K., Kawada, N., 2012. MicroRNA-221/222 upregulation indicates the activation of stellate cells and the progression of liver fibrosis. Gut 61, 1600–1609. https://doi.org/10.1136/gutjnl-2011-300717

Ogilvie, V.C., 2003. The highly related DEAD box RNA helicases p68 and p72 exist as heterodimers in cells. Nucleic Acids Research 31, 1470–1480. https://doi.org/10.1093/nar/gkg236

Ortega-Prieto, A.M., Cherry, C., Gunn, H., Dorner, M., 2019. *In Vivo* Model Systems for Hepatitis B Virus Research. ACS Infectious Diseases 5, 688–702. https://doi.org/10.1021/acsinfecdis.8b00223

Osseman, Q., Gallucci, L., Au, S., Cazenave, C., Berdance, E., Blondot, M.-L., Cassany, A., Bégu, D., Ragues, J., Aknin, C., Sominskaya, I., Dishlers, A., Rabe, B., Anderson, F., Panté, N., Kann, M., 2018. The chaperone dynein LL1 mediates cytoplasmic transport of empty and mature hepatitis B virus capsids. Journal of Hepatology 68, 441–448. https://doi.org/10.1016/j.jhep.2017.10.032

Padmanabhan, K., Robles, M.S., Westerling, T., Weitz, C.J., 2012a. Feedback Regulation of Transcriptional Termination by the Mammalian Circadian Clock PERIOD Complex. Science 337, 599–602. https://doi.org/10.1126/science.1221592

Paran, N., Ori, A., Haviv, I., Shaul, Y., 2000. A Composite Polyadenylation Signal with TATA Box Function. Molecular and Cellular Biology 20, 834–841. https://doi.org/10.1128/MCB.20.3.834-841.2000

Park, G.-S., Kim, H.-Y., Shin, H.-S., Park, S., Shin, H.-J., Kim, K., 2008. Modulation of hepatitis B virus replication by expression of polymerase-surface fusion protein through splicing: Implications for viral persistence. Virus Research 136, 166–174. https://doi.org/10.1016/j.virusres.2008.05.005

Pastor, F., Herrscher, C., Patient, R., Eymieux, S., Moreau, A., Burlaud-Gaillard, J., Seigneuret, F., de Rocquigny, H., Roingeard, P., Hourioux, C., 2019. Direct interaction between the hepatitis B virus core and envelope proteins analyzed in a cellular context. Scientific Reports 9. https://doi.org/10.1038/s41598-019-52824-z

Patient, R., Hourioux, C., Roingeard, P., 2009. Morphogenesis of hepatitis B virus and its subviral envelope particles. Cellular Microbiology 11, 1561–1570. https://doi.org/10.1111/j.1462-5822.2009.01363.x

Patient, R., Hourioux, C., Sizaret, P.-Y., Trassard, S., Sureau, C., Roingeard, P., 2007. Hepatitis B Virus Subviral Envelope Particle Morphogenesis and Intracellular Trafficking. Journal of Virology 81, 3842–3851. https://doi.org/10.1128/JVI.02741-06

Patil, D.P., Pickering, B.F., Jaffrey, S.R., 2018. Reading m6A in the Transcriptome: m6A-Binding Proteins. Trends in Cell Biology 28, 113–127. https://doi.org/10.1016/j.tcb.2017.10.001

Pentland, I., Campos-León, K., Cotic, M., Davies, K.-J., Wood, C.D., Groves, I.J., Burley, M., Coleman, N., Stockton, J.D., Noyvert, B., Beggs, A.D., West, M.J., Roberts, S., Parish, J.L., 2018a. Disruption of CTCF-YY1–dependent looping of the human papillomavirus genome activates differentiation-induced viral oncogene transcription. PLoS Biol 16, e2005752. https://doi.org/10.1371/journal.pbio.2005752

Pentland, I., Parish, J., 2015a. Targeting CTCF to Control Virus Gene Expression: A Common Theme amongst Diverse DNA Viruses. Viruses 7, 3574–3585. https://doi.org/10.3390/v7072791

Perfumo, S., Amicone, L., Colloca, S., Giorgio, M., Pozzi, L., Tripodi, M., 1992. Recognition efficiency of the hepatitis B virus polyadenylation signals is tissue specific in transgenic mice. Journal of Virology 66, 6819–6823. https://doi.org/10.1128/JVI.66.11.6819-6823.1992

Pol, J.G., Lekbaby, B., Redelsperger, F., Klamer, S., Mandouri, Y., Ahodantin, J., Bieche, I., Lefevre, M., Souque, P., Charneau, P., Gadessaud, N., Kremsdorf, D., Soussan, P., 2015. Alternative splicing-regulated protein of hepatitis B virus hacks the TNF- α -stimulated signaling pathways and limits the extent of liver inflammation. The FASEB Journal 29, 1879–1889. https://doi.org/10.1096/fj.14-258715

Pollicino, T., Belloni, L., Raffa, G., Pediconi, N., Squadrito, G., Raimondo, G., Levrero, M., 2006. Hepatitis B Virus Replication Is Regulated by the Acetylation Status of Hepatitis B Virus cccDNA-Bound H3 and H4 Histones. Gastroenterology 130, 823–837. https://doi.org/10.1053/j.gastro.2006.01.001

Pourcel, C., Louise, A., Gervais, M., Chenciner, N., Dubois, M.F., Tiollais, P., 1982. Transcription of the hepatitis B surface antigen gene in mouse cells transformed with cloned viral DNA. Journal of Virology 42, 100–105. https://doi.org/10.1128/JVI.42.1.100-105.1982

Pourcel, C., Summers, J., n.d. rmation of the Pool of Covalently Close ular Viral DNA in Hepadnavirus-Infected 10.

Prange, R., 2012. Host factors involved in hepatitis B virus maturation, assembly, and egress. Medical Microbiology and Immunology 201, 449–461. https://doi.org/10.1007/s00430-012-0267-9

Proudfoot, N.J., 2016. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. Science 352, aad9926–aad9926. https://doi.org/10.1126/science.aad9926

Qi, Y., Gao, Z., Xu, G., Peng, B., Liu, C., Yan, H., Yao, Q., Sun, G., Liu, Y., Tang, D., Song, Z., He, W., Sun, Y., Guo, J.-T., Li, W., 2016. DNA Polymerase κ Is a Key Cellular Factor for the Formation of Covalently Closed Circular DNA of Hepatitis B Virus. PLOS Pathogens 12, e1005893. https://doi.org/10.1371/journal.ppat.1005893

Rabe, B., Vlachou, A., Pante, N., Helenius, A., Kann, M., 2003. Nuclear import of hepatitis B virus capsids and release of the viral genome. Proceedings of the National Academy of Sciences 100, 9849–9854. https://doi.org/10.1073/pnas.1730940100

Rahman, M.M., Bagdassarian, E., Ali, M.A.M., McFadden, G., 2017. Identification of host DEAD-box RNA helicases that regulate cellular tropism of oncolytic Myxoma virus in human cancer cells. Scientific Reports 7. https://doi.org/10.1038/s41598-017-15941-1

Rajoriya, N., Combet, C., Zoulim, F., Janssen, H.L.A., 2017. How viral genetic variants and genotypes influence disease and treatment outcome of chronic hepatitis B. Time for an individualised approach? Journal of Hepatology 67, 1281–1297. https://doi.org/10.1016/j.jhep.2017.07.011

Rall, L.B., Standring, D.N., Laub, O., Rutter, W.J., 1983. Transcription of hepatitis B virus by RNA polymerase II. Molecular and Cellular Biology 3, 1766–1773. https://doi.org/10.1128/MCB.3.10.1766

Ramanathan, A., Robb, G.B., Chan, S.-H., 2016. mRNA capping: biological functions and applications. Nucleic Acids Research 44, 7511–7526. https://doi.org/10.1093/nar/gkw551

Ramanouskaya, T.V., Grinev, V.V., 2017. The determinants of alternative RNA splicing in human cells. Molecular Genetics and Genomics 292, 1175–1195. https://doi.org/10.1007/s00438-017-1350-0

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., Aiden, E.L., 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Rat, V., Seigneuret, F., Burlaud-Gaillard, J., Lemoine, R., Hourioux, C., Zoulim, F., Testoni, B., Meunier, J.-C., Tauber, C., Roingeard, P., de Rocquigny, H., 2019. BAY 41-4109-mediated aggregation of assembled and misassembled HBV capsids in cells revealed by electron microscopy. Antiviral Research 169, 104557. https://doi.org/10.1016/j.antiviral.2019.104557

Raviram, R., Rocha, P.P., Bonneau, R., Skok, J.A., 2014. Interpreting 4C-Seq data: how far can we go? Epigenomics 6, 455–457. https://doi.org/10.2217/epi.14.47

Ren, G., Jin, W., Cui, K., Rodrigez, J., Hu, G., Zhang, Z., Larson, D.R., Zhao, K., 2017. CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression. Molecular Cell 67, 1049-1058.e6. https://doi.org/10.1016/j.molcel.2017.08.026

Ren, J., Hu, J., Cheng, S., Yu, H., Wong, V.K.W., Law, B.Y.K., Yang, Y., Huang, Y., Liu, Y., Chen, W., Cai, X., Tang, H., Hu, Y., Zhang, W., Liu, X., Long, Q., Zhou, L., Tao, N., Zhou, H., Yang, Q., Ren, F., He, L., Gong, R., Huang, A., Chen, J., 2018. SIRT3 restricts hepatitis B virus transcription and replication through epigenetic regulation of covalently closed circular DNA involving suppressor of variegation 3-9 homolog 1 and SET domain containing 1A histone methyltransferases. Hepatology 68, 1260–1276. https://doi.org/10.1002/hep.29912

Rieger, A., Nassal, M., 1996. Specific Hepatitis B Virus Minus-Strand DNA Synthesis Requires Only the 5J Encapsidation Signal and the 3J-Proximal Direct Repeat DR1. J. VIROL. 70, 5.

Rivière, L., Gerossier, L., Ducroux, A., Dion, S., Deng, Q., Michel, M.-L., Buendia, M.-A., Hantz, O., Neuveut, C., 2015. HBx relieves chromatin-mediated transcriptional repression of

hepatitis B viral cccDNA involving SETDB1 histone methyltransferase. Journal of Hepatology 63, 1093–1102. https://doi.org/10.1016/j.jhep.2015.06.023

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 Research 172, 104618. https://doi.org/10.1016/j.antiviral.2019.104618

Roingeard, P., Lu, S., Sureau, C., Freschlin, M., Brigi, Essex, M., Romet-Lemonne, J.-L., 1990. Immunocytochemical and electron microscopic study of hepatitis B virus antigen and complete particle production in hepatitis B virus DNA transfected HepG2 cells. Hepatology 11, 277–285. https://doi.org/10.1002/hep.1840110219

Rokuhara, A., Matsumoto, A., Tanaka, E., Umemura, T., Yoshizawa, K., Kimura, T., Maki, N., Kiyosawa, K., 2006. Hepatitis B virus RNA is measurable in serum and can be a new marker for monitoring lamivudine therapy. Journal of Gastroenterology 41, 785–790. https://doi.org/10.1007/s00535-006-1856-4

Ruiz-Velasco, M., Kumar, M., Lai, M.C., Bhat, P., Solis-Pinson, A.B., Reyes, A., Kleinsorg, S., Noh, K.-M., Gibson, T.J., Zaugg, J.B., 2017a. CTCF-Mediated Chromatin Loops between Promoter and Gene Body Regulate Alternative Splicing across Individuals. Cell Systems 5, 628-637.e6. https://doi.org/10.1016/j.cels.2017.10.018

Ruiz-Velasco, M., Kumar, M., Lai, M.C., Bhat, P., Solis-Pinson, A.B., Reyes, A., Kleinsorg, S., Noh, K.-M., Gibson, T.J., Zaugg, J.B., 2017b. CTCF-Mediated Chromatin Loops between Promoter and Gene Body Regulate Alternative Splicing across Individuals. Cell Syst 5, 628-637.e6. https://doi.org/10.1016/j.cels.2017.10.018

Sainsbury, S., Bernecky, C., Cramer, P., 2015. Structural basis of transcription initiation by RNA polymerase II. Nature Reviews Molecular Cell Biology 16, 129–143. https://doi.org/10.1038/nrm3952

Saldana-Meyer, R., Gonzalez-Buendia, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., Reinberg, D., 2014. CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. Genes & Development 28, 723–734. https://doi.org/10.1101/gad.236869.113

Saldaña-Meyer, R., Rodriguez-Hernaez, J., Escobar, T., Nishana, M., Jácome-López, K., Nora, E.P., Bruneau, B.G., Tsirigos, A., Furlan-Magaril, M., Skok, J., Reinberg, D., 2019. RNA Interactions Are Essential for CTCF-Mediated Genome Organization. Mol. Cell 76, 412-422.e5. https://doi.org/10.1016/j.molcel.2019.08.015

Sandberg, R., Neilson, J.R., Sarma, A., Sharp, P.A., Burge, C.B., 2008. Proliferating Cells Express mRNAs with Shortened 3' Untranslated Regions and Fewer MicroRNA Target Sites. Science 320, 1643–1647. https://doi.org/10.1126/science.1155390

Sauvage, V., Boizeau, L., Candotti, D., Vandenbogaert, M., Servant-Delmas, A., Caro, V., Laperche, S., 2018. Early MinION™ nanopore single-molecule sequencing technology enables the characterization of hepatitis B virus genetic complexity in clinical samples. PLoS ONE 13, e0194366. https://doi.org/10.1371/journal.pone.0194366

Schaefer, S., 2007. Hepatitis B virus taxonomy and hepatitis B virus genotypes. World Journal of Gastroenterology 13, 14. https://doi.org/10.3748/wjg.v13.i1.14

Schmitz, A., Schwarz, A., Foss, M., Zhou, L., Rabe, B., Hoellenriegel, J., Stoeber, M., Panté, N., Kann, M., 2010. Nucleoporin 153 Arrests the Nuclear Import of Hepatitis B Virus Capsids in the Nuclear Basket. PLoS Pathogens 6, e1000741. https://doi.org/10.1371/journal.ppat.1000741

Schutz, T., Kairat, A., Schröder, C.H., 1996. DNA Sequence Requirements for the Activation of a CATAAA Polyadenylation Signal within the Hepatitis B Virus X Reading Frame: Rapid Detection of Truncated Transcripts. Virology 223, 401–405. https://doi.org/10.1006/viro.1996.0495

Sedlackova, L., Perkins, K.D., Lengyel, J., Strain, A.K., van Santen, V.L., Rice, S.A., 2008. Herpes Simplex Virus Type 1 ICP27 Regulates Expression of a Variant, Secreted Form of Glycoprotein C by an Intron Retention Mechanism. Journal of Virology 82, 7443–7455. https://doi.org/10.1128/JVI.00388-08

Seeger, C., Mason, W.S., 2000a. Hepatitis B Virus Biology. Microbiology and Molecular Biology Reviews 64, 51–68. https://doi.org/10.1128/MMBR.64.1.51-68.2000

Seeger, C., Mason, W.S., 2000b. Hepatitis B Virus Biology. Microbiology and Molecular Biology Reviews 64, 51–68. https://doi.org/10.1128/MMBR.64.1.51-68.2000

Sen, N., Cao, F., Tavis, J.E., 2004. Translation of Duck Hepatitis B Virus Reverse Transcriptase by Ribosomal Shunting. Journal of Virology 78, 11751–11757. https://doi.org/10.1128/JVI.78.21.11751-11757.2004

Shen, S., Park, J.W., Lu, Z., Lin, L., Henry, M.D., Wu, Y.N., Zhou, Q., Xing, Y., 2014. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci USA 111, E5593–E5601. https://doi.org/10.1073/pnas.1419161111

Sheu, S.Y., Lo, S.J., 1992. Preferential ribosomal scanning is involved in the differential synthesis of the hepatitis B viral surface antigens from subgenomic transcripts. Virology 188, 353–357. https://doi.org/10.1016/0042-6822(92)90764-G

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., Oberdoerffer, S., 2011a. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature 479, 74–79. https://doi.org/10.1038/nature10442

Sibley, C.R., Blazquez, L., Ule, J., 2016. Lessons from non-canonical splicing. Nature Reviews Genetics 17, 407–421. https://doi.org/10.1038/nrg.2016.46

Simonsen, C.C., Levinson, A.D., 1983. Analysis of processing and polyadenylation signals of the hepatitis B virus surface antigen gene by using simian virus 40-hepatitis B virus chimeric plasmids. Molecular and Cellular Biology 3, 2250–2258. https://doi.org/10.1128/MCB.3.12.2250

Sithole, N., Williams, C.A., Vaughan, A.M., Kenyon, J.C., Lever, A.M.L., 2018a. DDX17 Specifically, and Independently of DDX5, Controls Use of the HIV A4/5 Splice Acceptor Cluster and Is Essential for Efficient Replication of HIV. Journal of Molecular Biology 430, 3111–3128. https://doi.org/10.1016/j.jmb.2018.06.052

Sithole, N., Williams, C.A., Vaughan, A.M., Kenyon, J.C., Lever, A.M.L., 2018b. DDX17 Specifically, and Independently of DDX5, Controls Use of the HIV A4/5 Splice Acceptor Cluster and Is Essential for Efficient Replication of HIV. Journal of Molecular Biology 430, 3111–3128. https://doi.org/10.1016/j.jmb.2018.06.052

Sommer, G., Günther, S., Sterneck, M., Otto, S., Will, H., 1997. A New Class of Defective Hepatitis B Virus Genomes with an Internal Poly(dA) Sequence. Virology 239, 402–412. https://doi.org/10.1006/viro.1997.8898

Soussan, P., Garreau, F., Zylberberg, H., Ferray, C., Brechot, C., Kremsdorf, D., 2000. In vivo expression of a new hepatitis B virus protein encoded by a spliced RNA. Journal of Clinical Investigation 105, 55–60. https://doi.org/10.1172/JCI8098

Soussan, P., Pol, J., Garreau, F., Schneider, V., Pendeven, C.L., Nalpas, B., Lacombe, K., Bonnard, P., Pol, S., Kremsdorf, D., 2008a. Expression of Defective Hepatitis B Virus Particles Derived from Singly Spliced RNA Is Related to Liver Disease. The Journal of Infectious Diseases 198, 218–225. https://doi.org/10.1086/589623

Soussan, P., Tuveri, R., Nalpas, B., Garreau, F., Zavala, F., Masson, A., Pol, S., Brechot, C., Kremsdorf, D., 2003a. The expression of hepatitis B spliced protein (HBSP) encoded by a spliced hepatitis B virus RNA is associated with viral replication and liver fibrosis. Journal of Hepatology 38, 343–348. https://doi.org/10.1016/S0168-8278(02)00422-1

Stadelmayer, B., Diederichs, A., Chapus, F., Rivoire, M., Neveu, G., Alam, A., Fraisse, L., Carter, K., Testoni, B., Zoulim, F., 2020a. Full-length 5'RACE identifies all major HBV transcripts in HBV-infected hepatocytes and patient serum. Journal of Hepatology. https://doi.org/10.1016/j.jhep.2020.01.028

Staprans, S., Loeb, D.D., Ganem, D., 1991. Mutations affecting hepadnavirus plus-strand DNA synthesis dissociate primer cleavage from translocation and reveal the origin of linear viral DNA. Journal of Virology 65, 1255–1262. https://doi.org/10.1128/JVI.65.3.1255-1262.1991

Stuchell, M.D., Garrus, J.E., Müller, B., Stray, K.M., Ghaffarian, S., McKinnon, R., Kräusslich, H.-G., Morham, S.G., Sundquist, W.I., 2004. The Human Endosomal Sorting Complex Required for Transport (ESCRT-I) and Its Role in HIV-1 Budding. J. Biol. Chem. 279, 36059–36071. https://doi.org/10.1074/jbc.M405226200

Su, Q., Wang, S.-F., Chang, T.-E., Breitkreutz, R., Hennig, H., Takegoshi, K., Edler, L., Schroder, C.H., n.d. Circulating Hepatitis B Virus Nucleic Acids in Chronic Infection: Representation of Differently Polyadenylated Viral Transcripts during Progression to Nonreplicative Stages 12.

Su, T.-S., Lai, C.-J., Huang, J.-L., Lin, L.-H., Yauk, Y.-K., Chang, C., Lo, S.J., Han, S.-H., 1989. Hepatitis B Virus Transcript Produced by RNA Splicing. J. VIROL. 63, 8.

Suh, H., Ficarro, S.B., Kang, U.-B., Chun, Y., Marto, J.A., Buratowski, S., 2016. Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. Molecular Cell 61, 297–304. https://doi.org/10.1016/j.molcel.2015.12.021

Summers, J., O'Connell, A., Millman, I., 1975. Genome of hepatitis B virus: restriction enzyme cleavage and structure of DNA extracted from Dane particles. Proceedings of the National Academy of Sciences 72, 4597–4601. https://doi.org/10.1073/pnas.72.11.4597

Sun, S., Nakashima, K., Ito, M., Li, Y., Chida, T., Takahashi, H., Watashi, K., Sawasaki, T., Wakita, T., Suzuki, T., 2017. Involvement of PUF60 in Transcriptional and Post-transcriptional Regulation of Hepatitis B Virus Pregenomic RNA Expression. Scientific Reports 7. https://doi.org/10.1038/s41598-017-12497-y

Sunbul, M., 2014. Hepatitis B virus genotypes: Global distribution and clinical importance. World Journal of Gastroenterology 20, 5427. https://doi.org/10.3748/wjg.v20.i18.5427

Svejstrup, J.Q., 2012. Transcription: another mark in the tail: Transcription: another mark in the tail. The EMBO Journal 31, 2753–2754. https://doi.org/10.1038/emboj.2012.154

Tang, H., Delgermaa, L., Huang, F., Oishi, N., Liu, L., He, F., Zhao, L., Murakami, S., 2005. The Transcriptional Transactivation Function of HBx Protein Is Important for Its Augmentation Role in Hepatitis B Virus Replication. Journal of Virology 79, 5548–5556. https://doi.org/10.1128/JVI.79.9.5548-5556.2005

Tang, S., Patel, A., Krause, P.R., 2019. Hidden regulation of herpes simplex virus 1 pre-mRNA splicing and polyadenylation by virally encoded immediate early gene ICP27. PLoS Pathog 15, e1007884. https://doi.org/10.1371/journal.ppat.1007884

Tang, Z., Luo, O.J., Li, X., Zheng, M., Zhu, J.J., Szalaj, P., Trzaskoma, P., Magalska, A., Wlodarczyk, J., Ruszczycki, B., Michalski, P., Piecuch, E., Wang, P., Wang, D., Tian, S.Z., Penrad-Mobayed, M., Sachs, L.M., Ruan, X., Wei, C.-L., Liu, E.T., Wilczynski, G.M., Plewczynski, D., Li, G., Ruan, Y., 2015. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. Cell 163, 1611–1627. https://doi.org/10.1016/j.cell.2015.11.024

Tavis, J.E., Lomonosova, E., 2015. The hepatitis B virus ribonuclease H as a drug target. Antiviral Research 118, 132–138. https://doi.org/10.1016/j.antiviral.2015.04.002

Tempera, I., Klichinsky, M., Lieberman, P.M., 2011. EBV Latency Types Adopt Alternative Chromatin Conformations. PLoS Pathog 7, e1002180. https://doi.org/10.1371/journal.ppat.1002180

Tian, B., Manley, J.L., 2017. Alternative polyadenylation of mRNA precursors. Nature Reviews Molecular Cell Biology 18, 18–30. https://doi.org/10.1038/nrm.2016.116

Tian, Y., Yang, W., Song, J., Wu, Y., Ni, B., 2013. Hepatitis B Virus X Protein-Induced Aberrant Epigenetic Modifications Contributing to Human Hepatocellular Carcinoma Pathogenesis. Molecular and Cellular Biology 33, 2810–2816. https://doi.org/10.1128/MCB.00205-13

Tombácz, D., Moldován, N., Balázs, Z., Gulyás, G., Csabai, Z., Boldogkői, M., Snyder, M., Boldogkői, Z., 2019. Multiple Long-Read Sequencing Survey of Herpes Simplex Virus Dynamic Transcriptome. Front. Genet. 10, 834. https://doi.org/10.3389/fgene.2019.00834

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

Tropberger, P., Mercier, A., Robinson, M., Zhong, W., Ganem, D.E., Holdorf, M., 2015a. Mapping of histone modifications in episomal HBV cccDNA uncovers an unusual chromatin organization amenable to epigenetic manipulation. Proc Natl Acad Sci USA 112, E5715–E5724. https://doi.org/10.1073/pnas.1518090112

Tsutakawa, S.E., Classen, S., Chapados, B.R., Arvai, A.S., Finger, L.D., Guenther, G., Tomlinson, C.G., Thompson, P., Sarker, A.H., Shen, B., Cooper, P.K., Grasby, J.A., Tainer, J.A., 2011. Human Flap Endonuclease Structures, DNA Double-Base Flipping, and a Unified Understanding of the FEN1 Superfamily. Cell 145, 198–211. https://doi.org/10.1016/j.cell.2011.03.004

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

Verrier, E., Colpitts, C., Schuster, C., Zeisel, M., Baumert, T., 2016a. Cell Culture Models for the Investigation of Hepatitis B and D Virus Infection. Viruses 8, 261. https://doi.org/10.3390/v8090261

Verrier, E.R., Colpitts, C.C., Bach, C., Heydmann, L., Weiss, A., Renaud, M., Durand, S.C., Habersetzer, F., Durantel, D., Abou-Jaoudé, G., López Ledesma, M.M., Felmlee, D.J., Soumillon, M., Croonenborghs, T., Pochet, N., Nassal, M., Schuster, C., Brino, L., Sureau, C., Zeisel, M.B., Baumert, T.F., 2016. A targeted functional RNA interference screen uncovers glypican 5 as an entry factor for hepatitis B and D viruses: VIRAL HEPATITIS. Hepatology 63, 35–48. https://doi.org/10.1002/hep.28013

Vilborg, A., Sabath, N., Wiesel, Y., Nathans, J., Levy-Adam, F., Yario, T.A., Steitz, J.A., Shalgi, R., 2017. Comparative analysis reveals genomic features of stress-induced transcriptional readthrough. Proceedings of the National Academy of Sciences 114, E8362–E8371. https://doi.org/10.1073/pnas.1711120114

Villanueva, A., 2019. Hepatocellular Carcinoma. New England Journal of Medicine 380, 1450–1462. https://doi.org/10.1056/NEJMra1713263

Vivekanandan, P., Daniel, H.D.J., Kannangai, R., Martinez-Murillo, F., Torbenson, M., 2010. Hepatitis B Virus Replication Induces Methylation of both Host and Viral DNA. Journal of Virology 84, 4321–4329. https://doi.org/10.1128/JVI.02280-09

Vivekanandan, P., Thomas, D., Torbenson, M., 2009. Methylation Regulates Hepatitis B Viral Protein Expression. The Journal of Infectious Diseases 199, 1286–1291. https://doi.org/10.1086/597614

Wang, J., Shen, T., Huang, X., Kumar, G.R., Chen, X., Zeng, Z., Zhang, R., Chen, R., Li, T., Zhang, T., Yuan, Q., Li, P.-C., Huang, Q., Colonno, R., Jia, J., Hou, J., McCrae, M.A., Gao, Z., Ren, H., Xia, N., Zhuang, H., Lu, F., 2016. Serum hepatitis B virus RNA is encapsidated pregenome RNA that may be associated with persistence of viral infection and rebound. Journal of Hepatology 65, 700–710. https://doi.org/10.1016/j.jhep.2016.05.029

Washington, S.D., Musarrat, F., Ertel, M.K., Backes, G.L., Neumann, D.M., 2018. CTCF Binding Sites in the Herpes Simplex Virus 1 Genome Display Site-Specific CTCF Occupation, Protein Recruitment, and Insulator Function. J Virol 92, e00156-18. https://doi.org/10.1128/JVI.00156-18

Watanabe, T., Sorensen, E.M., Naito, A., Schott, M., Kim, S., Ahlquist, P., 2007. Involvement of host cellular multivesicular body functions in hepatitis B virus budding. Proceedings of the National Academy of Sciences 104, 10205–10210. https://doi.org/10.1073/pnas.0704000104

Wei, X., Peterson, D.L., 1996. Expression, Purification, and Characterization of an Active RNase H Domain of the Hepatitis B Viral Polymerase. Journal of Biological Chemistry 271, 32617–32622. https://doi.org/10.1074/jbc.271.51.32617

Weiss, E.R., Göttlinger, H., 2011. The Role of Cellular Factors in Promoting HIV Budding. Journal of Molecular Biology 410, 525–533. https://doi.org/10.1016/j.jmb.2011.04.055

Wenyu Cheng, Guohua Chen, Huaijie Jia, Xiaobing He, Zhizhong Jing, 2018a. DDX5 RNA Helicases: Emerging Roles in Viral Infection. International Journal of Molecular Sciences 19, 1122. https://doi.org/10.3390/ijms19041122

Wenyu Cheng, Guohua Chen, Huaijie Jia, Xiaobing He, Zhizhong Jing, 2018b. DDX5 RNA Helicases: Emerging Roles in Viral Infection. IJMS 19, 1122. https://doi.org/10.3390/ijms19041122

Wilson, B.J., Bates, G.J., Nicol, S.M., Gregory, D.J., Perkins, N.D., Fuller-Pace, F.V., 2004. [No title found]. BMC Mol Biol 5, 11. https://doi.org/10.1186/1471-2199-5-11

Wilusz, J.E., 2016. Long noncoding RNAs: Re-writing dogmas of RNA processing and stability. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1859, 128–138. https://doi.org/10.1016/j.bbagrm.2015.06.003

Wing, P.A., Davenne, T., Wettengel, J., Lai, A.G., Zhuang, X., Chakraborty, A., D'Arienzo, V., Kramer, C., Ko, C., Harris, J.M., Schreiner, S., Higgs, M., Roessler, S., Parish, J.L., Protzer, U., Balfe, P., Rehwinkel, J., McKeating, J.A., 2019. A dual role for SAMHD1 in regulating HBV cccDNA and RT-dependent particle genesis. Life Science Alliance 2, e201900355. https://doi.org/10.26508/lsa.201900355

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

Wu, H.-L., n.d. Characterization and Genetic Analysis of Alternatively Spliced Transcripts of Hepatitis B Virus in Infected Human Liver Tissues and Transfected HepG2 Cells. J. VIROL. 7.

Xie, Y., Ren, Y., 2019. Mechanisms of nuclear mRNA export: A structural perspective. Traffic 20, 829–840. https://doi.org/10.1111/tra.12691

Xing, Z., Ma, W.K., Tran, E.J., 2019a. The DDX5/Dbp2 subfamily of DEAD-box RNA helicases. Wiley Interdisciplinary Reviews: RNA 10, e1519. https://doi.org/10.1002/wrna.1519

Xu, Z., Bruss, V., Yen, T.S., 1997. Formation of intracellular particles by hepatitis B virus large surface protein. Journal of virology 71, 5487–5494. https://doi.org/10.1128/JVI.71.7.5487-5494.1997

Yan, H., Zhong, G., Xu, G., He, W., Jing, Z., Gao, Z., Huang, Y., Qi, Y., Peng, B., Wang, H., Fu, L., Song, M., Chen, P., Gao, W., Ren, B., Sun, Y., Cai, T., Feng, X., Sui, J., Li, W., 2012a. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. eLife 1. https://doi.org/10.7554/eLife.00049

Yang, C.-C., Huang, E.-Y., Li, H.-C., Su, P.-Y., Shih, C., 2014. Nuclear Export of Human Hepatitis B Virus Core Protein and Pregenomic RNA Depends on the Cellular NXF1-p15 Machinery. PLoS ONE 9, e106683. https://doi.org/10.1371/journal.pone.0106683

Yang, H.-C., Kao, J.-H., 2014. Persistence of hepatitis B virus covalently closed circular DNA in hepatocytes: molecular mechanisms and clinical significance. Emerg Microbes Infect 3, e64. https://doi.org/10.1038/emi.2014.64

Yang, L., Duff, M.O., Graveley, B.R., Carmichael, G.G., Chen, L.-L., 2011. Genomewide characterization of non-polyadenylated RNAs. Genome Biol 12, R16. https://doi.org/10.1186/gb-2011-12-2-r16

Yang, W., Summers, J., 1999. Integration of Hepadnavirus DNA in Infected Liver: Evidence for a Linear Precursor. Journal of Virology 73, 9710–9717. https://doi.org/10.1128/JVI.73.12.9710-9717.1999

Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R.D., Felsenfeld, G., 2010a. Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. Genes & Development 24, 2543–2555. https://doi.org/10.1101/gad.1967810

Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R.D., Felsenfeld, G., 2010b. Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. Genes & Development 24, 2543–2555. https://doi.org/10.1101/gad.1967810

Yao, Hongjie, Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R.D., Felsenfeld, G., 2010. Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. Genes Dev. 24, 2543–2555. https://doi.org/10.1101/gad.1967810

Yu, M.-W., Yeh, S.-H., Chen, P.-J., Liaw, Y.-F., Lin, C.-L., Liu, C.-J., Shih, W.-L., Kao, J.-H., Chen, D.-S., Chen, C.-J., 2005. Hepatitis B Virus Genotype and DNA Level and Hepatocellular Carcinoma: A Prospective Study in Men. JNCI: Journal of the National Cancer Institute 97, 265–272. https://doi.org/10.1093/jnci/dji043

Yu, X., Mertz, J.E., 1996. Promoters for Synthesis of the Pre-C and Pregenomic mRNAs of Human Hepatitis B Virus Are Genetically Distinct and Differentially Regulated. J. VIROL. 70, 8.

Zhang, H., Rigo, F., Martinson, H.G., 2015. Poly(A) Signal-Dependent Transcription Termination Occurs through a Conformational Change Mechanism that Does Not Require Cleavage at the Poly(A) Site. Molecular Cell 59, 437–448. https://doi.org/10.1016/j.molcel.2015.06.008

Zhang, H., Xing, Z., Mani, S.K.K., Bancel, B., Durantel, D., Zoulim, F., Tran, E.J., Merle, P., Andrisani, O., 2016a. RNA helicase DEAD box protein 5 regulates Polycomb repressive complex 2/Hox transcript antisense intergenic RNA function in hepatitis B virus infection and hepatocarcinogenesis: HEPATOLOGY, Vol. XX, No. X, 2016 ZHANG ET AL. Hepatology 64, 1033–1048. https://doi.org/10.1002/hep.28698

Zhao, Y., Garcia, B.A., 2015. Comprehensive Catalog of Currently Documented Histone Modifications. Cold Spring Harbor Perspectives in Biology 7, a025064. https://doi.org/10.1101/cshperspect.a025064

Zheng, Y., Li, J., Ou, J. -h., 2004. Regulation of Hepatitis B Virus Core Promoter by Transcription Factors HNF1 and HNF4 and the Viral X Protein. Journal of Virology 78, 6908–6914. https://doi.org/10.1128/JVI.78.13.6908-6914.2004

Ziebarth, J.D., Bhattacharya, A., Cui, Y., 2013. CTCFBSDB 2.0: a database for CTCF-binding sites and genome organization. Nucleic Acids Res. 41, D188-194. https://doi.org/10.1093/nar/gks1165

Zoulim, F., Saputelli, J., Seeger, C., n.d. Woodchuck Hepatitis Virus X Protein Is Required for Viral Infection In Vivo 5.

Zoulim, F., Seeger, C., 1994. Reverse Transcription in Hepatitis B Viruses Is Primed by a Tyrosine Residue of the Polymerase 68, 8.

Acknowledgements

Acknowledgements

I would like to greatly thank **Pr. Michael Nassal** and **Dr. Chistine Neuveut** for having accepted to review my thesis manuscript and for their helpful comments and feedbacks. I thank **Pr. Massimo Levrero** and **Dr. Perèz-del-Pulgar** for having accepted my invitation to participate to my thesis defense. Finally, I thank **Dr. Cyril Bourgeois** for having followed my work throughout these four years by participating to my thesis evaluation comitees and for his essential observations regarding my thesis projects.

Je souhaite par ailleurs remercier le **Pr. Fabien Zoulim** et le **Dr. Barbara Testoni** pour m'avoir fait confiance et m'avoir accueillie dans leur équipe. Je vous remercie d'avoir toujours pris le temps de discuter de mes résultats, et de m'avoir offert un cadre idéal pour réaliser cette thèse, avec un accès illimité à toutes les ressources dont j'ai eu besoin pour mener ce projet à bien, aussi bien professionnellement que personnellement.

Barbara, merci de m'avoir fait confiance pour cette thèse bien que tu m'aies retrouvée après seulement 2 semaines de stage de M2 en train de faire un p'tit somme devant ton projet.. Merci d'avoir toujours été derrière la porte de ton bureau que j'ai poussée bien trop souvent en pleurant ! Merci de m'avoir toujours reboostée quand j'en avais besoin et d'avoir pris le temps, toujours, de discuter du projet. Merci d'avoir accepté que je dise toujours ce qu'il me passe par la tête, même si c'est parfois farfelu, et de toujours avoir pris ça avec humour et bienveillance ! Merci d'avoir cette étincelle dans les yeux quand tu parles de science et que les informations dansent dans ta tête ! Ne perds jamais cette passion qui fait de toi une incroyable chercheuse ! Enfin, merci d'avoir cru en moi et d'avoir mis toute ton énergie au service de ce projet. Je suis très fière d'avoir été ta troisième bébé thésarde et je n'oublierai jamais ces 4 années !

Guillaume, merci de toujours dire « oui » pour la pause-café, qui en réalité veut toujours dire « heeeelp ». Merci de tout mettre en pause quand j'ai besoin de ton aide. Merci d'avoir donné un second souffle à ce projet et de nous avoir amené ton expertise. J'ai beaucoup appris cette dernière année grâce à toi et je t'en serai toujours reconnaissante ! Reste toujours un aussi bon encadrant pour les stagiaires et un incroyable grand frère pour les thésards. Les futures

recrues auront de la chance de t'avoir !! Je te souhaite tout le meilleur pour la vie ! Juste une dernière requête : deviens un super chef d'équipe, et embauche moi quand je reviendrai en France !

Guada, so many things to say! Thank you for being so microscopy enthusiastic! Thank you for always being nice, trustful, encouraging. Thank you for bringing love and colors in our lives. I would love to learn organization and perfect structure of daily life from you! Thank you for having helped me finding a post-doc, reading parts of my introduction, and draft of the paper. Thank you for giving me all the time you can to make my work and life better! And to share advices and tips! Muchas gracias para el dulce de leche et para los empenadas. Estoy realmente afortunada tenerte en mi vida !

Judith, aaaaah Judith ! Que dire ! Merci d'avoir toujours des tours dans tes tiroirs ! D'avoir toujours une idée, un protocole, une enzyme qui traine dans un congel, un carreau de chocolat, pour nous sauver la vie ! Tu es un peu le druide de cette équipe !! Merci d'avoir toujours pris le temps de répondre à mes questionnements, et de faire les mises au point des primers (chose que je déteste le plus au monde) et des anticorps, afin que nous ayons du matériel au top pour ce projet ! Merci d'être cette maman ours au grand cœur. Merci pour ta bienveillance. Tu vas me manquer, beaucoup.

Audrey, mon hibou. Merci d'être toi. Merci pour ces nombreuses soirées « juste un p'tit verre » qui finissaient à la fermeture du bar avec bien trop de mojitos dans le gosier. Merci d'avoir été là pour moi quand j'en avais besoin. J'espère te revoir souvent à Prague, aux US, à Lyon, ou n'importe où ailleurs pour se boire une p'tite binouse. Je n'aurais qu'un seul regret, celui de n'avoir jamais goutté tes fameux « petits pois sauce au roux ».

Maelle, que dire. Merci de m'avoir prise sous ton aile dès mon arrivée au CAT ce 15 février 2016. Merci de m'avoir intégrée à l'équipe et de m'avoir encadrée et appris les fondamentaux d'HBV lors de mon stage, et de ma première année de thèse aussi.. et encore après aussi ! Merci d'avoir toujours été patiente avec moi, d'avoir répondu au téléphone 35 fois par jours lorsque j'avais besoin de toi. Merci d'encore toujours répondre au téléphone 4 ans plus tard quand j'ai besoin de toi. Merci d'avoir toujours les mots pour me mettre des coups de pieds aux fesses, merci de toujours m'aider à gérer mes états émotionnels en montagne russe ! Merci de toujours m'apporter ton input intellectuel, peu importe le projet, professionnel ou

personnel. Merci de partager mes différentes passions, et de m'avoir fait découvrir les tiennes. Merci de ne m'avoir jamais laissée tomber, peu importe la situation. Tu es mon pilier. Je n'ai qu'une hâte, te revoir très vite en NC.

Merci à tous ceux qui ont fait du CAT un environnement incroyable ces dernières années :

Anaelle, merci de m'avoir offert un toit pour mes deux derniers mois de rédaction et d'avoir fait de cette période peu ordinaire des semaines fort agréables ! Merci de m'avoir supportée, et tendu la boîte de mouchoirs pour sécher mes larmes à chaque fois qu'on regardait « This is us ». Je ne te remercie pas de m'avoir quasiment rendue accros au chocolat, Schokobons, et Nutella.. Par contre je suis fière que tu aies appris à manger autant de légumes ! Merci à Nala d'avoir fait faire de l'exercice à Chatchou ! Il est presque svelte maintenant ! Je suis heureuse qu'on ait pu apprendre à se connaître plus durant ces dernières semaines et je te serai toujours reconnaissante de m'avoir accueillie sans réfléchir ! J'espère que les plantes poussent bien (à vue d'œil ?!) et que le balcon restera toujours aussi fleurit ! **Caroline P.**, merci de m'avoir toujours laissée jeter un œil à tes protocoles (voire carrément les photocopier) quand j'en avais besoin ! Merci pour ta gentillesse et de toujours prendre le temps pour répondre à nos questions. **Caroline C.**, à la plus farfelue des virologistes. Merci d'être toujours à 300%. Merci pour ta franchise et ton énergie. Merci d'être un génie de la science. Sans toi toute cette histoire de Nanopore n'en serait pas là où elle en est. Merci de nous aider à chaque fois qu'il t'en est possible et de faire preuve d'une détermination infaillible. Merci d'être toi. Ne change jamais. **Chloé**, thank you for being so you! Thank you for bringing us the Australian sun and energy! And of course, thank you for your help with the analysis! **Damien**, merci de faire découvrir tes p'tits plats aux saveurs asiatiques tous les midis ! Merci de toujours avoir une blague ou un mot pour rire ! Enfin, merci d'être un écolo implacable ! **Delphine**, merci d'être un savant mélange de geek, insta-addict, scientifique passionnée et défenseuse de ses idées ! Merci d'avoir toujours un mot rassurant et le temps de discuter. HBV a un joli avenir avec toi je n'en doute pas ! Et Barbara est chanceuse de t'avoir comme quatrième bébé thésarde ! Je suis sûre que tu parviendras à tout ce que tu veux dans la vie. Ta détermination et ton énergie t'amèneront loin. Ne perds pas cette force ! Et relax avec les stagiaires ! **Doohyun & Hyoseon**, thank you for being the kindest people ever! Thank you for always taking the time to discuss and share food specialties with us! Thank you also for always being smiling and caring to other people. **Héloise**, merci d'avoir cette énergie débordante qui te caractérise tant ! Merci de

propager ton rire dans les couloirs du CAT et à travers les murs de l'annexe ! Merci d'avoir toujours un mot gentil et une attention bienveillante ! **Manon**, ravie d'avoir partagé ce bureau de fin de thèse avec toi les dernières semaines. Je suis sûre qu'on aurait appris à mieux se connaître si on avait eu le temps de partager un peu plus de temps au labo, plutôt qu'en confinement ! Merci pour ton calme et ta gentillesse. **Isabelle B.**, merci de toujours, absolument toujours répondre à mes questions en prenant le temps et suivre mes délires écolos de poubelle liquide et poubelle de tri ! Merci pour ta disponibilité et pour tout ce que tu fais pour le labo. **Françoise B.**, merci d'être toujours gentille et bienveillante. A vous deux : ce sera pour sûr bizarre dans mon futur labo, d'être à ma future paillasse, de regarder à ma droite, et de ne pas vous voir.. Car oui, finalement c'est vous que j'aurais le plus vues tous les jours de cette thèse depuis ma paillasse ! **Julie**, tu sais que mes taquineries sont justes des signes de grande amitié ! Merci de toujours exposer les côtés positifs des joies du métier et de m'avoir donné l'envie de devenir une chercheuse passionnée ! Merci de nous ouvrir ta porte toutes les semaines pour que nous partagions ensemble de trop nombreuses coupes de prosecco en jouant au Tock, au Haricot, ou seulement en papotant. Merci de faire la meilleure des sauces salade ! Merci d'être toujours disponible et encourageante. Merci de nous offrir des moments inoubliables d'illusion d'optique ! **Manu**, l'équipe a de la chance de t'avoir recruté ! Merci d'être toujours d'un calme absolu et de bons conseils. Merci pour ta gentillesse et ta bienveillance ! **Marion**, COUCOUUUUUUUUUUUUUUUUUUUUUUU !!! Ici le CORM ! Ralalalalaaah Marion, grâce à toi la transition entre les « anciens du labo » (j'y reviendrai) et les « nouveaux du labo » a été bien moins compliquée que ce que je le pensais ! Merci d'être absolument toujours de bonne humeur ! Merci de me faire des bons p'tits plats, je pense notamment à l'épautrotto.. et de partager la passion de la nourriture avec moi ! Merci d'être toujours de bons conseils et motivée pour tout ! Ne perds jamais ton énergie ! **Maud**, merci de m'avoir appris le Northern-Blot et la bonne humeur !! Merci de faire de chaque jour un nouveau défilé de mode coloré ! Dans la vie, tout le monde a besoin d'une Maud ! Maud, Mauuuuuuuuuuud, simplement on t'dit, qu'on t'aime, t'aiiiiiiiiiiiiiiiiiiiime ! **Marie-Agnès**, MERCI de gérer absolument TOUTE la partie administrative en URGENCE (notamment ces dernières semaines où tu m'as bien sauvée la mise !). Merci d'avoir pris du temps pour moi à chaque fois que j'en ai eu besoin !! Sans toi, il y a bien longtemps que je me serais perdue dans des tas de papiers, et de documents incompréhensibles, je ne serais surement pas allée aux congrès pour cause d'oubli d'Ordre de mission, je n'aurais surement pas pu imprimer et

envoyer ma thèse, etc, etc… alors merci pour tout !! **Marie-Laure**, merci d'être d'une pédagogie infaillible quand il s'agit de montrer des manips (même si c'est pour me montrer 4 fois comment fonctionne le BioAnalyzer..). Merci d'être toujours franche et honnête (même s'il m'a fallu un p'tit temps pour m'y faire au début !). Merci de toujours prendre le temps pour aider et d'avoir un sens de l'humour et de l'autodérision à toute épreuve. **Olga**, thank you for having me got the chance to discover the person you are! Do not worry, protocols work! Let your stress go sometimes :). **Romain**, Rominou. Merci d'avoir toujours été là pour moi, du tout début à la toute fin de cette thèse. Merci d'avoir toujours des blagues à raconter. Merci pour ton calme légendaire et ta bienveillance. Merci pour les parties de squash ! Pour le Loiret où on passe toujours 3 jours de pur bonheur hors du temps. Merci de nous rincer au champagne et d'être un binôme de Tock de choc ! Belle vie en Allemagne, j'espère qu'on se verra souvent, d'un côté ou de l'autre de l'Atlantique !

Je remercie également tous les « anciens du labo », sans qui cette thèse n'aurait pas été une expérience personnelle incroyable, jonchée de trop nombreuses bières, mojitos, festivals, soirées endiablées et parties de rire.

Aurore, ma p'tite Aurore ! Merci d'avoir fait de ce début de thèse une période incroyable ! Merci d'être toujours motivée pour faire 1000 choses, merci de me motiver à aller courir, merci pour nos nombreuses soirées balcon / salade (parce que oui, on se donne bonne conscience !). Merci pour ton aide précieuse de ces derniers jours pré-soutenance et pour ton soutien infaillible. N'arrête jamais de te déchainer sur du N'to, n'arrête jamais de préparer des bons Gin-To. Je sais que tu vas réussir de belles choses dans la vie ! Et n'oublie pas de te faire exorciser ! **Brieux**, Bribri. Merci d'avoir été mon co-thésard pendant trois ans. Tu feras toujours partie des personnes importantes de ces dernières années, pendant lesquelles on aura grandi côte à côte au CAT. **Dulce**, même si nous n'avons jamais été au labo en même temps (ce que je regrette beaucoup), j'ai eu la chance de te connaître grâce aux autres ! Merci de toujours avoir un mot gentil qui fait du bien, qui rassure et qui apaise. Merci d'avoir toujours pris le temps de répondre à mes questions d'hypochondriaque.. (oups) et de ne m'avoir jamais jugée pour ça. **Emma**, merci de m'avoir initiée aux joies de l'enseignement et d'avoir toujours été là pour répondre à mes questions d'ordre TP/TDesques !! Merci aussi pour ta gentillesse. **Lea**, ma p'tite Léa ! Quelle joie d'avoir partagé ce bureau avec toi ces derniers mois, en tête à tête. Merci d'avoir toujours une anecdote sur tes neveux à raconter.

Merci d'être toujours bienveillante et attentive ! Merci pour ton honnêteté et ton naturel. Merci pour tes bonnes miches de pain et pour notre passion commune du fromage ! Merci pour la trotti qui m'aura permis de fuser à travers les rues de Monplaisir et ne presque plus être en retard ! **Loïc**, merci d'être le plus déjanté des humains imberbes de cette planète ! Merci d'avoir été un super coloc pendant deux ans. Merci d'être toujours attentif et bienveillant ! Merci d'avoir partagé tous ces bons moments, entre bars, concerts, festivals, randonnées sur les Volcans réunionnais et aménagement chaotique avec pour table un carton retournée dans le salon et un mini frigo sur le balcon ! C'aura été le meilleur aménagement de ma vie ! **Ludo**, mon champion n°1 !! Merci de m'avoir laissé ton titre et m'avoir fait gravir les échelons de vulgaire « 4 bis » à « Champion 1 » en titre ! Merci d'être génial et aussi drôle ! Merci pour tous ces p'tits rhum et toutes ces bières. Merci pour les parties de squash et ton frappé de balle nonchalant que seul toi maitrise à la perfection ! On se voit bientôt pour boire des pintes Pragoises ! **Laura**, mamie Lolo !!! Merci pour ces années où j'ai toujours pu compter sur toi ! Merci pour ta bonne humeur et ton rire ! Tu vas être une super maman ! **Petite Nat**, merci d'être toi ! Merci d'avoir été un rayon de soleil dans le labo ces dernières années ! Merci pour ta grande gentillesse et tes compliments quotidiens qui faisaient du bien au cœur ! **Susu**, sacrée Susu ! Je me rappellerai toujours mon premier jour au CAT, où j'ai carrément eu peur de toi (je pense que tu étais énervée ce jour-là !). Et puis j'ai très vite appris que derrière ce charisme incroyable, il y avait un immense cœur. Merci ma Susu d'avoir fait de mes premières années de thèse des années inoubliables. Merci pour ta franchise, et pour ton combat dans tes idées ! Reste toujours droite comme tu sais l'être, ça t'amènera très loin. Enfin, merci pour ton amitié ! **Thomas et Adrien**, merci d'avoir toujours pris le temps de répondre à mes questions « HBc » et d'avoir partagé avec moi votre expertise !

A tous les autres **Annexiens** et à toute l'**U1052**, merci d'avoir chacun participé à faire de cette thèse une expérience inoubliable ! Merci d'avoir répondu à mes questionnements, car je suis presque sûre que vous êtes tous passés sous mes questions au moins une fois ces dernières années ! Merci pour vos conseils et votre présence !

Enfin, merci à tous ceux qui sont là pour moi depuis des années.

Christa & Gloria, mes amies depuis l'enfance, avec qui on est toujours aux trois opposés de la planète mais avec qui on sera toujours proches par le cœur. Merci pour tous les moments communs de notre vie et pour tous les rires que nous avons partagés. Merci aussi pour tous

les moments heureux que nous partagerons encore. **Victoria**, merci d'être toi. Merci d'avoir été là à mes côtés depuis mon arrivée à Lyon et d'avoir été ma coloc de plus longue date. Ta présence à mes côtés ces dernières années en ont fait des années inoubliables. Merci pour ta gentillesse, et ta bienveillance. Merci de m'avoir appris qu'il y a toujours une explication à l'aigreur des gens et qu'il est important d'être compréhensif dans la vie lorsqu'on veut être quelqu'un de bien. **Inès**, toi qui es là depuis la 1ère année de licence, avec qui j'ai partagé la vie à Montpellier, puis la vie à Lyon. Merci d'avoir toujours été là à mes côtés, je sais qu'on n'est pas très câlins, mais ça n'entrave en rien ton amitié précieuse. Reste toujours la personne droite et honnête que tu es, et que j'admire. **Alice**, merci d'être là depuis tant d'années. Merci d'avoir partagé avec moi des moments forts, des moments tristes, des moments heureux, merci de m'avoir dit les choses que tu devais me dire quand il le fallait. Merci d'avoir toujours été franche. Merci de te battre pour tes convictions, d'être une femme exceptionnelle et une amie incroyable. **Anaïs, Anouk, Arthur**, merci de m'avoir accueillie à bras ouverts à l'ENS. Merci d'avoir toujours été là pour moi ces dernières années, peu importe la distance, peu importe le pays ! Merci Anaïs d'avoir partagé ces samedis rédaction qui faisaient du travail des moments toujours agréables ! Merci pour les nombreuses parties de « Times-Up », pour les « Scrabbles », pour les festivals à NYC. Merci Anouk pour tes desserts à tomber par terre et pour ton immense bienveillance, et merci Arthur pour tes conseils en mode implacables, hâte de te retrouver de l'autre côté de l'Atlantique ! Merci à vous trois d'être géniaux. Merci pour votre amitié. J'espère qu'on se reverra très vite tous les quatre pour nous construire encore de beaux souvenirs ensemble !

Pour terminer, je remercie ma petite famille. Dans un premier temps mes sœurs. **Carine**, merci de m'avoir ouvert la voie aux grandes études et d'avoir toujours été un modèle de sérieux et d'énergie ! **Céline**, merci d'être toujours de bonne humeur et partante pour tout ! Je n'aurai jamais assez d'énergie pour te suivre jusqu'au bout de la nuit ! Tu m'épateras toujours ! Merci à vous deux de m'avoir offert quatre super neveux qui ont bien de la chance de vous avoir comme mamans ! Merci à **mes parents**, qui ont toujours cru en moi et qui m'ont toujours soutenue dans tous mes choix. Merci à vous de m'avoir permis d'aller aussi loin que ce que je voulais, sans ne jamais y mettre de condition, sans ne jamais compter ni rien demander en échange. Merci de votre éducation qui fait de moi celle que je suis aujourd'hui. Merci de m'avoir donné votre confiance. Merci de m'avoir donné tout votre amour. Sans votre soutien

et votre présence, je n'aurais jamais réussi ce qui me tenait tant à cœur. Finalement, une pensée émue pour ma petite **Mamie**, qui serait fière de la petite Fleur. Merci de m'avoir donné tout ton amour, de m'avoir préparé des bons plats de pâtes à la tomate (du jardin !), et des tartes aux pommes toujours un peu brûlées qui en faisaient des délices !

Appendices

Appendices:

I. Chromatin Proteomics of cccDNA in HepG2-NTCP and PHH

Article in preparation

Landscape of hepatitis B virus minichromosome-human protein complexes functional interactions

Barbara Testoni[1,2*], David W. Avila[3], Jitao David Zhang[3], Alan James Mueller-Breckenridge[3], Sarah Maadadi[1], Johanna Walther[3], Fleur Chapus[1,2], Maëlle Locatelli[1,2], Hélène Meistermann[3], Audrey Diederichs[1], Judith Fresquet[1], Amel Neila Aberkane[1,2], Christoph Bieniossek[3], John Young[3], Souphalone Luangsay[3] and Fabien Zoulim[1,2,4*]

[1] INSERM U1052, CNRS UMR-5286, Cancer Research Center of Lyon (CRCL), Lyon, 69008, France;

[2] University of Lyon, Université Claude-Bernard (UCBL), 69008 Lyon, France;

[3] Hoffmann-La Roche Ltd, Roche Pharmaceutical Research and Early Development, Roche Innovation Center Basel, 4070 Basel, Switzerland;

[4]Hospices Civils de Lyon (HCL), 69002 Lyon, France.

*Correspondence: barbara.testoni@inserm.fr; fabien.zoulim@inserm.fr

Summary

Covalently-closed-circular (ccc)DNA accounts for the persistence of hepatitis B virus (HBV) infection. cccDNA stably resides in the nucleus of infected cells, unaffected by current antiviral therapies. Histones, HBV core protein and cccDNA are thought to be closely linked together to build a chromatin structure which serves as unique template for viral transcription.

We characterized the cccDNA-associated interactome combining cccDNA-specific Chromatin Immunoprecipitation to high-throughput Mass Spectrometry analysis in infected human hepatocytes. We identified a number of host complexes involved in nuclear matrix localization, transcriptional regulation and RNA processing/export hijacked by HBV for its replication, including the fused in sarcoma (FUS) protein, which was shown to be a master regulator of HBV lifecycle, by coupling viral transcription to RNA processing and export.

Our results provide a novel picture of host proteins specifically associated with the HBV genome and contribute an understanding of how the host machinery is manipulated to allow viral persistence and replication.

Keywords: HBV, cccDNA, HBc, chromatin, proteomics, transcription, RNA processing

Introduction

Human hepatitis B virus (HBV) genome persists in the nucleus of infected hepatocytes as a small episome, called covalently-closed-circular (ccc)DNA, and relies on both viral and host cellular machinery for its replication[1]. Since HBV genome arrives in the hepatocytes nucleus as a partially double stranded DNA (the so-called relaxed circular DNA, rcDNA), the host factors are required to repair and convert rcDNA into cccDNA, and to deposit histones and non-histone proteins to build a viral chromatin structure that serves as the unique template for viral RNAs synthesis. HBV genome replication involves the transcription of a longer-than-genome RNA (pre-genomic RNA, pgRNA) which is subsequently retrotranscribed into rcDNA, which is released from infected cells as progeny virions. Current available therapies target the HBV polymerase enzymatic activity, thus preventing pgRNA retrotranscription in the cytoplasm, but leave the nuclear cccDNA pool largely unaffected[2,3]. cccDNA persistence in the infected hepatocytes accounts for the chronicity of HBV infection and viral rebound after therapy withdrawal[4]. The chromatinized nuclear cccDNA consists, most likely, of the dynamic equilibrium of molecules with different levels of nucleosomes association, as demonstrated in the duck HBV model[5], translating into different levels of chromatin accessibility and biological activity[6,7]. The HBV core (HBc) protein is a structural component that associates to cccDNA, resulting in a rearrangement of the nucleosomal spacing of the HBV nucleoprotein complexes[7,8] and in alteration of cccDNA methylation profile[9]. Experiments using an adapted Chromatin Immunoprecipitation (ChIP) assay[10–14] and ChIP-Sequencing[15,16] approach demonstrated the recruitment to the HBV minichromosome of H3 and H4 histones, as well as several cellular and viral transcription factors and histone-modifying enzymes, which have been associated with modulation of cccDNA transcription[17].

Therefore, identifying the landscape of host proteins and complexes associated with the viral genome in the context of infected hepatocytes is crucial for the understanding of how HBV twists the host cellular machinery during the course of infection.

The application of proteomic approaches to virology studies has been increasing in the last decade, particularly focusing on the interactions between viral and host proteins or how viral infections affects host cell proteome[18]. The investigation of proteins associated with a viral minichromosome was performed to profile the histone post-translational modifications (PTMs) associated to the polyomavirus BK genome in virions[19]. In addition to localizing histone PTMs across the genome, pioneering studies coupling chromatin immunoprecipitation (ChIP) and mass-spectrometry analysis (ChIP-MS) have been conducted in yeast and Drosophila to elucidate the binding of transcription factors, coregulatory proteins and other chromatin components characterizing distinct chromatin regions without *a priori* assumptions[20].

Here, we developed a cccDNA-chromatin proteomics approach (cccDNA-ChroP) to detect host protein complexes specifically associated with the chromatinized HBV episome in the nucleus of infected transformed liver cells (HepG2-NTCP) and primary human hepatocytes. We identified a panel of 363 proteins associated with the viral minichromosome in HBV-infected cells. Gene Set Enrichment Analysis highlighted major differences in cccDNA-interactome composition between transformed and primary hepatocytes, but also revealed a common *core* network centered around the fused in sarcoma (FUS) protein. Suppression of FUS protein expression severely affects HBV RNA production and modifies cccDNA-associated RNA Polymerase II processivity and recruitment of proteins belonging to the RNA transcription/export complex (TREX), thus providing evidence on how HBV hijacks the host machinery to couple viral transcriptional regulation and RNA processing/export.

Experimental procedures

Production of HBV viral inoculum

HBV inoculum was prepared from filtered HepAD38 cells[21] supernatants by polyethylene-glycol-MW-8000 (PEG8000, SIGMA) precipitation (8% final) as previously described[22]. Viral stock with a titer reaching at least $1x10^{10}$ vge/mL was tested endotoxin free and used for infection.

Cell culture, HBV infection and siRNA transfection

HepG2-NTCP cells were seeded at 10^5 cells/cm2 in DMEM medium supplemented with penicillin (Life Technology), streptomycin (Life Technology), sodium pyruvate (Life Technology), 5% Fetal Calf Serum (FCS; Fetalclone IITM, PERBIO). The day after, medium was renewed and complemented with 2.5% DMSO (SIGMA). After 72h, cells were infected at a multiplicity of infection of 250 in the presence of 4% PEG800 for up to 16h and then extensively washed and cultured for the indicated time points in complete DMEM medium containing 2.5% DMSO until harvesting. Intracellular accumulation of viral RNA and DNA were monitored by RT-qPCR, qPCR and Southern Blotting.

siRNA against specific targets (Table 1) and the siRNA negative control (ON-TARGETplus non-targeting pool Catalog #D-001810, Dharmacon) were transfected at 10nM concentration in HepG2-NTCP cells and PHHs using Lipofectamine RNAiMAX reagent (ThermoFisher), following manufacturer's instructions. Two sequential transfections were performed (at 4 days and 6 days post-HBV infection) to maximize target expression inhibition. Cells were harvested 72h after the last transfection for subsequent analysis.

Target	Company
DDB1	
ALYREF	Dharmacon SMARTpool ON-TARGETplus L-012078-00
DDX39	Dharmacon SMARTpool ON-TARGETplus L-003805-00
LGALS3BP	Dharmacon SMARTpool ON-TARGETplus L-008016-02
OTUB1	Dharmacon SMARTpool ON-TARGETplus L-021061-00
PSMC6	Dharmacon SMARTpool ON-TARGETplus L-009570-01
SNRPD1	Dharmacon SMARTpool ON-TARGETplus L-012353-01
SRSF10	Dharmacon SMARTpool ON-TARGETplus L-012914-02
TLS/FUS	Dharmacon SMARTpool ON-TARGETplus L-009497-00

Table 1: siRNA list

Analysis of viral parameters during replication

Total DNA was purified from infected cells using MasterPure™ Complete DNA Purification Kit (Epicentre). Total HBV DNA was quantified using TaqMan® Gene Expression assay ID: Pa03453406_s1. To increase the specificity of HBV cccDNA detection, qPCR was preceded by a nuclease digestion using 10U of T5 exonuclease (Epicentre) for 500 ng of total DNA. Selective cccDNA qPCR was performed using primers and probes spanning the "gap region" in the HBV rcDNA (Table 2)[23,24]. Serial dilutions of an HBV monomer plasmid (pHBV-EcoRI) were used as standard for quantification. For normalization, the number of human hepatocytes was estimated by measuring human β-globin (TaqMan® assay ID Hs00758889_s1) while human genomic DNA (Roche Applied Science, Mannheim, Germany) was used as a standard curve for quantification. Total RNA was extracted from infected cells using NucleoSpin® RNA kit (Macherey-Nagel) and retro-transcribed using SuperScript III reverse transcriptase according to manufacturer's instructions (Invitrogen, Carlsbad, USA). 3.5 Kb viral RNA (comprising pgRNA and pre-core RNA) was amplified using previously described primers and probes (Table 2)[23,24]. Total HBV RNA was amplified using TaqMan® Gene Expression assay ID: Pa03453406_s1. The expression of the human housekeeping gene GUSB (TaqMan® assay ID Hs00939627_m1) was used for normalization. Real-time PCRs were performed using an Applied QuantStudio 7 machine and TaqMan Advanced Fast Master Mix or SYBR Green Master Mix (ThermoFischer).

Target	Name	Sequence
cccDNA	cccDNA Forward	CCGTGTGCACTTCGCTTCA
cccDNA	cccDNA Reverse	GCACAGCTTGGAGGCTTGA
cccDNA	cccDNA Taqman Probe	[6FAM]CATGGAGACCACCGTGAACGCCC[BBQ]
pgRNA	pgRNA Forward	GGAGTGTGGATTCGCACTCCT
pgRNA	pgRNA Reverse	AGATTGAGATCTTCTGCGAC
pgRNA	pgRNA Taqman Probe	[6FAM]AGGCAGGTCCCCTAGAAGAAGAACTCC[BBQ]

Table 2: Primers and probes sequences used for Taqman qPCR.

cccDNA-chromatin proteomics (ChroP) procedure

250x10^6 HepG2-NTCP cells or 50x10^6 PHHs at day 9 post HBV infection were used for each experiment. Cells were washed in phosphate buffered saline (PBS), and incubated for 15 minutes with 1% formaldehyde at 37°C and quenched with 0.125 M Glycine. For nuclear extracts preparation, cells were lysed 10 minutes on ice in lysis buffer (PIPES 5mM, KCl 85mM, NP-40 0,5%, PMSF 1mM, Protease inhibitor cocktail (PIC) 1X). After douncing (10 times) and centrifugation, nuclei were resuspended in Tris-EDTA buffer (10mM Tris-HCl pH 8, 1mM EDTA). 1/10 and 1/100 of nuclei resuspension was saved for Western Blotting and qPCR analysis, respectively. NaCl to a final concentration of 0.2M was added and incubated at room temperature for 15 minutes. After centrifugation at 2500 rpm for 10 minutes at room temperature, Hirt supernatant was collected and diluted ten times in ChIP RIPA buffer (10mM Tris-HCl pH 7,5, 140mM NaCl, 1mM EDTA, 0.5 mM EGTA, 1% Triton X100, 0.1% SDS, 0.1% Na-Deoxycholate) to be subjected to antibody immunoprecipitation overnight at 4°C using 100 µg of anti H3 antibody (Table 3). Immune complexes were, then, incubated for 2h with protein G agarose beads at 4 °C, washed, and eluted in 10mM DTT. The eluate was then diluted 10 times in ChIP RIPA buffer and used for immunoprecipitation overnight at 4°C with 100 µg of anti-HBc antibody (Table 3). Immune complexes were, then, incubated for 2h with protein G agarose beads at 4 °C, washed, and eluted in 50mM NaHCO3, 1% SDS to proceed to LC-MS analysis. 1/20 of the eluted immune complexes, together with saved aliquots from nuclei resuspension and Hirt supernatant were reversed crosslinked and, then, subjected either to proteinase K digestion and phenol/chloroform DNA extraction or to Western Blotting.

LC-MS procedure

Protein eluates from H3/HBc sequential ChIP were precipitated in 80% cold acetone for 90 min at -20°C. After centrifugation at 21'000xg for 10 min, protein pellets were re-suspended in 30 μL of SDS-PAGE sample buffer and heated at 90°C for 10 min. Samples were loaded on a Criterion 4-20% TGX gel (Bio-Rad) and allowed to migrate approximately 1 cm into the gel. Gels were stained with Colloidal Blue stain (Invitrogen) and each gel lane was cut into 4 equal slices then transferred to a 96-well MultiScreen-HTS filter plate (Millipore) for further processing. Gel slices were reduced with 50 mM DTT for 1 hour at room temperature and alkylated with 55 mM iodoacetamide for 30 min at room temperature in the dark. Gel slices were washed with 50 mM ammonium bicarbonate followed by overnight in-gel digestion with trypsin at room temperature. Peptides were extracted from the gel at high and low pH, then dried using a Speed-Vac. Samples were finally re-constituted in 5% formic acid/2% acetonitrile and analyzed by reverse-phase liquid chromatography-mass spectrometry (LC-MS) using an Easy-nLC 1000 with a 75 µm id x 50 cm C18 column coupled to Thermo Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific) in the data-dependent acquisition and positive ion mode. Raw LC-MS files were processed using Thermo Proteome Discoverer (version 2.1) and peak lists were searched through the Mascot search engine against the human Swiss-Prot database. The list of identified proteins was filtered using a False Discovery Rate (FDR) set at 1% for both peptides and proteins.

Chromatin immunoprecipitation (ChIP)

ChIP experiments were carried out at the indicated time points post-infection as previously described[25]. Briefly, cells were washed in phosphate buffered saline (PBS), and incubated for 15 minutes with 1% formaldehyde at 37°C and quenched with 0.125 M Glycine. For nuclear extracts preparation, cells were lysed in lysis buffer (PIPES 5mM, KCl 85mM, NP-40 0,5%, PMSF 1mM, Protease inhibitor cocktail (PIC) 1X). After douncing (10 times) and centrifugation, nuclei were resuspended in sonication buffer (SDS 1%, EDTA 10mM, Tris-HCl pH8 50mM, PMSF 1mM, PIC 1X). After sonication, chromatin was cleaned in ChIP RIPA buffer with Protein A Sepharose, and then subjected to overnight immunoprecipitation at 4 °C using 2–5 µg of

antibodies indicated in Table 3 or No Antibody. Immune complexes were, then, incubated for 2h with protein G agarose beads at 4 °C, washed, and eluted in Tris-HCl pH8 10mM, EDTA 5mM, NaCl 50mM, SDS 1%, Proteinase K 50μg, PIC 1X. Immunoprecipitated DNA was extracted and quantified by qPCR using cccDNA specific primers (see table 2). Samples were normalized to input DNA using the ΔCt method were ΔCt = Ct (input) – Ct (immunoprecipitation) and expressed as percentage of the input after normalization over No Antibody signal.

Target	Application	Company
HBc	ChIP and ChroP	Invitrogen SC2362651
H3pan	ChroP	Diagenode C15410324
RNA Polymerase II	ChIP	Diagenode C15200004
E2F-1 (KH95)	ChIP	Santa Cruz SC-251
ALYREF	ChIP	Abcam ab6141
BPIL1	ChIP	Abcam ab204804
DDX39	ChIP	Abcam ab50697
DHX15	ChIP	Abcam ab70454
EWSR1	ChIP	Invitrogen PA5-35366
FUSIP1	ChIP	Abcam ab77209
hnRNPH	ChIP	Abcam ab10374
hnRNPK	ChIP	Abcam ab39975
LGALS3BP	ChIP	Proteintech 10281-1-AP
NMT55	ChIP	Abcam ab70335
OTUB1	ChIP	Abcam ab175200
PSMC6	ChIP	Bethyl A303-824A-T
RALY	ChIP	Abcam ab170105
RENT1	ChIP	Abcam ab109363
SF3B3	ChIP	Abcam ab209402
SNRPD1	ChIP	Abcam ab236902
TARDBP	ChIP	Diagenode C15410266
TLS/FUS	ChIP	Abcam ab23439
TRA2B	ChIP	Abcam ab31353
RNA Polymerase II Ser 2p	ChIP	Diagenode C15200005
RNA Polymerase II Ser 5p	ChIP	Diagenode C15200004
H3K36me3	ChIP	Diagenode C15410192
beta-actin	WB	Abcam ab6276
HBc	WB	Dako B0586
Goat anti-mouse IgG (H+L) secondary antibody, HRP	WB	Invitrogen 31430
Goat anti-rabbit IgG (H+L) secondary antibody, HRP	WB	Invitrogen 31460

Table 3: Antibodies used in the study.

Western Blotting

Cells were lysed in RIPA buffer supplemented with PIC 1X and PMSF 1X. Proteins were migrated in 4-20% mini-PROTEAN$^{@}$ TGX stain-Free^{TM} Precast Gel (Bio-Rab Laboratories) and transferred onto a nitrocellulose membrane (Bio-Rab Laboratories). Blots were blocked 1 hour with 5% milk in TBS (1 x Tris Buffered Saline (Sigma)) and stained with primary antibody in blocking buffer overnight at 4 °C. After primary antibody incubation, blots were washed 3X with TBST (1× TBS with 0.1% Tween 20), stained with HRP-conjugated secondary antibodies for 1 hour at room temperature and washed again 3 times with TBST. Detection occurred using Biorad Clarity Western ECL and the ChemiDoc XRS system (Biorad). Antibodies are listed in Table 3.

Interaction graph

After removing keratin contaminants and converting to UniProt identifiers, the interaction network s was prepared using the Reactome functional interaction (FI) application (2016 annotation update) implemented within Cytoscape 3.4. Reactome pathway and biological process annotations (www.reactome.org) were determined by over-representation and FDR correction. Proteins within the original list that did not have interactions defined within the Reactome database were not included in the analysis.

Results

ChroP analysis reveals different landscapes of cccDNA-associated factors in transformed *vs* primary infected hepatocytes

To generate a comprehensive atlas of host proteins specifically associated with HBV cccDNA, we developed a technology coupling ChIP to MS profiling (called chromatin proteomics, cccDNA-ChroP) in infected hepatocytes. To specifically enrich for episomal chromatinized HBV DNA, we performed nuclei isolation followed by a modified Hirt extraction[26] (Figure 1A and Supplementary Figure 1). Sequential ChIP, first with an anti-histone H3 and then with an anti-HBc protein, allowed enrichment of the HBV genome organized around nucleosomes with respect to other viral DNA replicative intermediates or nuclear viral RNA (Figure 1A and Supplementary Figure 1). HBc is the structural component of the viral nucleocapsid that delivers the viral genome to the nucleoplasm during the early steps of infection and we found it associated to cccDNA as soon as the viral minichromosome is technically detectable in the nucleus of infected cells (Supplementary Figure 2). Moreover, anti-HBc ChIP was able to immunoprecipitate more than 70% of Hirt-extracted nuclear viral DNA across different time points in infected PHHs. This was confirmed in samples derived from liver resection of a chronically infected patient (Supplementary Figure 2B-C). After the sequential ChIP, eluates containing cccDNA-associated proteins were subjected to preparation for MS analysis, as detailed in Experimental Procedures section. ChroP analysis was performed 4 times in HBV-infected vs mock-infected HepG2-NTCP cells and on 5 different batches of HBV-infected vs mock-infected primary human hepatocytes (PHHs).

Analysis of HBc peptide coverage showed between 70% and 83% of HBc protein sequence coverage across ChroP experiments, confirming the efficiency of the methodology.

Raw LC-MS files were processed using Thermo Proteome Discoverer (version 2.1,vendor) and peak lists were searched through the Mascot search engine against the human Swiss-Prot database. The list of proteins identified by at least two different peptides was filtered using

a False Discovery Rate (FDR) set at 1% for both peptides and proteins. Only proteins identified in HBV-infected and not in mock-infected samples were considered for further analysis. After removing keratin contaminants and converting to UniProt identifiers, a total of 1588 proteins were detected in HBV-infected vs mock-infected samples (Supplementary Table 1 – *under embargo for pending patent*). 810 were found only in HepG2-NTCP cells, 511 only in PHHs, while 267 were detected in both cell types, corresponding to 17% of overall MS hits (Figure 1B).

By cross-referencing published interactions between single or small groups of proteins with cccDNA, we could identify among our cccDNA interactome RNA Polymerase II, transcription factors such as STATs[27], NF-kB[28] and transcription co-factors such as PRMT5[13], SIRT3[14], DDB1[29] and TARDBP[30], as already shown by others using different methods, even if they did not score in the top enriched cccDNA-associated factors identified by our approach. We also confirmed protein-cccDNA associations that were only investigated in non-infectious models, like HNRNPC/K[31,32], PARP1[33] and PUF60[34]. Independent cccDNA-ChIP experiments were performed to substantiate the recruitment of selected ChroP-identified proteins to cccDNA in both infected HepG2-NTCP cells and PHHs and further validated our approach (Figure 1C-D). 363 proteins detected in at least 2 ChroP experiments for each cell type (Supplementary Table 2 – *under embargo for pending patent*) were submitted to Gene Set Enrichment analysis (GSEA). Scatterplot visualization of GSEA results showed that the majority of the significantly enriched pathways were shared between HepG2-NTCP cells and PHH (Figure 2A) and comprised the GOslim categories RNA splicing, DNA damage, Ubiquitin, Nucleotide-binding process and transcription (Figure 2B). However, a consistent number of pathways were differentially enriched between the two cell models. In particular, in transformed hepatocytes, proteins belonging to GOslim categories angiogenesis, inflammation and TGFb and MAPK signaling were among the most enriched, while PHHs showed cccDNA-associated proteins mostly belonging to GOslim categories mRNA and histone dynamics, autophagy, kinase activity and glucose homeostasis (Figure 2B).

Connectivity map of cccDNA-associated factors based on protein-protein functional interactions

We next plotted the 363 cccDNA-associated proteins identified by ChroP analysis in a network representation where nodes identify individual proteins (Figure 2C). The interaction network was prepared using the Reactome functional interaction application implemented within Cytoscape 3.4 (see Experimental procedures for details). Using the Reactome database, interactions were found for 182 proteins that were derived either from HepG2-NTCP cells (nodes with white border), PHHs (nodes with grey border) or both (nodes with green border). The network was clustered using a network clustering algorithm and each cluster, or module, was defined by hypernodes. Extended description of associated pathway and biological process annotations (FDR <0.05) are presented in Supplementary Table 3. Consistently with results showed in Figure 2B, most of the proteins identified in both cell types belonged to modules 1 to 3 (RNA processing, ubiquitin and transcription). Highlighted protein names correspond to cccDNA-associated partners that belong to a *core network* of the Top 28 proteins identified by applying majority voting system, i.e. selecting proteins that were repeatedly detected across ChroP experiments (see Supplementary Table 2 for the complete list).

A FUS-centered network of proteins is associated to cccDNA *in vivo*

The 28 cccDNA-associated proteins identified by applying majority-voting system were further analyzed to select hits to be investigated in functional tests. The interaction with cccDNA of nine of them was confirmed by independent ChIP experiments in infected cell lines (Figure 1C-D). Repeatedly identified proteins (5 experiments out of 5 in PHH and 3 out of 4 in HepG2-NTCP cells) included BPI fold-containing family B, member 2 (BPIFB2) and galectin-3-binding protein (LGALS3BP) (Supplementary Table 2 – *under embargo for pending patent*). Both proteins have been described for their role in innate immune response against bacteria or viruses, but have never been investigated in HBV infections[35,36]. Interestingly, a cluster of proteins belonging to the protein ubiquitination pathway were also over-represented, including

proteasome components PSMC6, PSMD12 and PSMD13 and the deubiquitinase OTUB1. However, the majority of proteins (13 out of 28) displayed nucleic acid-binding activity; a subnetwork centered around the fused in sarcoma (FUS) protein was specifically highlighted by using the STRING software (Figure 3A). FUS has been shown to physically contact RNAP II and the basal transcription machinery to regulate RNAP II processivity, transcription-coupled RNA splicing and telescripting[37,38]. In addition, FUS has been shown to interact with DDX39, the central component of the TRanscription-EXport (TREX) complex, which also contains ALYREF, Tex1 and the THO complex[39,40]. The TREX complex is specifically recruited to activated genes during transcription and travels the entire length of the gene with RNAP II and exerts a conserved role in coupling transcription to RNA export[41]. Binding of FUS and interacting proteins to cccDNA was confirmed by independent ChIP experiments *in vivo*, in human liver resections derived from HBV chronically infected patients (Figure 3B).

Several additional proteins involved in splicing and RNA processing known to associate to FUS in HeLa cells[42] were enriched amongst the cccDNA-interactome dataset, namely DBIRD complex component ZNF326, ensuring integration between RNAP II elongation and splicing[43]; splicing factors SF3A1, SF3B1, EFTUD2, DHX15, SAFB, SAFB2, TRA2B, PRPF19 and PUF60; DEAD box helicases DDX5 and DDX17, coupling transcription to splicing[44]; ATP-dependent RNA helicase up-frameshift 1(UPF1/RENT1), a key player in nonsense-mediated mRNA decay (NMD) ensuring co-transcriptional RNA surveillance[45] (Supplementary Table 1 – *under embargo for pending patent*).

Intriguingly, three members of the tRNA splicing-ligase complex, FAM98A, DDX1 and RTCB, which were shown to interact with FUS in HEK293 cells[46], were also identified amongst cccDNA-associated proteins (Supplementary Table 1 - *under embargo for pending patent*).

FUS knockdown severely impairs cccDNA transcriptional activity in infected hepatocytes

RNAi experiments were performed in infected PHHs in order to investigate the functional impact on HBV replication of knockdown of selected proteins belonging to the 28 cccDNA-associated proteins identified by applying majority-voting system (Figure 4). Knockdown efficacy was controlled by mRNA expression of each target by qPCR (Figure 4A). cccDNA, 3.5Kb RNA and HBV replicative intermediates (totHBV DNA) were assessed by qPCR as described previously[24] and detailed in Experimental procedures. While cccDNA pool levels were not affected (Figure 4B), FUS knockdown significantly reduced both HBV 3.5Kb RNA and total HBV DNA levels (Figure 4C-D). Knockdown of the other proteins only reduced total HBV DNA levels, suggesting a post-transcriptional regulation mechanism.

cccDNA-ChIP analysis after FUS knockdown showed a reduction in ALYREF and DDX39 recruitment to cccDNA, together with a decrease in histone H3 lysine 36 trimethylation. Global RNAP II enrichment seemed not to be affected by TLS/FUS downregulation, as well as the levels of RNAP II CTD-phosphorylated Serine 5, while RNAP II CTD- phosphorylation of Serine 2 was significantly less detected after FUS downregulation. This suggests an alteration in RNAP II processivity and in RNA splicing/export factors recruitment on cccDNA in the absence of FUS.

Discussion

HBV is classified as a para-retrovirus, since it replicates its genomic DNA via the synthesis of an RNA intermediate[47]. The unique template for viral RNA transcription, cccDNA, is formed in the nucleus of infected hepatocytes by completion of the rcDNA genome contained in the virion[48]. It is generally accepted that cccDNA activity heavily resides on host factors and cofactors regulating its transcription and chromatin accessibility[17]. Chromosome conformation capture experiments (Hi-C) indicated that cccDNA contacts host cell chromatin preferentially in regions corresponding to CpG islands, suggesting that HBV might take advantage of open host chromatin regions to hijack the host transcriptional machinery and efficiently replicate itself[49].

Chromatin is a large subcellular structure dynamically and differentially ordered to allow physiological levels of gene expression[50]. By coupling Chromatin Immunoprecipitation and Mass spectrometry analysis, we described a new approach to investigate the interactome of a chromatinized episome without *a priori* assumptions. In particular, we were able to draw the landscape of HBV cccDNA-interacting proteins in the context of natural infection, a condition where, after viral inoculation, cccDNA intrinsically settles into the nucleus of infected cells and transcribes the viral RNAs.

We found that different categories of proteins were associated to cccDNA in infected hepatoma cells *vs* primary hepatocytes. In HepG2-NTCP infected cells, cccDNA-interacting proteins belonged to pathways associated to epithelial-mesenchimal transition, Notch signaling and inflammation, while in PHH, mostly to metabolic pathways. Published RNA-Seq profiling of HepG2 cells *vs* liver tissue revealed that the differentially regulated pathways were largely associated to cancer aggressiveness and immune system processes (blood plasma and immune system components protein production) in HepG2 and to metabolism and response to stimulus in liver tissue[51]. Therefore, assuming that RNA expression mirrors, at least to some extents, chromatin accessibility, our data support the notion that cccDNA contacts open chromatin regions enriched for different transcription factors and cofactors which are specific

to the different transcriptional programs of distinct cell models, pointing out the importance of performing such studies in relevant biological systems.

Interestingly enough, cccDNA-associated proteins involved in nucleotide binding processes (transcription, DNA damage), RNA processing and protein catabolism were shared between hepatoma and primary hepatocytes models. These data strongly suggest that HBV minichromosome, similarly to human genome, also adopts a protein-DNA-RNA structure to ensure its proper and efficient activity. In particular, we identified different cccDNA-associated complexes coupling basal transcription machinery to RNA processing and export, all of them containing the DNA/RNA binding protein FUS.

FUS belongs to the FET family of structurally related proteins[52], together with TAF15 and EWSR1, this latter also found in our cccDNA-interactome dataset. FUS is an essential component of paraspeckles, protein-rich nuclear membraneless organelles built around the long non-coding (lnc)RNA nuclear paraspeckle assembly transcript 1 (NEAT1)[53]. Beyond FUS, most of the proteins essential for paraspeckles formation were found in the cccDNA interactome, namely HNRNPK, NONO, SFPQ and RBM14[53] (Figure 1 and Supplementary Table 1 – *under embargo for pending patent*). Paraspeckels are dynamic structures modulated by changes in cellular metabolic activity and dependent on active transcription[54]. Their precise function is still not fully understood, but probably resides on the capacity of being a platform for retaining and putting in contact proteins involved in RNA processing and export, while active RNAP II and nascent RNAs are found at the edges of paraspekles[55]. NEAT1 levels are increased following infection with several RNA viruses[56,57] and NEAT1 and other paraspeckles components associate with HSV-1 genomic DNA increasing viral gene expression and viral replication[58]. According to our data, we cannot exclude that cccDNA associates with paraspeckles *in vivo*, since the results strongly suggest that macromolecular complexes are gathered together in proximity of the HBV genome to couple viral transcription to RNA processing and export. Indeed, a number of U1 snRNP components and hnRNPs proteins were found in our dataset. Most of these factors were identified as FUS-interacting partners

by MS analysis in HEK293 or HeLa cells[42,46] and only one of them, PUF60, was already shown to be involved in HBV RNA processing[34].

Similarly, the TREX component ALYREF, coupling transcription, splicing and mRNA export interacts with FUS via DDX39[39]. ALYREF has been also shown to interact with HBc and to be involved in pgRNA nuclear export via NFX1-p15 pathway[59]. Albeit TREX complex has been shown to be recruited to a plasmid containing the HBV PRE sequence by the zinc-finger ZC3H18 in HeLa cells[40], in our model ZC3H18 could not be identified among the cccDNA-associated proteins and downregulation of FUS expression significantly affected TREX components association with cccDNA. Phosphorylation on Serine 2 of cccDNA-associated RNAP II CTD was also decreased following FUS knockdown, together with the global levels of H3K36me3, suggesting that the processivity of the enzyme would be affected, resulting in a decrease of the viral RNA levels.

Ultimately, we developed a technique that could potentially allow to identifying proteins co-opted by other pathogens, thus favoring the characterization of better therapeutic targets for future studies. Further work will be required to determine whether and how the remaining host factors identified by our approach can impinge on HBV function.

References

1. Allweiss, L. & Dandri, M. The Role of cccDNA in HBV Maintenance. *Viruses* **9**, (2017).
2. Werle-Lapostolle, B. *et al.* Persistence of cccDNA during the natural history of chronic hepatitis B and decline during adefovir dipivoxil therapy. *Gastroenterology* **126**, 1750–1758 (2004).

3. Boyd, A. *et al.* Decay of ccc-DNA marks persistence of intrahepatic viral DNA synthesis under tenofovir in HIV-HBV co-infected patients. *J. Hepatol.* **65**, 683–691 (2016).
4. Lampertico, P. & Liaw, Y. F. New perspectives in the therapy of chronic hepatitis B. *Gut* **61 Suppl 1**, i18-24 (2012).
5. Newbold, J. E. *et al.* The covalently closed duplex form of the hepadnavirus genome exists in situ as a heterogeneous population of viral minichromosomes. *J. Virol.* **69**, 3350–3357 (1995).
6. Bock, C. T., Schranz, P., Schröder, C. H. & Zentgraf, H. Hepatitis B virus genome is organized into nucleosomes in the nucleus of the infected cell. *Virus Genes* **8**, 215–229 (1994).
7. Bock, C. T. *et al.* Structural organization of the hepatitis B virus minichromosome. *J. Mol. Biol.* **307**, 183–196 (2001).
8. Hatton, T., Zhou, S. & Standring, D. N. RNA- and DNA-binding activities in hepatitis B virus capsid protein: a model for their roles in viral replication. *J. Virol.* **66**, 5232–5241 (1992).
9. Guo, Y.-H., Li, Y.-N., Zhao, J.-R., Zhang, J. & Yan, Z. HBc binds to the CpG islands of HBV cccDNA and promotes an epigenetic permissive state. *Epigenetics* **6**, 720–726 (2011).
10. Belloni, L. *et al.* Nuclear HBx binds the HBV minichromosome and modifies the epigenetic regulation of cccDNA function. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19975–19979 (2009).
11. Pollicino, T. *et al.* Hepatitis B virus replication is regulated by the acetylation status of hepatitis B virus cccDNA-bound H3 and H4 histones. *Gastroenterology* **130**, 823–837 (2006).
12. Benhenda, S. *et al.* Methyltransferase PRMT1 is a binding partner of HBx and a negative regulator of hepatitis B virus transcription. *J. Virol.* **87**, 4360–4371 (2013).
13. Zhang, W. *et al.* PRMT5 Restricts Hepatitis B Virus Replication via Epigenetic Repression of cccDNA Transcription and Interference with pgRNA Encapsidation. *Hepatol. Baltim. Md* (2017) doi:10.1002/hep.29133.

14. Ren, J.-H. *et al.* SIRT3 restricts hepatitis B virus transcription and replication through epigenetic regulation of covalently closed circular DNA involving suppressor of variegation 3-9 homolog 1 and SET domain containing 1A histone methyltransferases. *Hepatol. Baltim. Md* (2018) doi:10.1002/hep.29912.
15. Tropberger, P. *et al.* Mapping of histone modifications in episomal HBV cccDNA uncovers an unusual chromatin organization amenable to epigenetic manipulation. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5715-5724 (2015).
16. Flecken, T. *et al.* Mapping the Heterogeneity of Histone Modifications on Hepatitis B Virus DNA Using Liver Needle Biopsies Obtained from Chronically Infected Patients. *J. Virol.* **93**, (2019).
17. Hong, X., Kim, E. S. & Guo, H. Epigenetic regulation of hepatitis B virus covalently closed circular DNA: Implications for epigenetic therapy against chronic hepatitis B. *Hepatol. Baltim. Md* **66**, 2066–2077 (2017).
18. Cristea, I. M. & Graham, D. Virology meets Proteomics. *Proteomics* **15**, 1941–1942 (2015).
19. Fang, C.-Y. *et al.* Global profiling of histone modifications in the polyomavirus BK virion minichromosome. *Virology* **483**, 1–12 (2015).
20. Wierer, M. & Mann, M. Proteomics to study DNA-bound and chromatin-associated gene regulatory complexes. *Hum. Mol. Genet.* **25**, R106–R114 (2016).
21. Ladner, S. K. *et al.* Inducible expression of human hepatitis B virus (HBV) in stably transfected hepatoblastoma cells: a novel system for screening potential inhibitors of HBV replication. *Antimicrob. Agents Chemother.* **41**, 1715–1720 (1997).
22. Luangsay, S. *et al.* Early inhibition of hepatocyte innate responses by hepatitis B virus. *J. Hepatol.* **63**, 1314–1322 (2015).
23. Allweiss, L. *et al.* Proliferation of primary human hepatocytes and prevention of hepatitis B virus reinfection efficiently deplete nuclear cccDNA in vivo. *Gut* **67**, 542–552 (2018).

24. Testoni, B. *et al.* Serum hepatitis B core-related antigen (HBcrAg) correlates with covalently closed circular DNA transcriptional activity in chronic hepatitis B patients. *J. Hepatol.* **70**, 615–625 (2019).
25. Testoni, B. *et al.* Ribavirin restores IFNα responsiveness in HCV-infected livers by epigenetic remodelling at interferon stimulated genes. *Gut* (2015) doi:10.1136/gutjnl-2014-309011.
26. Hirt, B. Selective extraction of polyoma DNA from infected mouse cell cultures. *J. Mol. Biol.* **26**, 365–369 (1967).
27. Belloni, L. *et al.* IFN-α inhibits HBV transcription and replication in cell culture and in humanized mice by targeting the epigenetic regulation of the nuclear cccDNA minichromosome. *J. Clin. Invest.* **122**, 529–537 (2012).
28. Natoli, G. *et al.* Characterization of the hepatitis B virus preS/S region encoded transcriptional transactivator. *Virology* **187**, 663–670 (1992).
29. Kim, W., Lee, S., Son, Y., Ko, C. & Ryu, W.-S. DDB1 Stimulates Viral Transcription of Hepatitis B Virus via HBx-Independent Mechanisms. *J. Virol.* **90**, 9644–9653 (2016).
30. Makokha, G. N. *et al.* Regulation of the Hepatitis B virus replication and gene expression by the multi-functional protein TARDBP. *Sci. Rep.* **9**, 8462 (2019).
31. Tay, N., Chan, S. H. & Ren, E. C. Identification and cloning of a novel heterogeneous nuclear ribonucleoprotein C-like protein that functions as a transcriptional activator of the hepatitis B virus enhancer II. *J. Virol.* **66**, 6841–6848 (1992).
32. Ng, L. F. P. *et al.* Host heterogeneous ribonucleoprotein K (hnRNP K) as a potential target to suppress hepatitis B virus replication. *PLoS Med.* **2**, e163 (2005).
33. Ko, H.-L. & Ren, E.-C. Novel poly (ADP-ribose) polymerase 1 binding motif in hepatitis B virus core promoter impairs DNA damage repair. *Hepatol. Baltim. Md* **54**, 1190–1198 (2011).
34. Sun, S. *et al.* Involvement of PUF60 in Transcriptional and Post-transcriptional Regulation of Hepatitis B Virus Pregenomic RNA Expression. *Sci. Rep.* **7**, 12874 (2017).

35. Bingle, C. D. & Craven, C. J. PLUNC: a novel family of candidate host defence proteins expressed in the upper airways and nasopharynx. *Hum. Mol. Genet.* **11**, 937–943 (2002).
36. Loimaranta, V., Hepojoki, J., Laaksoaho, O. & Pulliainen, A. T. Galectin-3-binding protein: A multitask glycoprotein with innate immunity functions in viral and bacterial infections. *J. Leukoc. Biol.* **104**, 777–786 (2018).
37. Yu, Y. & Reed, R. FUS functions in coupling transcription to splicing by mediating an interaction between RNAP II and U1 snRNP. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8608–8613 (2015).
38. Kaida, D. *et al.* U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664–668 (2010).
39. Sugiura, T., Sakurai, K. & Nagano, Y. Intracellular characterization of DDX39, a novel growth-associated RNA helicase. *Exp. Cell Res.* **313**, 782–790 (2007).
40. Chi, B. *et al.* Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. *Nucleic Acids Res.* **41**, 1294–1306 (2013).
41. Strässer, K. *et al.* TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417**, 304–308 (2002).
42. Chi, B. *et al.* Interactome analyses revealed that the U1 snRNP machinery overlaps extensively with the RNAP II machinery and contains multiple ALS/SMA-causative proteins. *Sci. Rep.* **8**, 8755 (2018).
43. Close, P. *et al.* DBIRD complex integrates alternative mRNA splicing with RNA polymerase II transcript elongation. *Nature* **484**, 386–389 (2012).
44. Giraud, G., Terrone, S. & Bourgeois, C. F. Functions of DEAD box RNA helicases DDX5 and DDX17 in chromatin organization and transcriptional regulation. *BMB Rep.* **51**, 613–622 (2018).
45. Hong, D., Park, T. & Jeong, S. Nuclear UPF1 Is Associated with Chromatin for Transcription-Coupled RNA Surveillance. *Mol. Cells* **42**, 523–529 (2019).

46. Kamelgarn, M. *et al.* Proteomic analysis of FUS interacting proteins provides insights into FUS function and its role in ALS. *Biochim. Biophys. Acta* **1862**, 2004–2014 (2016).
47. Seeger, C. & Mason, W. S. Hepatitis B virus biology. *Microbiol. Mol. Biol. Rev. MMBR* **64**, 51–68 (2000).
48. Nassal, M. HBV cccDNA: viral persistence reservoir and key obstacle for a cure of chronic hepatitis B. *Gut* **64**, 1972–1984 (2015).
49. Moreau, P. *et al.* Tridimensional infiltration of DNA viruses into the host genome shows preferential contact with active chromatin. *Nat. Commun.* **9**, 4268 (2018).
50. Tai, P. W. L. *et al.* The dynamic architectural and epigenetic nuclear landscape: developing the genomic almanac of biology and disease. *J. Cell. Physiol.* **229**, 711–727 (2014).
51. Tyakht, A. V. *et al.* RNA-Seq gene expression profiling of HepG2 cells: the influence of experimental factors and comparison with liver tissue. *BMC Genomics* **15**, 1108 (2014).
52. Schwartz, J. C., Cech, T. R. & Parker, R. R. Biochemical Properties and Biological Functions of FET Proteins. *Annu. Rev. Biochem.* **84**, 355–379 (2015).
53. Fox, A. H., Nakagawa, S., Hirose, T. & Bond, C. S. Paraspeckles: Where Long Noncoding RNA Meets Phase Separation. *Trends Biochem. Sci.* **43**, 124–135 (2018).
54. Fox, A. H. *et al.* Paraspeckles: a novel nuclear domain. *Curr. Biol. CB* **12**, 13–25 (2002).
55. Fox, A. H. & Lamond, A. I. Paraspeckles. *Cold Spring Harb. Perspect. Biol.* **2**, a000687 (2010).
56. Imamura, K. *et al.* Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. *Mol. Cell* **53**, 393–406 (2014).
57. Zhang, Q., Chen, C.-Y., Yedavalli, V. S. R. K. & Jeang, K.-T. NEAT1 long noncoding RNA and paraspeckle bodies modulate HIV-1 posttranscriptional expression. *mBio* **4**, e00596-00512 (2013).
58. Wang, Z. *et al.* NEAT1 modulates herpes simplex virus-1 replication by regulating viral gene transcription. *Cell. Mol. Life Sci. CMLS* **74**, 1117–1131 (2017).

59. Yang, C.-C., Huang, E.-Y., Li, H.-C., Su, P.-Y. & Shih, C. Nuclear export of human hepatitis B virus core protein and pregenomic RNA depends on the cellular NXF1-p15 machinery. *PloS One* **9**, e106683 (2014).

Figure legends

Figure 1. Chromatin proteomics represents an unbiased approach to identify HBV minichromosome interactome A) Schematic representation of ChroP protocol. B) Venn diagram of proteins specifically detected in HBV-infected and not in mock-infected HepG2-NTCP cells or PHH. C-D) Indipendent ChIP-qPCR experiments were done after 9 days of infection in HepG2-NTCP (C) and in PHH (D) using antibodies against proteins identified by ChroP screening. NoAb and E2F ChIP served as negative controls, while RNAP II ChIP served as positive control for active cccDNA transcription. cccDNA-ChIP protocol is detailed in Experimental Procedures. Graphs represent the mean ± SD of three independent experiments.

Figure 2: ChroP analysis reveals different landscapes of cccDNA-associated factors in transformed versus primary infected hepatocytes. A) Pathway enrichment scatterplot. In yellow and green, pathways specifically enriched in HepG2-NTCP and PHH, respectively; in violet, pathways enriched in both cell models; in light grey pathways that did not reach statistical significance. B) Barchart of enrichment scores of selected pathways, significant either in PHH, or in HepG2, or in both cell models. The p-value of enrichment in HepG2-NTCP cells and PHH is given as Benjamini-Hochberg adjusted p-value of Fisher's exact test. Monte-Carlo p-values are calculated by randomly resampling the genes without replacement and by counting the frequency of enrichment. C) Functional interaction network for 182 proteins found to associate to HBV cccDNA was built using the Reactome database. The network was clustered using a network clustering algorithm (spectral partition based network clustering by Newman 2006) into modules, which were colored, numbered and highlighted by hypernodes. Module associated pathway and biological process annotations (FDR <0.05) are presented in Supplementary Table 2. Remaining proteins with no interactions defined in the Reactome database are not shown. Small modules (fewer than 4 proteins) were not annotated. Nodes in the network represent individual proteins; edges are presented as undirected connections between nodes, but are based upon the Reactome functional interactions. The

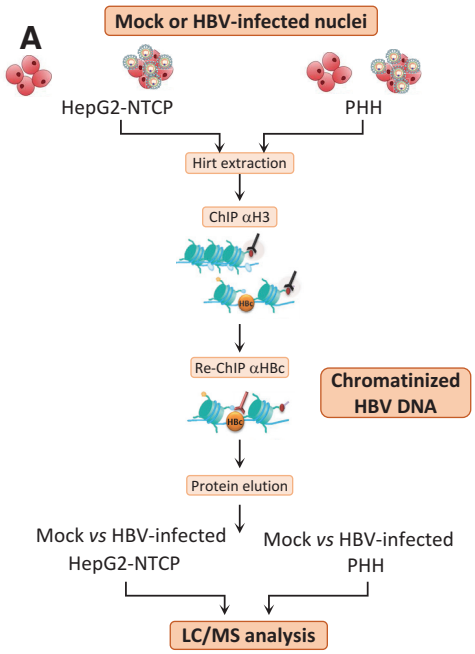
source of the identified proteins is indicated by the node border color: white, HepG2; grey, PHH; bright green, nodes found in both cell types or by majority voting. Only nodes defined as common or frequently identified by majority voting (Top 28 cccDNA-associated partners) are annotated with the official gene symbol. A summary annotation is provided for each hypernode. ER, endoplasmic reticulum; ERAD, Endoplasmic-reticulum-associated protein degradation.

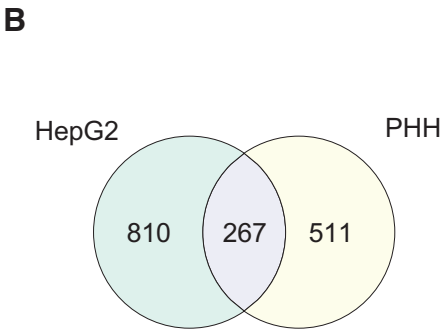
Figure 3. A FUS-interacting network of proteins is associated to cccDNA *in vivo*. A) Connectivity map based on protein-protein interactions with the FUS protein among the Top 28 cccDNA-associated partners obtained using the STRING software. B) cccDNA ChIP-qPCR validation of the FUS-interacting network *in vivo*, in liver tissue derived from chronic HBV-infected patients. Liver resections from 3 chronic hepatitis B patients were processed using micro-ChIP technology and cccDNA was quantified by qPCR as detailed in Experimental Procedures. Anti-RNAP II and E2F immunoprecipitations served as positive and negative control, respectively. Graph represents mean ± SD.

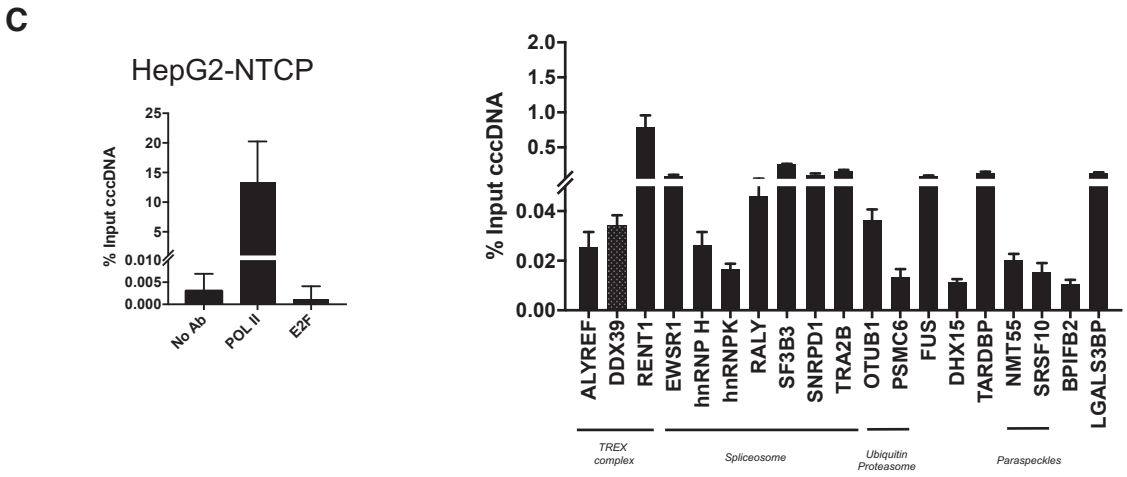
Figure 4. FUS knockdown severely impairs cccDNA transcriptional activity. Expression knockdown of 8 selected factors among the Top 28 cccDNA-associated proteins was performed using siRNA approach. PHH cells were transfected 5 days after infection and gene expression and viral parameters were analyzed 72h later. A) mRNA expression of the 8 selected targets after siRNA transfection. cccDNA (B), pgRNA (C) and HBV DNA intermediates (totHBV DNA, D) quantification after siRNA against indicated proteins. E) FUS mRNA (left panel) and protein levels (right panel) after FUS siRNA in HepG2-NTCP cells. F) cccDNA-ChIP analysis of FUS, ALYREF, DDX39, RNAP II, serine 2 or 5-phosphorylated RNAP II recruitment and histone H3 lysine 36 trimethylation enrichment on cccDNA after siRNA against FUS in HepG2-NTCP cells. Knock-down of target genes in infected PHHs or HepG2-NTCP cells was obtained by a double siRNA transfection performed 4 and 6 days after HBV infection. Cells were harvested 72h after the second transfection. Relative RNA values for targets mRNAs and HBV pgRNA were normalized over *GUSB* housekeeping gene

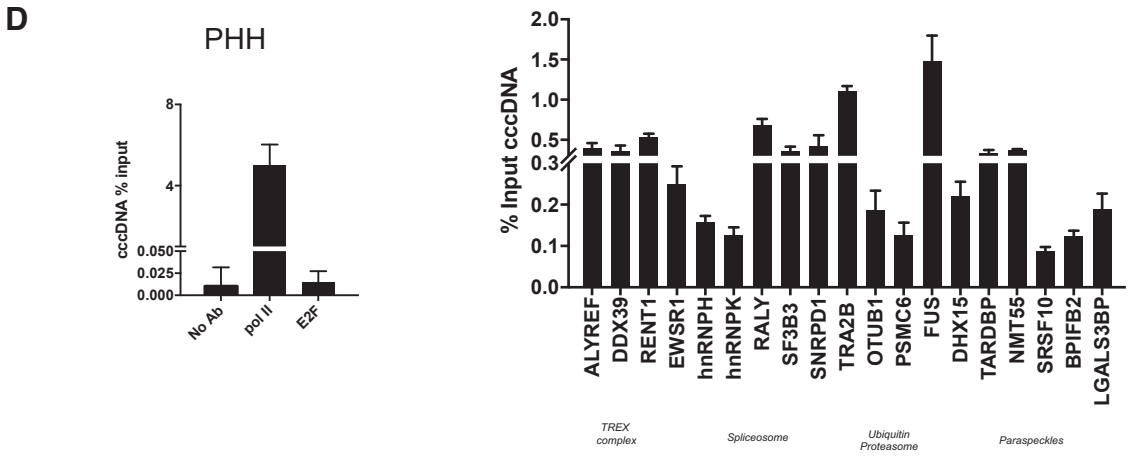
expression. Total HBV DNA and cccDNA copies/nucleus were obtained after normalization on bGlobin content. 1h T5 DNase digestion was performed before cccDNA-specific qPCR (see Experimental Procedures). Cells for cccDNA-ChIP were crosslinked and harvested 72h after the second siRNA transfection and submitted to ChIP as detailed in Experimental Procedures. Barplots represent mean ± SD of at least 3 independent experiments. *p<0.05; **p<0.005; ***p<0.001 Two-tailed p-values were calculated for a risk threshold = 0.05 using the Kruskal-Wallis test respect to siCTL condition.

Testoni et al. Figure 1

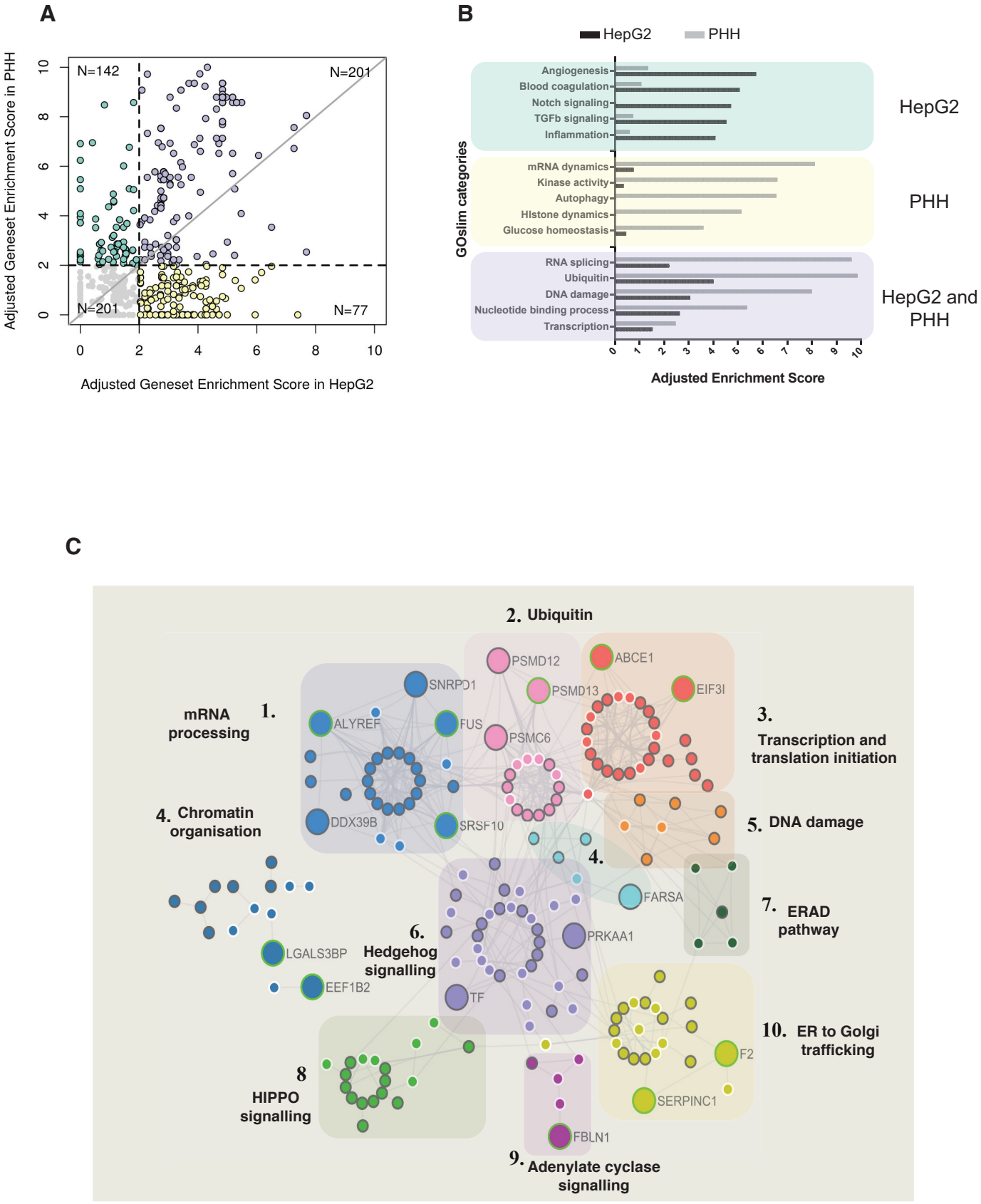




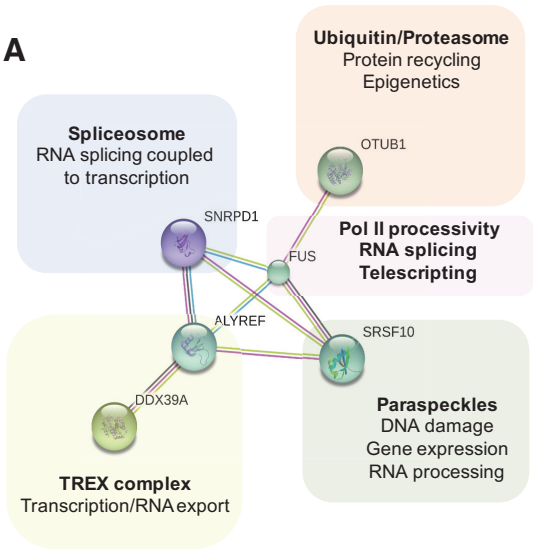


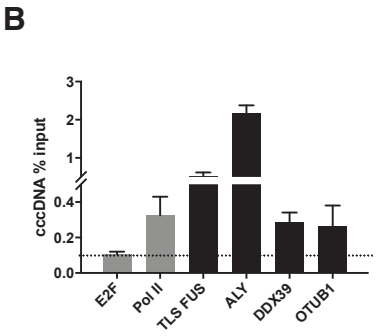


Testoni et al., Figure 2



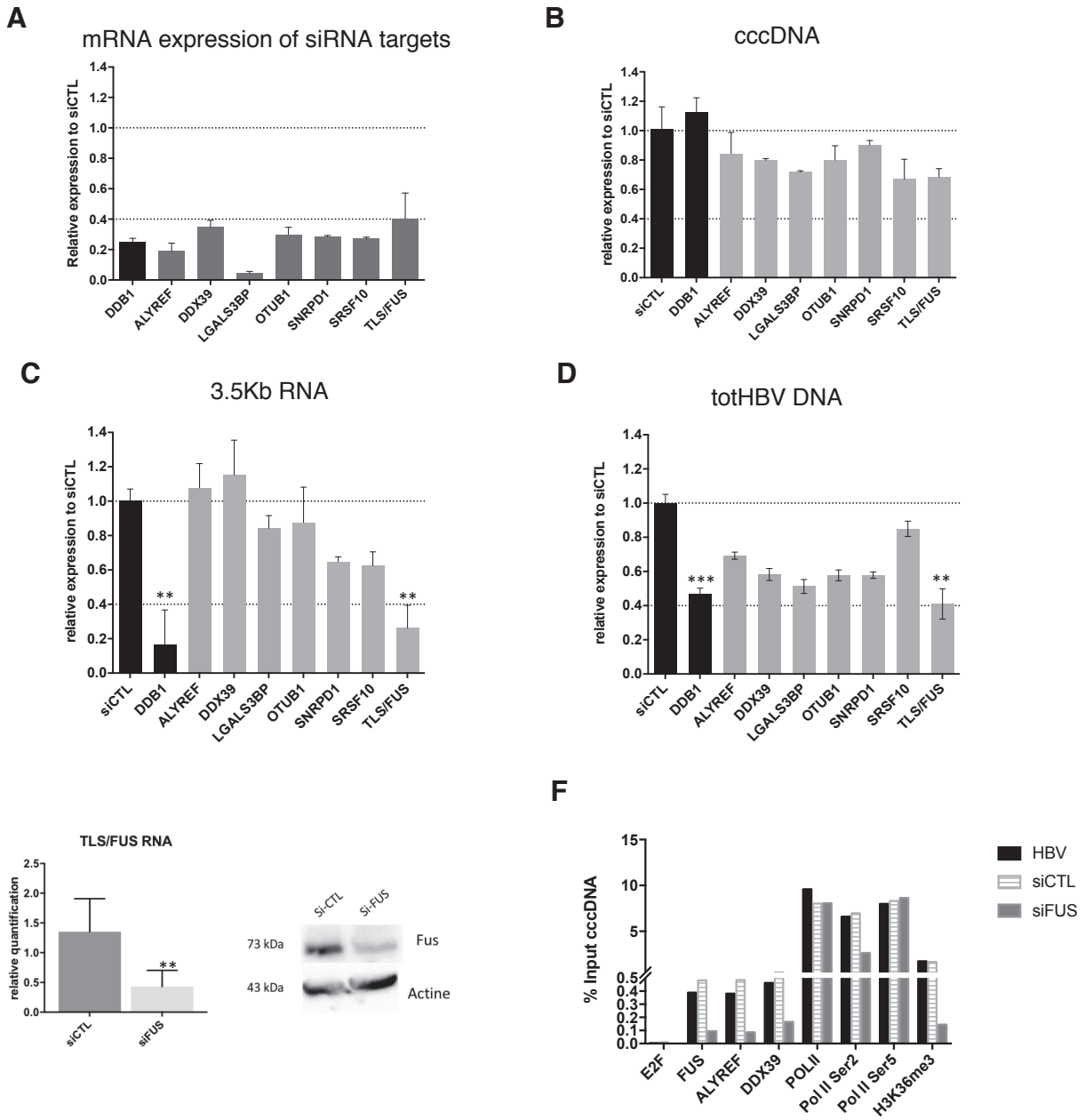
Testoni et al., Figure 3

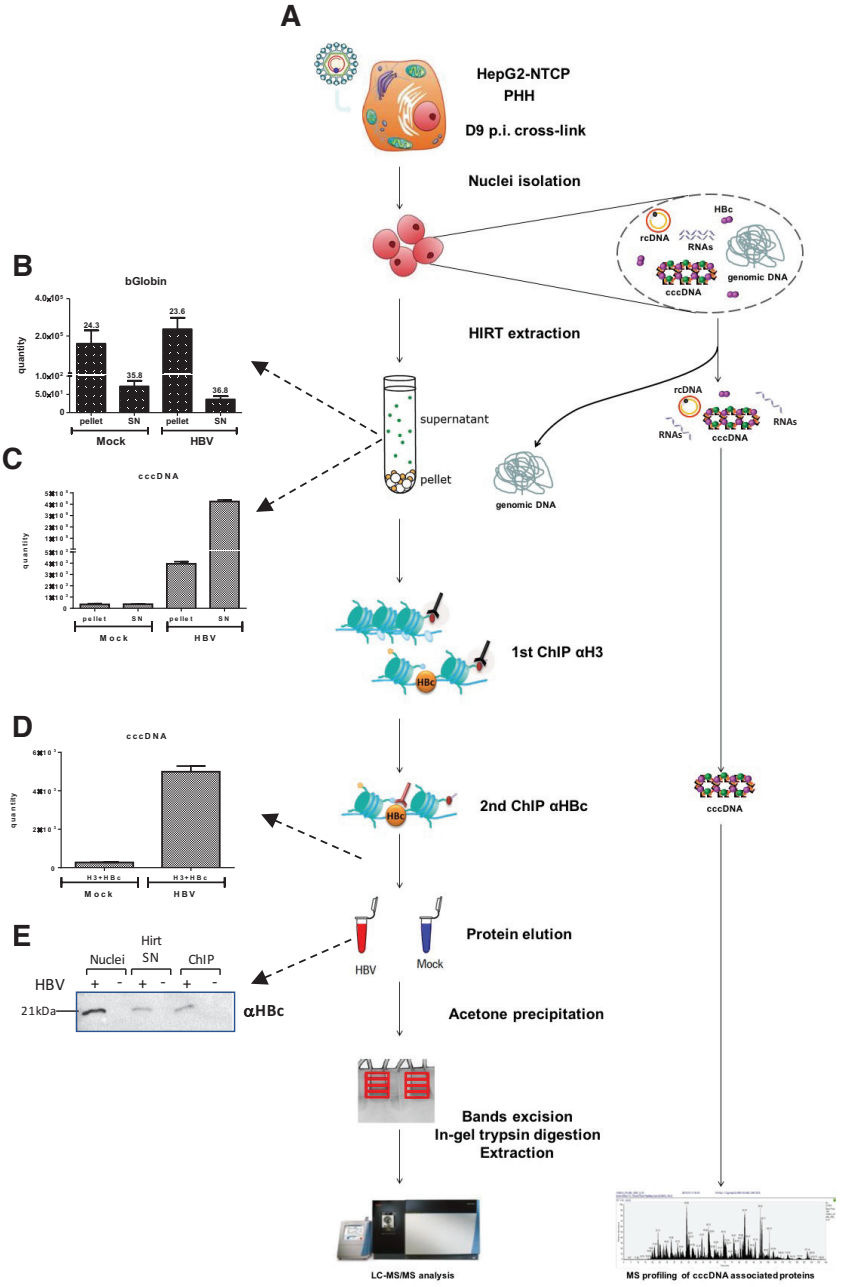




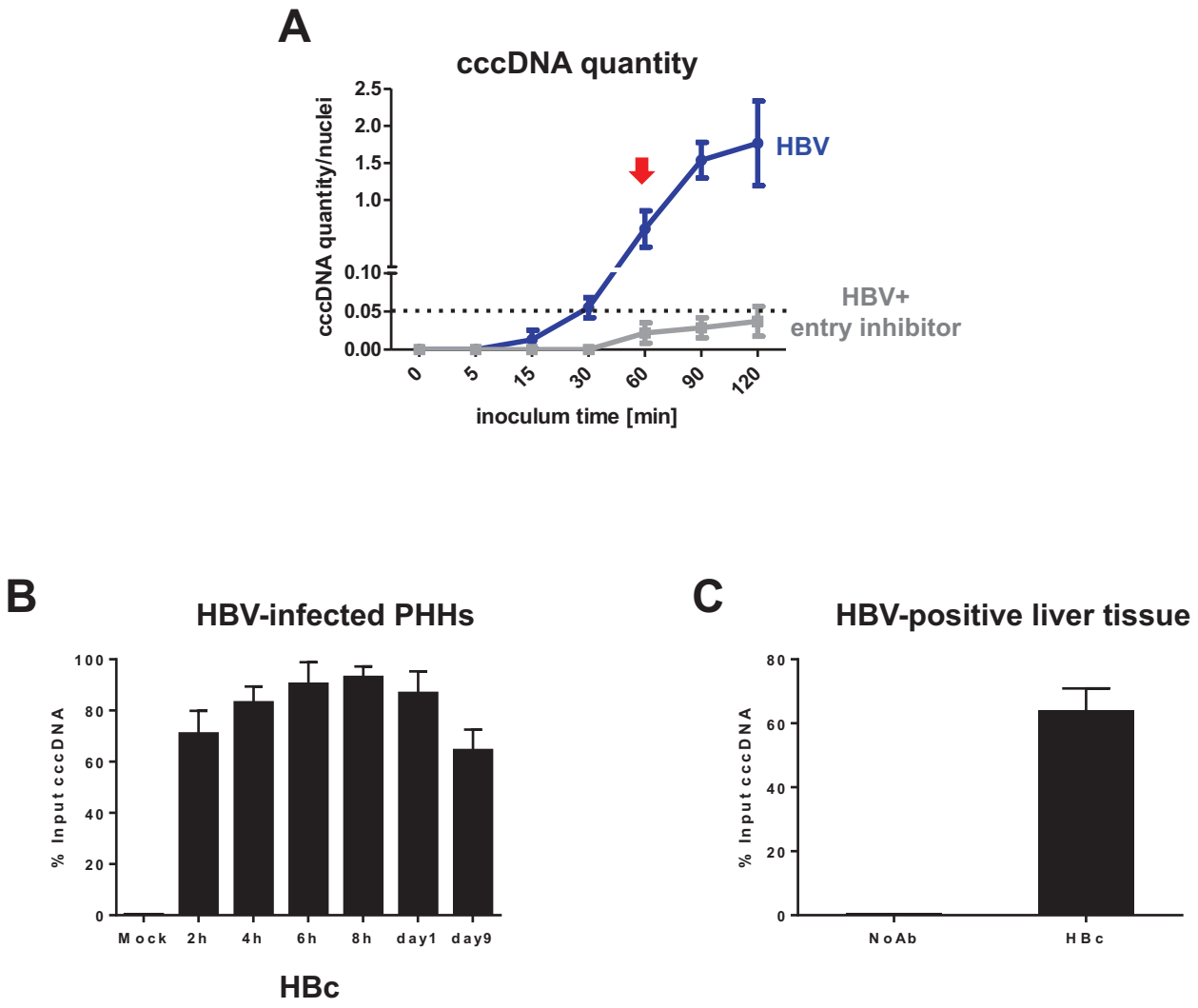
Liver tissue

Testoni et al., Figure 4





Supplementary Figure 1. Technical controls for cccDNA-ChroP approach. (A) Extended scheme of cccDNA-ChroP workflow. (B) and (C) qPCR quantification of β-globin gene and cccDNA in pellet and supernatant derived from Hirt extraction showing significant depletion of host cell genomic DNA and enrichment of cccDNA in Hirt-derived supernatant fraction. As expected, mock-infected and HBV-infected samples showed comparable quantities of β-globin gene, but cccDNA was detectable only in HBV-infected samples. (D) qPCR for cccDNA quantification after sequential ChIP. (E) western blotting analysis showing recovery of HBc protein during ChroP protocol in nuclei extracts, supernatant of HIRT extraction and after H3+HBc sequential ChIP. Data for one representative ChroP experiment performed in HepG2-NTCP cells are shown. qPCR for β -Globin and cccDNA quantification were performed as in[23,24]. The full cccDNA-ChroP protocol is detailed in Experimental Procedures.



Supplementary Figure 2. HBV core protein associates to cccDNA along HBV infection time course. (A) cccDNA quantification at early time points post-infection. PHHs were inoculated with HBV for the indicated time points and, then, extensively washed before harvesting. preS1-mimiking peptide was used as HBV entry inhibitor and administered 2h before and during HBV inoculation at 100nM concentration. B-C) cccDNA-ChIP using an antibody against HBc was performed at the indicated time points post-infection in PHHs (B) and *in vivo*, in a liver resection derived from a chronic HBV-infected patient. cccDNA quantity/nucleus was calculated after normalization of β-Globin expression as in[23,24]. cccDNA-ChIP protocol is detailed in Experimental Procedures. Graphs represent the mean ± SD of three independent experiments.

II. Full-length 5′RACE identifies all major HBV transcripts in HBV-infected hepatocytes and patient serum

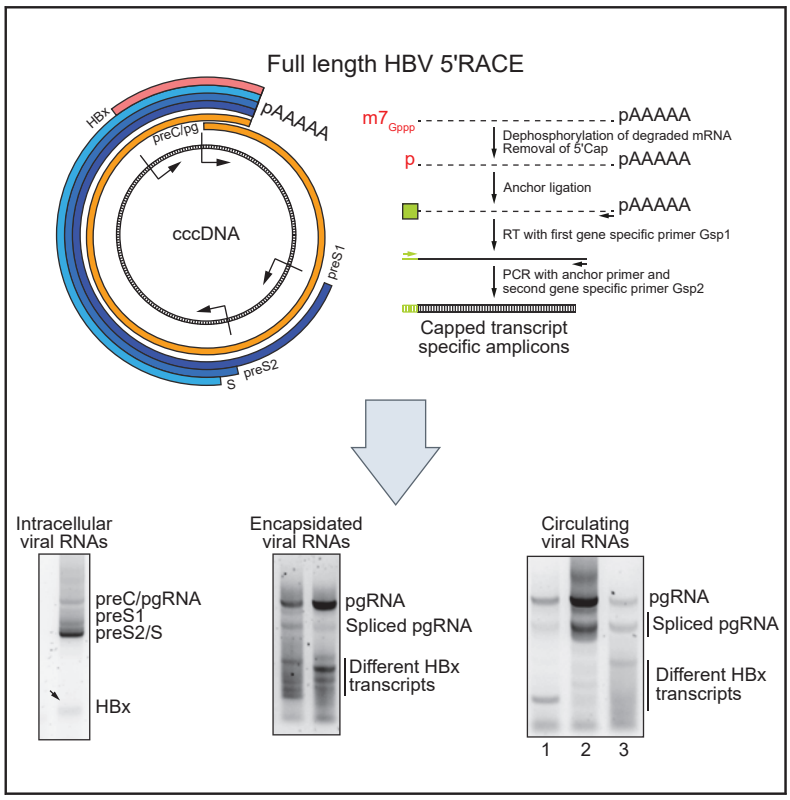
Published article

Research Article
Experimental and Translational Hepatology

JOURNAL
OF HEPATOLOGY

Full-length 5'RACE identifies all major HBV transcripts in HBV-infected hepatocytes and patient serum

Graphical abstract



Highlights

- HBV full-length 5'RACE discriminates all viral transcripts including HBx.
- Viral particles contain pgRNA, pgRNA splice variants and different HBx RNAs.
- HBx RNAs in viral particles are both capped and uncapped.
- The composition of circulating viral RNA species can vary among patients.

Authors

Bernd Stadelmayer, Audrey Diederichs, Fleur Chapus, …, Kara Carter, Barbara Testoni, Fabien Zoulim

Correspondence

bernd.stadelmayer@inserm.fr (B. Stadelmayer), fabien.zoulim@inserm.fr (F. Zoulim).

Lay summary

Especially under infection conditions, it has been difficult to study the different hepatitis B virus transcripts in depth. This study introduces a new method that can be used in any standard lab to discriminate all hepatitis B viral transcripts in cell culture and in the serum of patients.

https://doi.org/10.1016/j.jhep.2020.01.028

Research Article
Experimental and Translational Hepatology

JOURNAL OF HEPATOLOGY

Full-length 5'RACE identifies all major HBV transcripts in HBV-infected hepatocytes and patient serum

Bernd Stadelmayer[1,2,*], Audrey Diederichs[1,2], Fleur Chapus[1,2], Michel Rivoire[3], Gregory Neveu[4], Antoine Alam[4], Laurent Fraisse[4], Kara Carter[4], Barbara Testoni[1,2], Fabien Zoulim[1,2,5,*]

[1]INSERM U1052, CNRS UMR-5286, Cancer Research Center of Lyon (CRCL), Lyon, 69008, France; [2]University of Lyon, Université Claude-Bernard (UCBL), 69008 Lyon, France; [3]INSERM U1032, Centre Léon Bérard (CLB), 69008 Lyon, France; [4]Evotec, 1541 Avenue Marcel Mérieux, 69280 Marcy l'Etoile, France; [5]Hospices Civils de Lyon (HCL), 69002 Lyon, France

Background & Aims: Covalently closed circular DNA (cccDNA) is the episomal form of the HBV genome that stably resides in the nucleus of infected hepatocytes. cccDNA is the template for the transcription of 6 major viral RNAs, *i.e.* preC, pg, preS1/2, S and HBx RNA. All viral transcripts share the same 3' end and are all to various degrees subsets of each other. Especially under infection conditions, it has been difficult to study in depth the transcription of the different viral transcripts. We thus wanted to develop a method with which we could easily detect the full spectrum of viral RNAs in any lab.

Methods: We set up an HBV full-length 5'RACE (rapid amplification of cDNA ends) method with which we measured and characterized the full spectrum of viral RNAs in cell culture and in chronically infected patients.

Results: In addition to canonical HBx transcripts coding for full-length X, we identified shorter HBx transcripts potentially coding for short X proteins. We showed that interferon-β treatment leads to a strong reduction of preC and pgRNAs but has only a moderate effect on the other viral transcripts. We found pgRNA, 1 spliced pgRNA variant and a variety of HBx transcripts associated with viral particles generated by HepAD38 cells. The different HBx RNAs are both capped and uncapped. Lastly, we identified 3 major categories of circulating RNA species in patients with chronic HBV infection: pgRNA, spliced pgRNA variants and HBx.

Conclusions: This HBV full-length 5'RACE method should significantly contribute to the understanding of HBV transcription during the course of infection and therapy and may guide the development of novel therapies aimed at targeting cccDNA.

Lay summary: Especially under infection conditions, it has been difficult to study the different hepatitis B virus transcripts in depth. This study introduces a new method that can be used in any standard lab to discriminate all hepatitis B viral transcripts in cell culture and in the serum of patients.

Keywords: HBV transcripts; HBV viral particles; Serological HBV RNA; Hepatitis B; cccDNA.
Received 31 July 2019; received in revised form 10 January 2020; accepted 15 January 2020; available online xxx
* Corresponding author. Address: INSERM U1052, CNRS UMR-5286, Cancer Research Center of Lyon (CRCL), Lyon, 69008, France.
E-mail addresses: bernd.stadelmayer@inserm.fr (B. Stadelmayer), fabien.zoulim@inserm.fr (F. Zoulim).
https://doi.org/10.1016/j.jhep.2020.01.028

Introduction

Hepatitis B remains a major public health problem worldwide despite the availability of prophylactic vaccination and antiviral treatments.[1] HBV persists as a covalently closed circular DNA (cccDNA) of approximately 3.2 kb that embeds into the chromatin of hepatocytes as an episomal entity.[2] The amount of cccDNA and its transcriptional activity vary along the natural course of infection and are a main determinant of viral persistence and reactivation. From 4 promoters, 6 major viral RNAs are expressed: preCore (preC) RNA, pregenomic (pg)RNA, large surface protein (preS1) RNA, middle surface protein (preS2) RNA, small surface protein (S) RNA, X protein RNA (HBx) (Fig. 1A). preC and pgRNA transcripts are characterized by a 100-base pair (bp) redundancy at their 3' ends and consequently contain 2 epsilon stem-loop structures at their 5' and 3' ends (Fig. 1A). Every viral RNA is translated into proteins that are essential for the viral life cycle. In this regard, the pgRNA is unique in the sense that apart from encoding the 2 viral proteins C (core) and P (polymerase), it serves as the template for the viral DNA synthesis. Genomic viral DNA synthesis is mediated by the reverse transcriptase activity of the viral P protein once pgRNA and the P protein are encapsidated into the nucleocapsids formed by the C protein.[3,4] preC RNA codes for the HBe protein that has a potential immunoregulatory function.[5] preS1/2 and S mRNAs encode the 3 viral surface proteins L, M and S, respectively, that build up the viral envelope. The shortest of the 6 transcripts is the HBx transcript that gives rise to a 154 amino-acid long cytoplasmic and nuclear X protein.[6–10] X directly interacts with a multitude of host proteins explaining its diverse transactivating effects in host cells.[11–14]

The organization of the HBV genome is highly condensed and all transcripts are to various degrees subsets of each other (Fig. 1A). Indeed, all HBV transcripts share the same 3' end and, thus, the HBx sequence constitutes the 3' end of every viral transcript (Fig. 1A). Therefore, most HBV RNAs are indistinguishable by quantitative real-time PCR (qPCR). At present, only the larger viral RNAs can be differentiated by northern blotting during the course of infection. Most probably because of low expression levels, HBx transcripts are undetectable by northern blotting under infection conditions. Accordingly, not much is known about HBx transcription during viral infection. Moreover,

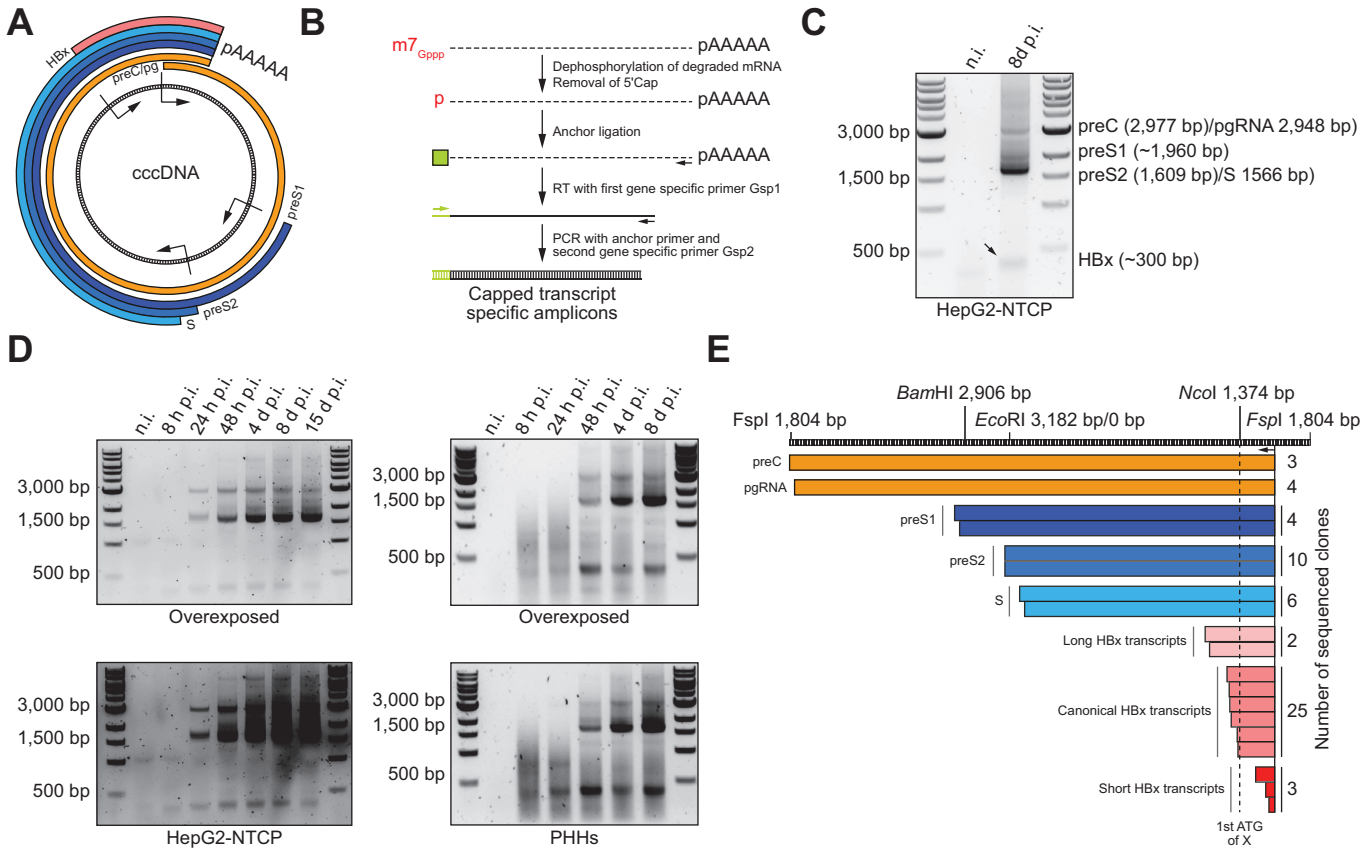


Fig.1. Set up of the HBV full-length 5'RACE. (A) Illustration of the HBV genome with its 4 promoters (arrows) and 6 viral transcripts: preC RNA, pgRNA, preS1/2 and S RNA, HBx and cccDNA. (B) Overview of the capped RNA-specific HBV full-length 5'RACE protocol. In a first step, RNAs are dephosphorylated to assure that degraded and truncated RNAs are excluded from the 5'RACE assay. 5' capped full-length RNAs are protected from the dephosphorylation reaction. Subsequently, the 5' cap structure (^{m7}Gppp) is removed from the full-length RNAs by a de-capping reaction resulting in a free phosphate group (p) at their 5' ends to which an RNA anchor (green box) can be ligated in the following step. Then, the viral RNAs are reverse transcribed using the gene specific primer Gsp1 (black arrow) integrating the anchor sequence into the cDNA. Finally, the different viral cDNAs are amplified by PCR using the second gene specific primer Gsp2 (black arrow) and an anchor specific primer (green arrow) in the same tube. (C) Transcript-specific 5'RACE amplicons using HepG2-NTCP cells 8 days post infection (p.i.) as compared to non-infected (n.i.) cells. Arrow indicates HBx. Individual amplicons are reduced in length by approximately 337 bp with respect to template viral RNAs because of the usage of gene specific primers. All amplicons were sequenced by subcloning and Sanger sequencing (Table S1 and Fig. 1E for overview). (D) HBx is the first viral transcript detected after infection in HepG2-NTCP cells and PHHs. (E) Alignment of sequenced amplicons with respect to HBV genome genotype D (reference sequence GenBank U95551.1). pgRNA and preC RNA in orange. preS1 in dark blue, preS2 in blue and S in light blue. Long HBx transcripts in very light red, canonical HBx transcripts in light red and short HBx transcripts in red. To the right, total number of sequenced clones per transcript category. HBx, X protein RNA; pg, pregenomic; PHHs, primary human hepatocytes; preC, preCore; preS1, large surface protein; preS2, middle surface protein; RACE, rapid amplification of cDNA ends; S, small surface protein.

there is a knowledge gap about the viral RNA species, per se, found in cell culture, but also in viral particles and in chronically HBV-infected patients.

To this end, we set up a strategy called HBV full-length 5'RACE (rapid amplification of cDNA ends) with which we characterized all major intra- and extracellular HBV RNAs during viral infection of cultured hepatocytes. With the HBV full-length 5' RACE approach, we followed the expression of the different viral RNAs over the course of infection and additionally studied how they are modulated by various drug treatments. Furthermore, we identified the viral RNAs that are associated with viral particles produced by HepAD38 cells. Finally, we analyzed the full spectrum of circulating HBV RNAs in patients with chronic HBV infection.

Materials and methods

Cell lines, viral inoculum and infection conditions

For infection purpose, HBV particles were concentrated from the supernatant of HepAD38 (HBV genotype D) cells by filtering and PEG precipitation as described previously.[15] The HepAD38 cell line was a kind gift of Dr C. Seeger (Fox Chase Cancer Center, Philadelphia, USA).[16] The HepG2-NTCP cell line was a kind gift of Dr S. Urban (Heidelberg University, Germany).[17] Primary human hepatocytes (PHHs) were isolated from surgically removed liver sections, cultured, and infected as described previously.[15] HepG2-NTCP cells and PHHs were infected at a multiplicity of infection of 100 vge/cell in culture media supplemented with 4% PEG (polyethylene glycol 8000).

Serum samples

Serum samples were collected in the Lyon Biobank after written informed consent from the patients and the protocol was approved by the Lyon institutional Ethic Committees (#DC-2008-235, CPP 11/040 20110530 A11-168, CNIL n°1789480v0).

RNA ligase-mediated 5'RACE

RNAs were isolated using a guanidinium thiocyanate–phenol–chloroform extraction protocol (TRI reagent Sigma[18]). 5'RACE was essentially performed as described in the GeneRacer Kit manual (ThermoFisher Scientific) except Tobacco Acid Pyrophosphatase was substituted by RNA 5' Pyrophosphohydrolase (New England Biolabs) and SuperScript reverse transcriptase III by SuperScript reverse transcriptase IV (ThermoFisher Scientific). The reverse transcription reaction was performed using 3' HBV specific Gsp1 primer 5'-GGTGCGCAGACCAATTTATG-3'. For the 5'RACE PCR reaction Platinum™ SuperFi™ DNA Polymerase (ThermoFisher Scientific), GeneRacer 5' primer and HBV specific nested primer Gsp2 5'-GTGCACACGGTCCGGCAGATG-3' were used (Fig. 1B). 5'RACE PCR was run in a iTOUCH 1000 BIORAD thermocycler using the following touch down PCR program: Initial denaturation step 98°C 3 min >5x (98°C 10 s; 72°C 3 min) >5x (98°C 10 s; 70°C 3 min) >25x (98°C 10 s; 64,4 20 s; 72°C 3min) >72°C 10min. pg and preC specific PCRs were run using primer Gsp2.2 5'-GCTTCCCGATACAGAGCTGAGG-3' in combination with either pg ACTGAAGGAGTAGAAAAACTTTTTCACCTCTG or preC ATGGACTGAAGGAGTAGAAAATAAATTGGTCTGCG specific primer in a iTOUCH 1000 BIORAD thermocycler using the following 2-step PCR program: Initial denaturation step 98°C 3 min >34x (98°C 10 s; 72°C 30 s) >72°C 10 min.

Cloning of 5'RACE amplicons

5'RACE amplicons were A-tailed and cloned into the pGEM-T vector (Promega) using standard procedures. Clones were verified by digestion and sequenced by Sanger sequencing at Eurofins Genomics.

Drug treatments of culture hepatocytes

Interferon-β (PBL Assay Science) was used at 500 IU/ml, preS1-myr (Myrcludex-like) peptide (sequence: GTNLSVPNPLGFFPDHQLDPA-FRANSNNPDWDFNPN KDHWPEANKVG, synthesized by GeneScript) was used at 500 nM, tenofovir (Sigma Aldrich) was used at 100 μM. Drug regimens are detailed in the corresponding figures.

ELISA for viral antigens

Supernatants were 10x diluted prior to ELISAs. ELISAs for HBeAg detection in cell supernatants were performed according to the manufacturer's protocol using the CLIA kit from Autobio Diagnostic.

Protein isolation and western blotting

Whole cell proteins were isolated by a first lysis with buffer A (20 mM Tris-HCL pH7.5; 0. 5mM EDTA; 0.1% TritonX100; 100 mM NaCl; 10% Glycerol; 2 mM MgCl2; 10 mM mercaptoethanol; 0. 5mM PMSF [phenylmethylsulfonyl fluoride]) for 30 min at 4°C. Subsequently, the lysate was centrifuged for 5 min at 4,000 rpm using a tabletop centrifuge. Supernatant was transferred to a fresh tube and the pellet was incubated with 5 volumes of buffer B (20 mM Tris-HCL pH7.5; 0. 5mM EDTA; 0.1% TritonX100; 400 mM NaCl; 10% Glycerol; 2 mM MgCl2; 10 mM mercaptoethanol; 0. 5mM PMSF) for 30 min on 4°C and then centrifuged for 10 min at 13,000 rpm. Both supernatants were pooled. One optical density per condition determined by Bradford assay was diluted in Laemmli buffer and resolved on an SDS-PAGE gel.

For the detection of surface proteins in gradient fractions, 10 μl of fractions were directly diluted in Laemmli buffer and resolved on an SDS-PAGE gel. Proteins were electroblotted onto a nitrocellulose membrane using the Transblot system (Biorad). Surface proteins L/M/S were detected by the monoclonal antibody H166,[19] a kind gift of Abbott Laboratories.

Iodixanol gradient sedimentation analysis

To analyze viral particles, 200 ml of HepAD38 cell supernatant was centrifuged for 5 min at 500 g to sediment cell debris. This was followed by an additional centrifugation step for 45 min at 2,000 g. Supernatant was laid over a 20% sucrose cushion and centrifuged at 90,000 g for 2 h using a Beckmann 32Ti rotor to pellet viral particles. The pellet was resuspended in 1 ml PBS and material was loaded onto a 10 ml 10% to 50% Iodixanol gradient in PBS and centrifuged at 175,000 g for 14 h in a Beckmann SW41 rotor. 900 μl fractions were collected from the top of the gradient.

MNase digestion of HepAD38 supernatants

To isolate encapsidated viral RNAs we performed MNase (Micrococcal Nuclease) digestions of the HepAD38 supernatants. MNase is a DNA and RNA endonuclease that degrades double- and single-stranded DNA and RNA. 200 ml of HepAD38 supernatant was centrifuged for 5 min at 500 g to sediment cell debris. This was followed by an additional centrifugation step for 45 min at 2,000 g. The supernatant was laid over a sucrose cushion and centrifuged at 90,000 g for 2 h using a Beckmann 32Ti rotor to pellet viral particles. Viral particles were resuspended in MNase buffer, 2 μg of a non-relevant plasmid (pcDNA3 vector + insert) was added to monitor digestion (Fig. S5). MNase digestion was performed for 25 min at 37°C using 30 gel units (New England Biolabs) in a 150 μl reaction. Reactions were stopped by Trizol.

qPCRs

RNA or DNA was extracted using the MasterPure complete kit (Epicentre) following manufacturer's instructions. qPCRs were performed essentially as described in Lebossé *et al.*[20] Three kilobase RNAs were detected by the primer pair ggagtgtg-gattcgcactcct (A) and agattgagatcttctgcgac (B) together with the Taqman hybridization probe ([6FAM]-AGGCAGGTCCCCTAGAA-GAAGAACTCC-[BHQ1]).[21] Total RNAs were measured using primers and Taqman probe Pa03453406_s1 (ThermoFisher Scientific). Signals were normalized by the housekeeping gene GUSB detected with primers and Taqman probe Hs999 99908_m1 (ThermoFisher Scientific). Total DNA was measured using primers and Taqman probe Pa03453406_s1 (Life Technologies). Serial dilutions of a plasmid that contains a HBV monomer (pHBVEcoR1) served as an external quantification standard. The amount of total DNA was normalized to the cell number using primers and Taqman probe Hs00758889_s1 (ThermoFisher Scientific) that are specific for the β-globin locus. PCRs were performed in duplicates. Amplification was carried out on the QuantStudio 7 Flex with the following PCR conditions: Initial denaturation step 95°C 20 s >40x (95°C 1 s; 60°C 20 s).

Accession code for sequencing data
All Sanger sequencing data is available at the European Nucleotide Archive at http://www.ebi.ac.uk/ena/data/view/PRJEB36101 (accession code PRJEB36101).

Results

Detection of all major viral RNAs by HBV full-length 5'RACE

In the context of the HBV genome, transcription of 5' positioned genes obscures the measurement of transcripts starting further downstream because all transcripts share the same 3' ends (Fig. 1A). Currently, the measurement of viral RNAs relies on 3 techniques: Northern blotting, reverse transcription (RT)-qPCR and RNA-next generation sequencing (NGS). Northern blotting allows the resolution of viral RNAs according to their molecular weights on an agarose gel but misses the HBx transcript under infection conditions.[22] qRT-PCR cannot distinguish between the different viral transcripts and hence it is commonly used exclusively for measuring total RNAs *vs.* 3. 5kb RNAs. Similarly, standard short read RNA-NGS approaches cannot distinguish the different viral transcripts as soon as more than 1 viral transcript are present. To our knowledge, there is only 1 publication in which the authors tried to discriminate all the different viral transcripts expressed in cell culture and in patients by a strategy called cap analysis of gene expression (CAGE).[23] However, the major limitation of this study was that only the first 50 nucleotides of the viral transcripts were sequenced and thus no conclusions could be drawn about viral full-length transcripts and especially pgRNA splice variants.

Thus, to specifically detect and characterize the different intra- and extracellular HBV RNAs in full-length we set up the HBV full-length 5'RACE strategy. The technique follows a standard RNA ligase-mediated 5'RACE protocol including initial dephosphorylation and de-capping steps to ensure the amplification of only full-length transcripts with respect to 5' ends. Truncated, uncapped RNA species are eliminated from the subsequent 5'RACE amplification process because they miss the 5' phosphate group that is essential for anchor ligation (Fig. 1B). After performing the HBV full-length 5'RACE, the different HBV transcripts are resolved on an agarose gel according to their molecular weights.

The initial experiment was carried out in HepG2-NTCP cells that had been infected for 8 days. As shown in Fig. 1C, we found all major transcripts including HBx migrating at their expected sizes on the gel. Their relative sizes are reduced in length by approximately 338 bp with respect to template viral RNAs because of the use of gene specific 3' primer 2 (Gsp2) for PCR. PCR products are specific to viral RNAs as shown by performing the 5'RACE without the reverse transcription reaction (Fig. S1A) and by their absence in the non-infected condition (Fig. 1C). In addition, PCR products are specific to capped viral RNAs as demonstrated in Fig. S1C and further below. Of note, the proportional quantities of the different viral RNAs are not directly translated by the 5'RACE PCR reaction. This is because the viral transcripts exhibit an important heterogeneity in their respective lengths and, hence, are subject to different amplification efficiencies during the 5'RACE PCR reaction, *i.e.*, gel band intensities of the 5'RACE PCR product do not provide absolute information on the original proportional quantities of the different viral RNA species.

We then investigated the expression kinetics of the different viral RNA species during the course of infection. In both HepG2-NTCP cells and PHHs we found HBx to be the first transcript detectable by full-length 5'RACE as early as 8 h post-infection (Fig. 1D). The remaining transcripts are detected later during infection, in HepG2-NTCP cells from 24 h onwards and in PHHs after a longer delay (not earlier than 48 h, see additional experiment Fig. S2). The variance in the kinetics of viral RNAs observed for the 2 experiments shown for PHHs (Fig. 1D and Fig. S2) probably resulted from the heterogeneity of cells retrieved from 2 different donors. However, these results are similar to the data published in Niu *et al.* where the authors followed viral RNAs in PHHs during the course of infection by an RNA-NGS approach.[22]

Short versions of HBx transcripts are produced in HepG2-NTCP cells and PHHs

To confirm their identity, the amplicons obtained were fully characterized by subcloning and Sanger sequencing (Fig. 1E and PRJEB36101). Sequencing of the 5'RACE amplicons revealed an important diversity in transcript start sites (TSSs) for the HBx transcripts expressed in HepG2-NTCP cells and PHHs (red bars in Fig. 1E). In contrast, all sequenced clones for the preS2 transcripts were identical and share the same 5' end in PHHs and HepG2-NTCP cells (TSS nucleotide [nt] 3,159 with respect to reference sequence GenBank U95551.1, blue bars in Fig. 1E). The same was true for pgRNA (TSS nt1,820) and preC (TSS nt1,791). Surprisingly, we were not able to detect spliced pgRNA variants, most likely because of their low proportional copy numbers with respect to the other viral RNA species in HepG2-NTCP cells and PHHs.

In HepG2-NTCP cells, HBx transcripts can be subdivided into 2 categories: i) "canonical" transcripts initiating closely upstream (TSSs nt1,243 to nt1,338) of the first ATG of HBx (light red) coding for the full-length X protein (154 amino acids) and ii) transcripts starting after the first ATG of HBx (dark red) presumably coding for a short version of X potentially using 2 ATGs further downstream as start codons (Met79/Met103). In PHHs, we additionally sequenced a third category of longer HBx transcripts (TSS nt1065 and nt1151, very light red Fig. 1E) that we did not detect in HepG2-NTCP cells. The longer transcripts include ATGs from the P protein open reading frame and could potentially code for its RNAseH subdomain. The number of sequenced clones positive for shorter and longer HBx transcripts was much smaller compared to clones harboring canonical HBx (see Fig. 1E), indicating that shorter and longer HBx transcripts are most likely less abundant in cells.

Tenofovir treatment primarily increases the pgRNA transcript in PHHs and HepG2-NTCP cells

Subsequently, to further validate our approach, we investigated the effect of treatment interventions on viral transcripts in HBV-infected HepG2-NTCP cells and PHHs using the full-length 5'RACE assay. We first exposed the cells to a Myrcludex-like peptide (hereafter Myrcludex) and functionally well-characterized drugs interfering with the viral life cycle at different stages. Myrcludex is a viral entry inhibitor peptide binding to the HBV receptor sodium taurocholate co-transporting polypeptide (NTCP) which efficiently prevents viral infection.[17,24] Tenofovir, on the other hand, is a nucleotide analog that interferes with DNA chain elongation during the reverse transcription reaction and, accordingly, inhibits viral

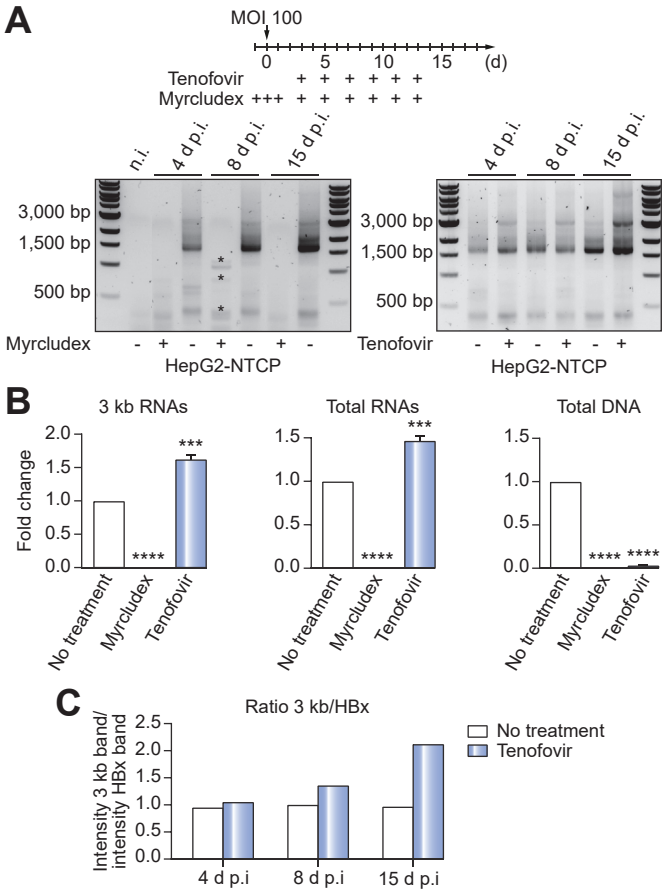


Fig. 2. Tenofovir treatment increases full-length transcripts in HepG2-NTCP cells. (A) Full-length 5'RACE amplicons using HepG2-NTCP cells after 4, 8 and 15 days of infection. Myrcludex drug regime was initiated 1 day before infection. Tenofovir treatment was started 3 days p.i. No specific signals are detected by 5'RACE in the Myrcludex condition as verified by sequencing (asterisks: non-specific signals). Tenofovir leads to an increase in full-length transcripts as compared to non-treated cells. (B) Quantitative PCRs measuring 3 kb RNAs (preC and pgRNA), total RNAs and total DNA in treated *vs.* control cells. (C) 3 kb ratio (intensity 3 kb band/intensity HBx band/lane) is increased and further increases over time in the tenofovir condition as compared to untreated cells. One representative experiment is shown. Experiments were performed in triplicates. Student's *t* test *p* values with respect to no treatment: ***<0.001, ****<0.0001. HBx, X protein RNA; pg, pregenomic; preC, preCore; RACE, rapid amplification of cDNA ends.

genome replication.[25] For HepG2-NTCP experiments, cytotoxicity of drug regimens was excluded by visual inspection. For experiments in PHHs, the effect of drug regimens on cell viability was measured by the CellTiter-Glo approach (Fig. S3). As shown in Fig. 2A, we measured viral RNAs by 5'RACE at 4, 8 and 15 days post infection. Myrcludex treatment was initiated 1 day before infection to saturate NTCP receptors, tenofovir treatment was started 3 days post infection (see timescale in Fig. 2A). After 8 days of infection, conventional qPCRs were performed to monitor their effect on 3 kb RNAs (pg and preC RNA), total RNAs and total DNA (Fig. 2B). In the Myrcludex condition, no viral RNAs and DNAs were detectable by qPCRs. Tenofovir treatment, on the other hand, led to a ~1.5-fold increase in 3 kb RNAs and total RNAs and to a strong decrease in viral DNAs.

Concordantly, full-length 5'RACEs showed no specific signal when Myrcludex blocks viral entry (Fig. 2A, upper gel). The bands visible in the Myrcludex lane 8 days post infection are non-specific, as verified by cloning and sequencing (data not shown). However, tenofovir treatment leads to an increase in 3 kb 5'RACE signals indicating more preC/pgRNA than in the non-treated control conditions (Fig. 2A, lower gel). Tenofovir administration had no effect on the S and HBx transcripts, especially from day 8 on. Since the experiment is exclusively controlled by using equal amounts of input RNA for the different conditions, we normalized the quantity of the 3 kb band (preC/pgRNA) in each lane by its corresponding HBx band. This procedure produces the 3 kb ratio (intensity 3 kb band/intensity HBx band) and shows that, for each time-point, the 3 kb ratios are elevated, *i.e.* more preC/pgRNAs are found in tenofovir treated than untreated cells (Fig. 2C). Moreover, we observed an increase of the 3 kb ratio over time supporting the concept that viral full-length transcripts accumulate during tenofovir treatment in HepG2-NTCP cells.

Like the experiments in HepG2-NTCP cells, no 5'RACE signals are detected in PHHs treated with Myrcludex (Fig. 3A,B). In PHHs too, tenofovir treatment causes an exclusive increase in the 3 kb band compared to the other 5'RACE products (*i.e.* S transcripts and HBx transcripts) during infection (Fig. 3A-C). Next, in order to elucidate whether the increase in 3 kb transcripts reflects an augmentation of pgRNA, preC RNA or both, we designed primers that are specific for pg- or preC RNA using full-length 5'RACE cDNAs (for specificity and outline of strategy see Fig. S4). pg/preC RNA specific 5'RACE PCRs indicate that the effect of tenofovir is much more pronounced for pgRNA than for preC RNA on day 8 post infection, an observation that is consistent with the mode of action of tenofovir as a reverse transcription inhibitor (Fig. 3A, gels pg and preC). To support the results, we quantified HBeAg by ELISA as another read out for preC using the supernatants of treated and untreated cells. Consistently, we did not detect any effect of tenofovir on extracellular HBeAg by ELISA (Fig. 3D). Furthermore, we showed by western blotting that cellular S protein levels arenot affected by tenofovir administration (Fig. 3E). In summary, these results validated the HBV full-length 5'RACE technique and further confirmed that tenofovir treatment primarily increases pgRNA but has no effect on preC RNA, S transcripts and HBx RNA.

Interferon-β treatment reduces both pg and preC transcripts without affecting other viral transcripts in PHHs

Patients with chronic Hepatitis B are treated with interferon-α and nucleos(t)ides analog leading to the suppression of viral replication.[25] In HepG2 or chicken hepatoma cell lines that express viral RNAs from transfected or integrated plasmids, interferon-α efficiently downregulates viral RNA transcription by reducing active chromatin markers like H3K9ac and H3K27ac from HBV and DHBV chromatin.[26,27] We wanted to take advantage of the full-length 5'RACE and investigate how interferon-β affects the different viral RNAs under infection conditions in PHHs. Interferon-β is a type I interferon which utilizes the same receptor as interferon-α, which has been shown to inhibit HBV genome replication in hepatocyte culture.[28,29] Interferon-β treatment was started on day 3 post infection, 5'RACEs (Fig. 4A) and control qPCRs (Fig. 4B) were carried out on day 8 post infection. Surprisingly, we found that interferon-β treatment predominantly reduces viral 3 kb transcripts *i.e.* preC- and pgRNA without significantly affecting the other viral transcripts.

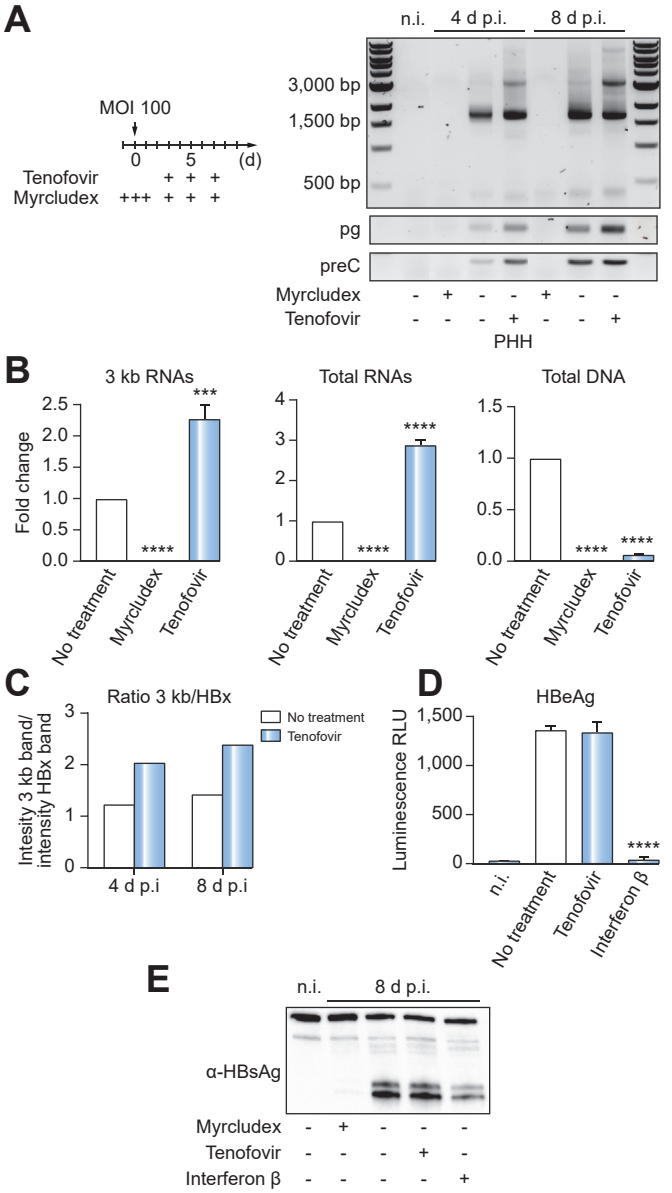


Fig. 3. Tenofovir treatment primarily augments the pgRNA transcript without affecting the other viral transcripts in PHHs. (A) Full-length 5'RACE and preC/pg RNA-specific 5'RACE amplicons using PHHs after 4 and 8 days of infection. Myrcludex drug regime was initiated 1 day before infection. Tenofovir treatment was started 3 days p.i. No specific signals are detected by both the 5'RACEs in the Myrcludex condition. Tenofovir primarily increases the pgRNA transcript (see preC/pg RNA-specific 5'RACEs) without affecting the other viral transcripts (see full-length 5'RACEs). (B) Quantitative PCRs measuring 3 kb RNAs (preC and pgRNA), total RNAs and total DNA in treated *vs.* control cells. (C) 3 kb ratio (intensity 3 kb band/intensity HBx band/lane) is increased and further increases over time in the tenofovir condition as compared to untreated cells. One representative experiment is shown. (D) ELISAs against HBeAg with supernatants of PHHs treated with tenofovir and interferon-β as compared to n.i. and untreated PHHs. Tenofovir treatment does not change HBeAg levels whereas interferon-β strongly reduces HBeAg. (E) Western blot against surface proteins (HBsAgs) using antibody H166 (Abbott) recognizing a conserved epitope in the antigenic loop after 8 days of Myrcludex, tenofovir or interferon-β treatment. Tenofovir treatment does not affect surface protein levels. Interferon-β treatment moderately affects surface protein levels. Experiments were performed in triplicates. Student's *t* test *p* values with respect to no treatment: ***<0.001, ****<0.0001. HBx, X protein RNA; pg, pregenomic; PHHs, primary human hepatocytes; preC, preCore; RACE, rapid amplification of cDNA ends.

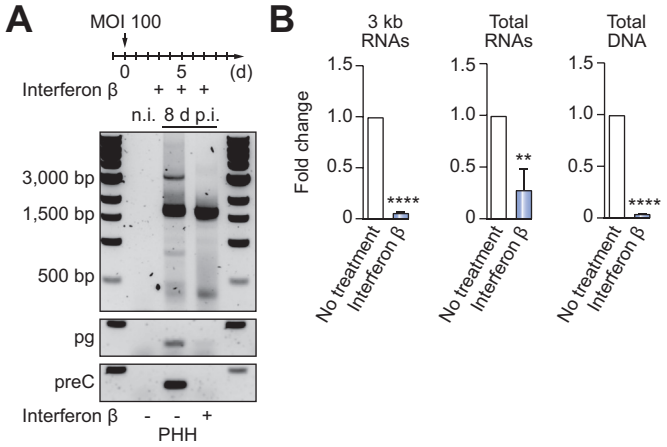


Fig. 4. Interferon-β treatment reduces both pg and preC transcripts without affecting the other viral transcripts in PHHs. (A) Full-length 5'RACE and preC/pg RNA-specific 5'RACE amplicons using PHHs after 8 days of infection. Interferon-β treatment was initiated 3 days p.i. Interferon-β treatment decreases both pg- and preC- transcripts (see preC/pg RNA-specific 5'RACEs) without affecting the other viral transcripts, like HBx (see full-length 5'RACE). (B) Quantitative PCRs measuring 3 kb RNAs (preC and pgRNA), total RNAs and total DNA in treated *vs.* control cells. Experiments were performed in triplicates. Student's *t* test *p* values with respect to no treatment: **<0.01, ****<0.0001. HBx, X protein RNA; pg, pregenomic; PHHs, primary human hepatocytes; preC, preCore; RACE, rapid amplification of cDNA ends.

This prominent effect is observed by full-length 5'RACE and preC/pgRNA-specific 5'RACE PCRs (Fig. 4A).

To complement the 5'RACE analysis, we showed by western blotting that the S protein levels are only moderately affected in the interferon condition (Fig. 3E). In contrast, extracellular HBeAg was strongly reduced, as determined by ELISA (Fig. 3D). We omitted measuring secreted HBsAgs in supernatants by ELISA since it had been shown that Interferons inhibit HBV virion secretion by inducing tetherin expression.[30]

Viral particles produced in HepAD38 cells contain different HBx transcripts, at least 1 spliced pgRNA variant and pgRNA

Recently it was shown that viral particles produced in cell culture or circulating in the blood of chronically infected patients may harbor viral RNA species.[22,31,32] To gain more insight into the type of viral RNAs that may be contained in these particles, we performed 5'RACEs with material derived from supernatants of HepAD38 cells (see Fig. 5A for purification steps). We focused our work on viral particles produced by the HepAD38 cell line since HepG2-NTCP cells and PHHs in infection assays (Fig. 1-Fig. 4) were infected with viral inocula generated by the same cell line.

Prior to full-length 5'RACE analysis, the pelleted material was digested with MNase to detect exclusively protected RNAs (see Fig. S5 for digestion efficacies). As shown in Fig. 5B/C in the supernatant of HepAD38 cells, we found pgRNA, 1 pgRNA spliced variant (known as splice variant 6 (SP6),[33] splice donor site nt2,471/nt2,472; splice acceptor site nt488/nt499, Fig. 5C) and HBx transcripts with different lengths (long, canonical, and short) that are protected from MNase digestion. For their identification, we characterized all RNAs that were protected from MNase degradation (Fig. 5B) by cloning and Sanger sequencing

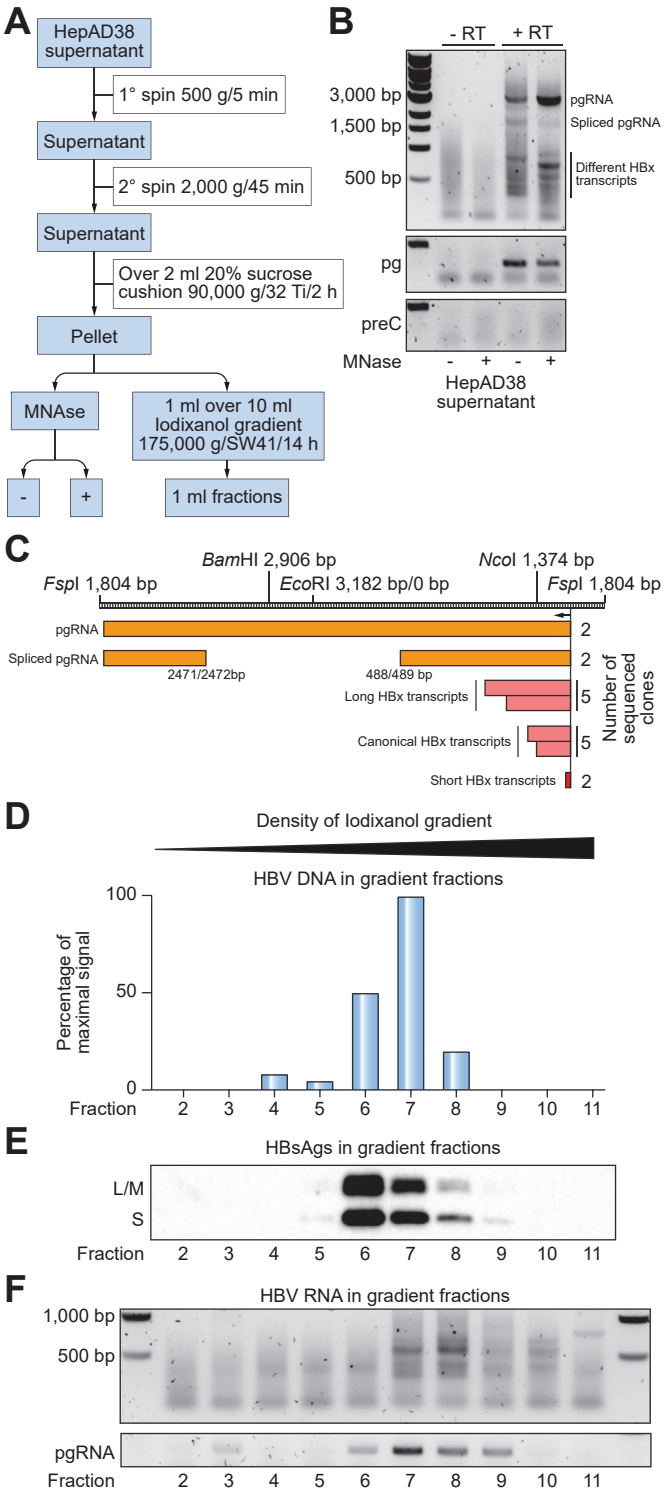


Fig. 5. Viral particles produced in HepAD38 cells contain different HBx transcripts, at least one spliced pgRNA variant and pgRNA. (A) Schematic diagram of procedures used to obtain viral particles. (B) Full-length 5'RACE and preC/pg RNA-specific 5'RACE amplicons using HepAD38 supernatant precipitated through sucrose cushion, with/without MNase digestion, with/without RT (reverse transcriptase) domain. All amplicons were sequenced by subcloning and Sanger sequencing (Table S2). Different HBx RNAs (long, normal and short), pgRNA and one pgRNA splice variant were identified. Neither preC RNA nor S transcripts were detected. (C) Alignment of sequenced amplicons with respect to HBV genome genotype D (reference sequence GenBank U95551.1). To the right, total number of sequenced clones per transcript category. (D) Total viral DNA in gradient fractions of Iodixanol gradient. Viral DNA fractionated between fractions 4 and 8 with a sharp peak in fraction 7. (E) HBsAgs in Iodixanol gradient. All 3 HBsAgs co-sedimented between fractions 5 and 9 with the strongest intensities found in fractions 6 and 7 as shown by western blotting. (F) Full-length 5'RACE detects the majority of long, normal and short HBx transcripts in gradient fractions 7 and 8 co-fractionating with viral cDNA and S proteins. pgRNA-specific 5'RACE detects pgRNA in gradient fractions 6 to 9 with the highest abundance in fraction 7 similar to cDNA. HBx, X protein RNA; MNase, micrococcal nuclease; pg, pregenomic; PHHs, primary human hepatocytes; preC, preCore; RACE, rapid amplification of cDNA ends; S, small surface protein.

(Fig. 5C and PRJEB36101). Neither S transcripts nor preC RNA were detected by sequencing. Furthermore, by pgRNA and preC RNA specific 5'RACE analysis, we confirmed that protected RNAs did not include preC RNA (Fig. 5B). Importantly, the 5'RACE products were specific to viral RNAs since there was no signal in the negative control reactions without RT.

To confirm that the MNase-protected RNAs are contained in viral particles, we performed Iodixanol gradient sediment analysis. Overall, 12 fractions were collected from the top, from which fractions 2 to 11 were characterized in detail with respect to viral DNA (Fig. 5D), S antigens (HBAgs, Fig. 5E) and viral 5'RACE RNA content (Fig. 5F). Viral DNA sedimented between fractions 4 and 8 with a sharp peak in fraction 7. All 3 HBsAgs co-sedimented between fractions 5 and 9 with the strongest intensities found in fractions 6 and 7.

By full-length 5'RACE we could detect the different HBx transcripts similar to the MNase approach. The majority of HBx transcripts were detected in fractions 7 and 8, co-fractionating with viral DNA and HBAgs. Most probably because of RNA quantity issues, full-length 5'RACE neither detected pgRNA nor the pgRNA splice variant found by the MNase approach. However, by performing pgRNA-specific 5'RACE analysis using the pgRNA-specific primer, pgRNA was detected in gradient fractions 6 to 9 with the highest abundance in fraction 7 similar to viral DNA.

Taken together, we found pgRNA, 1 pgRNA-derived splice variant (SP6) and different HBx transcripts co-fractionating with HBAgs in gradient fractions where viral DNA is present (Fig. 5C). HBx transcripts are of different lengths: long (TSSs nt1,065 to nt1,198), canonical (TSSs nt1,243 to nt1,338bp) and short (TSSs nt1,418bp to nt1,533).

HBx RNAs in viral particles are both capped and uncapped

Capped RNAs are protected against dephosphorylation, the first step of the full-length-5'RACE protocol that is necessary to exclude degraded or uncapped RNAs from the full-length 5'RACE (Fig. 1B). The interpretation of our data, *i.e.* that the different RNAs detected by the 5'RACE in viral particles are indeed capped, depends on the efficiency of this dephosphorylation step. An inefficient dephosphorylation step could leave uncapped RNAs phosphorylated and thus could produce false positive signals during the full-length 5'RACE. However, uncapped/degraded viral RNAs that are generated by the RNAse H activity of the P protein during cDNA synthesis must be present in viral particles, as discussed by Bai *et al.*[31]

To clarify these issues, we carried out control 5'RACE experiments that specifically detect uncapped RNAs using again MNAse-digested material from HepAD38 cells derived supernatants. The uncapped RNA-specific 5'RACE protocol lacks the

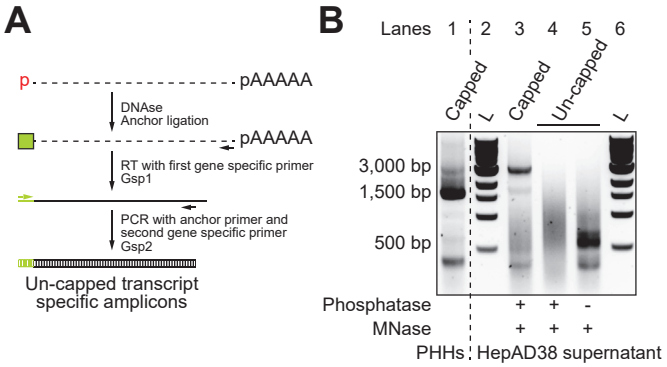


Fig. 6. HBx RNAs in viral particles are both capped and uncapped. (A) Overview of the uncapped RNA-specific HBV full-length 5'RACE protocol. The uncapped RNA-specific 5'RACE protocol lacks the de-capping and dephosphorylation steps from the standard HBV full-length 5'RACE. In a first step, isolated RNAs are digested by DNAse to remove residual DNA molecules. In a next step, the RNA anchor (green box) is ligated to the free phosphate group of uncapped RNAs. Then, the viral RNAs are reverse transcribed using the gene specific primer Gsp1 (black arrow) integrating the anchor sequence into the cDNA. Finally, the different viral cDNAs are amplified by PCR using the second gene specific primer Gsp2 (black arrow) and an anchor specific primer (green arrow) in the same tube. (B) Capped RNA-specific 5'RACE using total RNAs of PHHs (lane 1). Capped and uncapped RNA-specific 5'RACEs using MNAse-digested supernatant of HepAD38 cells (lane 3-5). Presence of different uncapped HBx RNA (amplicons sequenced by subcloning and Sanger sequencing, data not shown) in viral particles (lane 5). The dephosphorylation step of the 5'RACE protocol is very effective (compare lane 4 with lane 5) demonstrating the authenticity of the capped RNA species identified by the HBV full-length 5'RACE in viral particles and in cells (lanes 1 and 3). L (1 kb ladder (NEB)). MNase, micrococcal nuclease; PHHs, primary human hepatocytes; RACE, rapid amplification of cDNA ends.

de-capping and dephosphorylation steps from the standard full-length 5'RACE and thus capped RNAs are not ligated to the anchor sequence and consequently are not amplified by the final PCR reaction (Fig. 6A). To control dephosphorylation efficacy, we included or not a dephosphorylation step in the uncapped RNA-specific 5'RACE protocol (compare lane 4 with lane 5 in Fig. 6B). Simultaneously, we performed the standard full-length 5'RACE (capped RNA-specific) on the same material (Fig. 6B, lane 3) and on cellular RNAs derived from PHHs (Fig. 6B, lane 1).

Collectively, the data in Fig. 6B demonstrates the authenticity of the capped RNAs found in viral particles, since all uncapped RNAs contained in viral particles are dephosphorylated during the capped RNA-specific full-length 5'RACE approach (compare lane 4 with lane 5 in Fig. 6B). However, we also confirm that uncapped RNA species up to 800 bp exist in the MNAse-digested material (Fig. 6B, lane 5). Subcloning and sequencing identified the uncapped RNAs as long, canonical and short HBx transcripts (data not shown). We detected neither pgRNA nor the pgRNA splice variant as uncapped RNAs (Fig. 6B, lane 5).

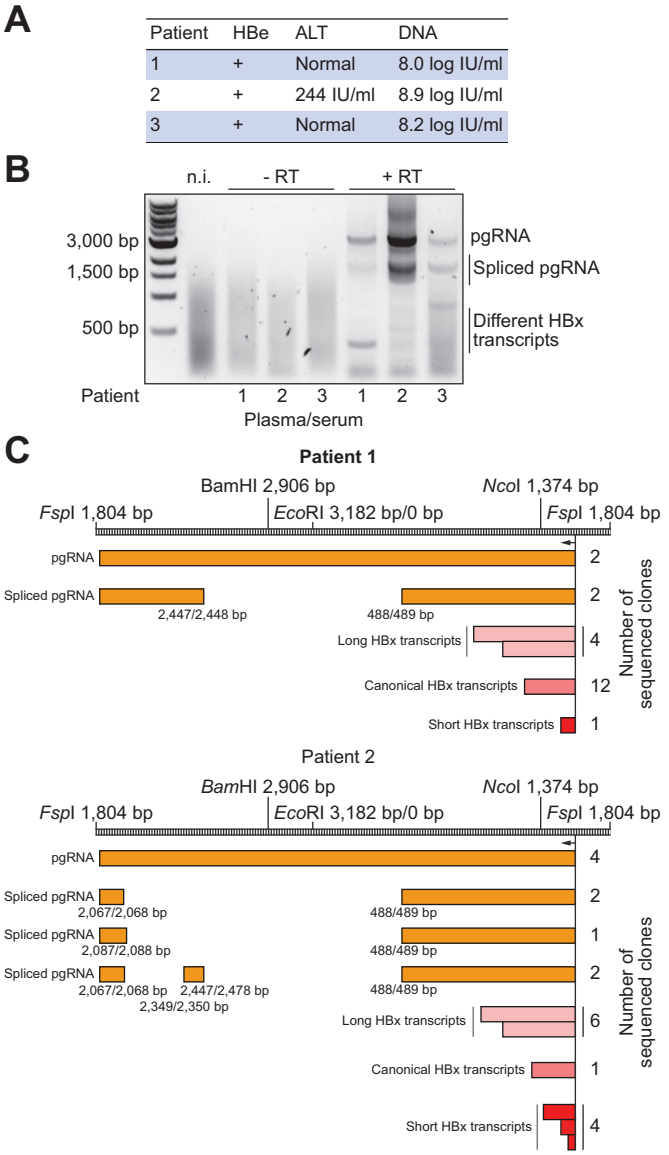


Fig. 7. Full-length 5'RACE identifies 4 major categories of circulating RNA species in plasma and sera of patients. (A) Clinical characteristics associated with the blood samples of patients with chronic hepatitis B used in this study. (B) Full-length 5'RACE amplicons using patient plasmas (patient #1 and #2) and serum (patient #3) plus minus the RT (reverse transcriptase) domain. All amplicons for patients 1 and 2 were sequenced by subcloning and Sanger sequencing (Table S3A,B). Sequences of serological RNAs belong to 1 of the 3 categories: i) pgRNA, ii) pgRNA splice variant, iii) HBx transcript (long, normal and short HBx transcripts). (C) Alignment of sequenced amplicons in patient #1 and #2 with respect to HBV genome genotype D. To the right, total number of sequenced clones per transcript category. HBx, X protein RNA; pg, pregenomic; preC, preCore; RACE, rapid amplification of cDNA ends.

Full-length 5'RACE identifies 3 major categories of circulating RNA species in patients with chronic HBV infection

An accurate monitoring of intrahepatic cccDNA levels and activity during patient management is limited by the need for invasive liver biopsy procedures. So far, no single serum parameter has been shown to accurately reflect the transcriptional activity of the cccDNA pool in the liver.[34] Moreover, it is still an open question, which serological viral RNA species exist in the plasma/serum of patients. Indeed, pgRNA has been detected in sera of patients, spliced pgRNA variants and HBx transcripts as well, but there is not much information about whether and how these transcripts co-exist.[22,32,35–38] To gain deeper insight into these questions, we characterized the major circulating viral RNA species found in 3 highly viremic chronically infected patients (Fig. 7A) by full-length 5'RACE (Fig. 7B). All 3 patients were HBe-positive and patient #2 had elevated alanine aminotransferase (ALT) levels. Fig. 7B shows the different 5'RACE profiles for the 3 patients and visual inspection revealed

a differential pattern between the patients. Importantly, for all patients 5'RACE signals were specific to RNAs (Fig. 7B).

To characterize the RNA species found in plasma, we subcloned and sequenced the 5'RACE products of patient 1 and 2 (Fig. 7C and PRJEB36101). We did not detect any S or preC transcripts in the plasma samples. For both patients 1 and 2, we found co-circulating pgRNA and pgRNA splice variants (known as SP1/2/3/5,[33] see Fig. 7C for splice donor/acceptor sites). We again identified an important heterogeneity for the HBx transcripts including long (TSSs nt991-nt1,197), canonical (TSSs nt1,278-nt1,350) and short (TSSs nt1397-nt1558) HBx transcripts similar to the HBx transcripts associated with *in vitro* produced viral particles. Interestingly, the pgRNA splice variants observed in patients #1 and #2 were different (see Fig. 7C for alignments of transcripts in individual patients). In summary, HBV RNAs found in plasma samples of patients belong to 1 of the 3 categories: i) pgRNA, ii) pgRNA splice variant, iii) HBx transcript (long, canonical and short HBx transcripts). Considering the sequencing data together with the patient specific 5'RACE profiles, the results presented here open the possibility that patients exhibit qualitative and/or quantitative differences in circulating HBV RNAs.

Discussion

We set up an HBV full-length 5'RACE approach to overcome the current limitations for the measurement and discrimination of the different HBV RNAs. The results show that our HBV full-length 5'RACE detects all major viral RNAs including pgRNA splicing variants and most importantly, HBx. The full-length 5'RACE is a qualitative approach, though sensitive enough to demonstrate changes during the course of infection (Fig. 1) and in experimental conditions where cultured hepatocytes were treated with tenofovir (Fig. 2A and 3A) or interferon-β (Fig. 4A).

Contrary to preC/pgRNA and the S transcripts, we found a variety of transcripts for HBx that are characterized by heterogeneous TSSs (Fig. 1D). Of special interest are the long and short HBx transcripts that were found in cells, viral particles produced by cultured hepatocytes and the plasma of patients, indicating their biological relevance. Some of the long and short HBx transcripts could potentially code for other proteins than full-length canonical X. Shorter than canonical HBx transcripts and short X proteins have been discussed in the literature.[23,39,40] Short X starting from Met 103 of the canonical X would lack the DDB1-binding domain of the full-length protein[41] and thus would have lost its co-regulatory activity of the Cul4A E3 ubiquitin ligase complex.[12,41,42] In addition, the C-terminal domain of full-length X is required for p53 interaction[13] and consequently, short X could be a protein that predominantly interacts with p53.

On the other hand, long HBx transcripts have not yet been described. We could detect them in plasma, viral particles produced by the HepAD38 cell line and in cell extracts from PHHs, but surprisingly not in HepG2-NTCP cells. Long HBx transcripts could code for the RNAseH domain of the P protein since the transcripts include ATGs upstream of HBx. Whether the long transcripts code for RNAseH and what function the RNAseH domain has when separated from its reverse transcriptase domain requires further investigation. Interestingly, it has been shown that the HBV RNAseH domain is a stable protein without the reverse transcriptase domain,[43–46] a characteristic that is not true for all the RNAseH domains of other reverse transcriptases.[47,48]

Besides coding for proteins, there is also the possibility that the short and long HBx RNAs function as regulatory non-coding RNAs (ncRNAs). Many DNA and RNA viruses express ncRNAs that target either host or viral factors.[49] Especially long non-coding RNAs (lncRNAs) are an interesting class of ncRNAs that regulate host or viral gene expression at the level of chromatin, transcription, post-transcription and translation. Interestingly, for the Kaposi's sarcoma-associated herpesvirus, Japanese encephalitis virus, dengue virus, and West Nile virus it has been shown that lncRNAs antagonize the antiviral interferon response.[50–55]

The capacity of the full-length 5'RACE to discriminate all the major viral transcripts allowed us to show that viral RNAs can be differentially regulated. We demonstrated that interferon-β reduced preC RNA and pgRNA without affecting the other viral transcripts (Fig. 4A). In overexpression assays[26,27,56] and mouse model infection systems,[57] this differential phenotype could not be differentiated from a general downregulation of viral transcripts. Moreover, in combination with the qPCR assays (Fig. 4B), the full-length 5'RACE data also indicates that preC- and pgRNAs constitute the majority of viral transcripts in infected PHHs confirming CAGE-NGS results recently published by Altinel *et al.*[23]

It will be important to elucidate how interferon-β induces the specific downregulation of preC/pgRNAs. One possible scenario could be that interferon-β treatment induces the specific degradation of the full-length viral transcripts in PHHs. Interestingly, interferon stimulated gene 20 (ISG20) degrades HBV viral RNAs by binding to the epsilon stem-loop structure.[58] If the targeting/degradation activity of ISG20 in PHHs would be specific for the epsilon structure at the 5' end of an RNA molecule, only full-length viral RNAs (preC/pgRNAs) would be degraded. However, Liu *et al.* demonstrated by over-expression assays that ISG20 degrades viral RNAs irrespective of whether the epsilon structure is at the 5' or 3' end of the transcripts. The authors additionally demonstrated that ISG20 does not act alone but requires co-factors for its target specificity. Thus, either another ISG displaying different targeting characteristics is induced upon interferon-β treatment or, the specificity of ISG20 is modulated by different co-factors in PHHs.

Another possibility could be that interferon-β specifically reduces the activity of the HBV core promoter regulating preC/pgRNAs expression. It has been shown that the core HBV promoter is negatively regulated by the NF-κβ factor p65,[59,60] another rapidly induced ISG, which could preferentially downregulate the HBV core promoter once activated by interferon-β. Alternatively, the ISG TRIM22 can suppress HBV core promoter activity in HepG2 cells. TRIM22 is a member of the tripartite motif (TRIM) superfamily whose members have been shown to be expressed in response to interferons to restrict viral activity, especially with respect to retroviruses.[61]

We also characterized the full spectrum of viral RNAs associated with viral particles and gained a detailed overview of their composition (Figs. 5, 6). Our results integrate earlier published results, but at the same time provide new insights into the diversity of particle-associated viral RNAs. For instance, in addition to pgRNA we also detect spliced pgRNA variants in viral particles. Spliced pgRNA variants were first described in cell lines and infected livers[62–65] and were later shown to be incorporated into the nucleocapsids and then reverse transcribed into HBV DNA to generate defective HBV particles.[33,66–69] Interestingly, it also has been shown that pgRNA splice variants can code for additional

viral proteins.[70,71] Furthermore, our results substantiate the notion that viral particles contain HBx RNA that may have an important role for the establishment of infection as proposed by Niu *et al.*[22] In this regard, we found that different HBx RNAs exist as capped and uncapped molecules in viral particles, an observation that raises the question of whether one is the product of the other (Fig. 6B). As depicted in Fig. 1A, the 6 viral RNAs are transcribed from 4 different promoters. After productive initiation of polymerase II, nascent transcripts are co-transcriptionally capped in the nucleus by different capping enzymes. Capping of transcripts promotes translation and protects against RNA degradation by cellular exonucleases. The capped non-canonical HBx RNAs are puzzling with respect to their 5'ends since, to our knowledge, promoters that initiate transcription from these positions in the HBV genome have not yet been molecularly characterized. In addition, transcripts that lack the 5' epsilon stem-loop structure should be excluded from viral capsids.[72–75] Therefore, the question arises whether the capped HBx transcripts in viral particles could be produced post- rather than co-transcriptionally. The hypothetical post-transcriptional generation of encapsidated HBx RNAs would require 2 major steps: i) an endolytic cut to generate the 5'end of the RNA molecule and ii) the addition of the cap structure. A scenario would be that the RNAse H domain of the P protein produces specific endolytic cuts in the pgRNA during reverse transcription in viral particles giving rise to the 5' ends of the different HBx RNAs. Interestingly, the RNAse H domain of the avian myeloblastosis virus reverse transcriptase *e.g.* initially cuts every 100–200 nucleotides during reverse transcription. Whereas for HIV-1 and Moloney murine leukemia virus, the cleavage frequency centers around 100–120 nucleotides.[48,76] Consequently, it will be important to determine the cutting frequency of the HBV RNAse H domain and whether it correlates with the molecular sizes of the different HBx transcripts (400–800 nts). Furthermore, the post-transcriptional generation of capped non-canonical HBx RNAs would also require that the capping of uncapped 5' ends takes place inside viral particles. Since HBV does not code for its own capping enzymes, viral particles would have to contain host-capping enzymes that fulfill this task. Notably, for RNA alphaviruses it has been shown that viral particles that contain non-capped viral RNAs induce innate immune responses in host cells and therefore are less infectious than viral particles carrying capped RNAs.[77] It is intriguing to speculate that HBV might have found a way to cap the different RNAs arising during the maturation of viral particles to hide them from the innate immune system of host cells. Importantly, it is very likely that other pgRNA-derived RNA species are generated during the reverse transcription process in addition to the different HBx RNAs characterized in this study. However, the HBV full-length 5'RACE method is specific to RNA molecules that are intact until nt1,810 (Gsp1 annealing point) of the HBV genome (reference sequence GenBank U95551.1). Our approach is blind to RNA species that do not include this distal 3' sequence, *i.e.* to all RNAs that have been released from the pgRNA by 2 cuts upstream of this reference point during reverse transcription. Whether these RNA species are also capped, similar to the HBx transcripts that were identified in this study, remains to be investigated.

In recent years, many studies evaluated circulating pgRNA as a biomarker for disease development and treatment efficacy. A good correlation between serological and intrahepatic pgRNA levels, as well as between circulating pgRNA and the ratio between intracellular pgRNA and cccDNA have been observed. Therefore, serological pgRNA could represent a non-invasive read out for monitoring intrahepatic cccDNA transcriptional activity, even during nucleos(t)ide analog treatment.[37,78] To get a deeper understanding of the complexity of the circulating viral RNAs, we characterized the majority of serological viral RNAs in 3 patients using the HBV full-length 5'RACE (Fig. 7). Similar to the RNA species associated with viral particles, circulating RNAs belong to 1 of the 3 categories: i) pgRNA, ii) pgRNA splice variant, iii) HBx transcripts (long, canonical and short). Likewise, we detected neither preC- nor S-transcripts in the plasma of patients. Considering the possibility that the different HBx RNAs are post-transcriptionally generated inside viral particles these results support the concept that viral RNAs found in the blood circulation are all derived from packaged pgRNAs and protected against degradation by nucleases in the blood of patients. Consequently, all viral RNAs that are not packaged into viral particles, *i.e.* preC and S transcripts, are not detected by the full-length 5'RACE. At this point, it should be emphasized that the capped RNA-specific 5'RACE approach only detects RNAs that are non-degraded. It excludes all the partially digested viral RNAs that might be present in the blood of patients, especially of those with high ALT levels. However, with our limited subcloning and sequencing approach, we cannot categorically exclude the presence of S or preC RNAs in the plasma of patients. Yet, they might represent minor viral RNA species since we readily detected them by the same cloning and sequencing approach in infected cells.

At last, the 5'RACE profiles of the different patients exhibit differential patterns of PCR bands. The differential PCR patterns together with the identification of specific pgRNA splice variants in different patients demonstrate that the composition of the viral RNA species can vary among patients. Our observations create new avenues for research to determine the biological and clinical relevance of circulating viral RNAs, as novel biomarkers of treatment efficacy, and their potential to decrease the pool of intrahepatic cccDNA.

Abbreviations

ALT, alanine aminotransferase; CAGE, cap analysis of gene expression: cccDNA, covalently closed circular DNA; HBx, X protein RNA; lncRNAs, long non-coding RNAs; MNase, micrococcal nuclease; NGS, next generation sequencing; pg, pregenomic; PHHs, primary human hepatocytes; preC, preCore; preS1, large surface protein; preS2, middle surface protein; RACE, rapid amplification of cDNA ends; RT-qPCR, reverse transcription quantitative PCR; S, small surface protein; TSS, transcript start sites.

Financial support

Evotec and ANRS grants to FZ and BT.

Conflict of interest

This work was part of a collaborative research agreement between INSERM and Evotec.

Please refer to the accompanying ICMJE disclosure forms for further details.

Authors' contribution

B.S. conceived the HBV full-length 5'RACE assay. B.S. performed 5'RACEs, western blots, qPCRs, ELISA, Iodixanol gradient sedimentations, infections of HepG2-NTCP cells and PHHs, cloning of cDNAs, analyzed and interpreted the data. A.D. performed infections of PHHs, qPCRs and analyzed the data. F.C. helped in performing cell culture experiments. M.R. provided liver resections. FZ provided serum samples from patients. G.N., A.A., L.F., K.C., B.T. and F.Z. interpreted the data and supervised the work. B.S. wrote the manuscript with editing input from all authors.

Acknowledgements

The authors would like to thank Maud Michelet, Jennifer Molle, Anaelle Dubois, and Océane Floriot, for their help in the isolation of primary human hepatocytes, as well as Prof. Michel Rivoire's surgical staff for providing liver resections.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhep.2020.01.028.

References

Author names in bold designate shared co-first authorship

[1] Zoulim F, Lebossé F, Levrero M. Current treatments for chronic hepatitis B virus infections. Curr Opin Virol 2016;18:109–116.
[2] Moreau P, Cournac A, Palumbo GA, Marbouty M, Mortaza S, Thierry A, et al. Tridimensional infiltration of DNA viruses into the host genome shows preferential contact with active chromatin. Nat Commun 2018;9:4268.
[3] Nassal M. HBV cccDNA: viral persistence reservoir and key obstacle for a cure of chronic hepatitis B. Gut 2015;64:1972–1984.
[4] Seeger C, Mason WS. Molecular biology of hepatitis B virus infection. Virology 2015;479–480:672–686.
[5] Mitra B, Wang J, Kim ES, Mao R, Dong M, Liu Y, et al. Hepatitis B virus precore protein p22 inhibits interferon-alpha signaling by blocking STAT nuclear translocation. J Virol 2019;93(13). https://doi.org/10.1128/JVI.00196-19.
[6] Cha M-Y, Ryu D-K, Jung H-S, Chang H-E, Ryu W-S. Stimulation of hepatitis B virus genome replication by HBx is linked to both nuclear and cytoplasmic HBx expression. J Gen Virol 2009;90:978–986.
[7] Dandri M, Petersen J, Stockert RJ, Harris TM, Rogler CE. Metabolic labeling of woodchuck hepatitis B virus X protein in naturally infected hepatocytes reveals a bimodal half-life and association with the nuclear framework. J Virol 1998;72:6.
[8] Doria M, Klein N, Lucito R, Schneider RJ. The hepatitis B virus HBx protein is a dual specificity cytoplasmic activator of Ras and nuclear activator of transcription factors. EMBO J 1995;14:4747–4757.
[9] Kornyeyev D, Ramakrishnan D, Voitenleitner C, Livingston CM, Xing W, Hung M, et al. Spatiotemporal analysis of hepatitis B virus X protein in primary human hepatocytes. J Virol 2019;93. https://doi.org/10.1128/JVI.00248-19.
[10] Murakami S. Hepatitis B virus X protein: a multifunctional viral regulator. J Gastroenterol 2001;36:651–660.
[11] Benhenda S, Ducroux A, Riviere L, Sobhian B, Ward MD, Dion S, et al. Methyltransferase PRMT1 is a binding partner of HBx and a negative regulator of hepatitis B virus transcription. J Virol 2013;87:4360–4371.
[12] Decorsière A, Mueller H, van Breugel PC, Abdul F, Gerossier L, Beran RK, et al. Hepatitis B virus X protein identifies the Smc5/6 complex as a host restriction factor. Nature 2016;531:386–389.
[13] Elmore LW, Hancock AR, Chang S-F, Wang XW, Chang S, Callahan CP, et al. Hepatitis B virus X protein and p53 tumor suppressor interactions in the modulation of apoptosis. Proc Natl Acad Sci 1997;94:14707–14712.
[14] Truant R, Antunovic J, Greenblatt J, Prives C, Cromlish JA. Direct interaction of the hepatitis B virus HBx protein with p53 leads to inhibition by HBx of p53 response element-directed transactivation. J Virol 1995;69:9.
[15] Lucifora J, Arzberger S, Durantel D, Belloni L, Strubin M, Levrero M, et al. Hepatitis B virus X protein is essential to initiate and maintain virus replication after infection. J Hepatol 2011;55:996–1003.
[16] Ladner SK, Otto MJ, Barker CS, Zaifert K, Wang GH, Guo JT, et al. Inducible expression of human hepatitis B virus (HBV) in stably transfected hepatoblastoma cells: a novel system for screening potential inhibitors of HBV replication. Antimicrob Agents Chemother 1997;41:1715–1720.
[17] Ni Y, Lempp FA, Mehrle S, Nkongolo S, Kaufman C, Fälth M, et al. Hepatitis B and D viruses exploit sodium taurocholate co-transporting polypeptide for species-specific entry into hepatocytes. Gastroenterology 2014;146:1070–1083.e6.
[18] Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 1987;162:156–159.
[19] Peterson DL, Paul DA, Lam J, Tribby II, Achord DT. Antigenic structure of hepatitis B surface antigen: identification of the "d" subtype determinant by chemical modification and use of monoclonal antibodies. J Immunol 1984;132(2):920–927.
[20] **Lebossé F**, **Testoni B**, Fresquet J, Facchetti F, Galmozzi E, Fournier M, et al. Intrahepatic innate immune response pathways are downregulated in untreated chronic hepatitis B. J Hepatol 2017;66:897–909.
[21] Volz T, Lutgehetmann M, Wachtler P, Jacob A, Quaas A, Murray JM, et al. Impaired intrahepatic hepatitis B virus productivity contributes to low viremia in most HBeAg-negative patients. Gastroenterology 2007;133:843–852.
[22] Niu C, Livingston CM, Li L, Beran RK, Daffis S, Ramakrishnan D, et al. The Smc5/6 complex restricts HBV when localized to ND10 without inducing an innate immune response and is counteracted by the HBV X protein shortly after infection. PLoS One 2017;12:e0169648.
[23] Altinel K, Hashimoto K, Wei Y, Neuveut C, Gupta I, Suzuki AM, et al. Single-nucleotide resolution mapping of hepatitis B virus promoters in infected human livers and hepatocellular carcinoma. J Virol 2016;90:10811–10822.
[24] Lempp FA, Urban S. Inhibitors of hepatitis B virus attachment and entry. Intervirology 2014;57:151–157.
[25] Zoulim F, Locarnini S. Hepatitis B virus resistance to nucleos(t)ide analogues. Gastroenterology 2009;137:1593–1608.e2.
[26] Belloni L, Allweiss L, Guerrieri F, Pediconi N, Volz T, Pollicino T, et al. IFN-α inhibits HBV transcription and replication in cell culture and in humanized mice by targeting the epigenetic regulation of the nuclear cccDNA minichromosome. J Clin Invest 2012;122(2):529–537.
[27] Liu F, Campagna M, Qi Y, Zhao X, Guo F, Xu C, et al. Alpha-interferon suppresses hepadnavirus transcription by altering epigenetic modification of cccDNA minichromosomes. Plos Pathog 2013;9:e1003613.
[28] Lucifora J, Durantel D, Testoni B, Hantz O, Levrero M, Zoulim F. Control of hepatitis B virus replication by innate response of HepaRG cells. Hepatology 2010;51:63–72.
[29] Shen F, Li Y, Wang Y, Sozzi V, Revill PA, Liu J, et al. Hepatitis B virus sensitivity to interferon-α in hepatocytes is more associated with cellular interferon response than with viral genotype. Hepatology 2018;67:1237–1252.
[30] Yan R, Zhao X, Cai D, Liu Y, Block TM, Guo J-T, et al. The interferon-inducible protein tetherin inhibits hepatitis B virus virion secretion. J Virol 2015;89:9200–9212.
[31] Bai L, Zhang X, Kozlowski M, Li W, Wu M, Liu J, et al. Extracellular hepatitis B virus RNAs are heterogeneous in length and circulate as capsid-antibody complexes in addition to virions in chronic hepatitis B patients. J Virol 2018;92. https://doi.org/10.1128/JVI.00798-18.
[32] Wang J, Shen T, Huang X, Kumar GR, Chen X, Zeng Z, et al. Serum hepatitis B virus RNA is encapsidated pregenome RNA that may be associated with persistence of viral infection and rebound. J Hepatol 2016;65:700–710.
[33] Günther S, Sommer G, Iwanska A, Will H. Heterogeneity and common features of defective hepatitis B virus genomes derived from spliced pregenomic RNA. Virology 1997;238:363–371.
[34] **Testoni B**, **Lebossé F**, Scholtes C, Berby F, Miaglia C, Subic M, et al. Serum hepatitis B core-related antigen (HBcrAg) correlates with covalently closed circular DNA transcriptional activity in chronic hepatitis B patients. J Hepatol 2019;70:615–625.
[35] Hacker HJ, Zhang W, Tokus M, Bock T, Schröder CH. Patterns of circulating hepatitis B virus serum nucleic acids during lamivudine therapy. Ann N Y Acad Sci 2004;1022:271–281.
[36] Lam AM, Ren S, Espiritu C, Kelly M, Lau V, Zheng L, et al. Hepatitis B virus capsid assembly modulators, but not nucleoside analogs, inhibit the production of extracellular pregenomic RNA and spliced RNA variants.

Antimicrob Agents Chemother 2017;61. https://doi.org/10.1128/AAC.00680-17.
[37] Liu S, Zhou B, Valdes JD, Sun J, Guo H. Serum hepatitis B virus RNA: a new potential biomarker for chronic hepatitis B virus infection. Hepatology 2019;69:1816–1827.
[38] Wang J, Sheng Q, Ding Y, Chen R, Sun X, Chen X, et al. HBV RNA virion-like particles produced under nucleos(t)ide analogues treatment are mainly replication-deficient. J Hepatol 2018;68:847–849.
[39] Kwee L. Alternate translation initiation on hepatitis B virus X mRNA produces multiple polypeptides that differentially transactivate class II and III promoters. J Virol 1992;66:8.
[40] Zheng Y, Riegler J, Wu J, Yen TS. Novel short transcripts of Hepatitis B virus X gene derived from intragenic promoter. J Biol Chem 1994;269(36):22593–22598.
[41] Hodgson AJ, Hyser JM, Keasler VV, Cang Y, Slagle BL. Hepatitis B virus regulatory HBx protein binding to DDB1 is required but is not sufficient for maximal HBV replication. Virology 2012;426:73–82.
[42] Leupin O, Bontron S, Schaeffer C, Strubin M. Hepatitis B virus X protein stimulates viral genome replication via a DDB1-dependent pathway distinct from that leading to cell death. J Virol 2005;79:4238–4245.
[43] Choi J, Kim EE, Park YI, Han YS. Expression of the active human and duck hepatitis B virus polymerases in heterologous system of Pichia methanolica. Antiviral Res 2002;55:279–290.
[44] Potenza N, Salvatore V, Raimondo D, Falanga D, Nobile V, Peterson DL, et al. Optimized expression from a synthetic gene of an untagged RNase H domain of human hepatitis B virus polymerase which is enzymatically active. Protein Expr Purif 2007;55:93–99.
[45] Tavis JE, Cheng X, Hu Y, Totten M, Cao F, Michailidis E, et al. The hepatitis B virus ribonuclease H is sensitive to inhibitors of the human immunodeficiency virus ribonuclease H and integrase enzymes. Plos Pathog 2013;9:e1003125.
[46] Wei X, Peterson DL. Expression, purification, and characterization of an active RNase H domain of the hepatitis B viral polymerase. J Biol Chem 1996;271:32617–32622.
[47] Champoux JJ, Schultz SJ. Ribonuclease H: properties, substrate specificity and roles in retroviral reverse transcription: retroviral RNases H. FEBS J 2009;276:1506–1516.
[48] Schultz SJ, Champoux JJ. RNase H activity: structure, specificity, and function in reverse transcription. Virus Res 2008;134:86–103.
[49] Tycowski KT, Guo YE, Lee N, Moss WN, Vallery TK, Xie M, et al. Viral noncoding RNAs: more surprises. Genes Dev 2015;29:567–584.
[50] Bidet K, Dadlani D, Garcia-Blanco MA. G3BP1, G3BP2 and CAPRIN1 are required for translation of interferon stimulated mRNAs and are targeted by a dengue virus non-coding RNA. PLoS Pathog 2014;10:e1004242.
[51] Chang R-Y, Hsu T-W, Chen Y-L, Liu S-F, Tsai Y-J, Lin Y-T, et al. Japanese encephalitis virus non-coding RNA inhibits activation of interferon by blocking nuclear translocation of interferon regulatory factor 3. Vet Microbiol 2013;166:11–21.
[52] Qiu L, Wang T, Tang Q, Li G, Wu P, Chen K. Long non-coding RNAs: regulators of viral infection and the interferon antiviral response. Front Microbiol 2018;9:1621.
[53] Rossetto CC, Pari G. KSHV PAN RNA associates with demethylases UTX and JMJD3 to activate lytic replication through a physical interaction with the virus genome. PLoS Pathog 2012;8:e1002680.
[54] Rossetto CC, Pari GS. Kaposi's sarcoma-associated herpesvirus noncoding polyadenylated nuclear RNA interacts with virus- and host cell-encoded proteins and suppresses expression of genes involved in immune modulation. J Virol 2011;85:13290–13297.
[55] Yang X, Li H, Sun H, Fan H, Hu Y, Liu M, et al. Hepatitis B virus-encoded MicroRNA controls viral replication. J Virol 2017;91. https://doi.org/10.1128/JVI.01919-16.
[56] Rang A, Günther S, Will H. Effect of interferon-α on hepatitis B virus replication and gene expression in transiently transfected human hepatoma cells. J Hepatol 1999;31:791–799.
[57] Pasquetto V, Wieland SF, Uprichard SL, Tripodi M, Chisari FV. Cytokine-sensitive replication of hepatitis B virus in immortalized mouse hepatocyte cultures. J Virol 2002;76:5646–5653.
[58] Liu Y, Nie H, Mao R, Mitra B, Cai D, Yan R, et al. Interferon-inducible ribonuclease ISG20 inhibits hepatitis B virus replication through directly binding to the epsilon stem-loop structure of viral RNA. PLOS Pathog 2017;13:e1006296.
[59] Lin Y-C, Hsu E-C, Ting L-P. Repression of hepatitis B viral gene expression by transcription factor nuclear factor-kappaB. Cell Microbiol 2009;11:645–660.
[60] Quasdorff M, Protzer U. Control of hepatitis B virus at the level of transcription. J Viral Hepat 2010;17:527–536.
[61] Gao B, Duan Z, Xu W, Xiong S. Tripartite motif-containing 22 inhibits the activity of hepatitis B virus core promoter, which is dependent on nuclear-located RING domain. Hepatology 2009;50:424–433.
[62] Chen P-J, Chen C-R, Sung J-L, Chen D-S. Identification of a doubly spliced viral transcript joining the separated domains for putative protease and reverse transcriptase of hepatitis B virus. J Virol 1989;63:7.
[63] Su T-S, Lui W-Y, Lin L-H, Han S-H, P'eng F-K. Analysis of hepatitis B virus transcripts in infected human livers. Hepatology 1989;9:180–185.
[64] Su T-S, Lai C-J, Huang J-L, Lin L-H, Yauk Y-K, Chang C, et al. Hepatitis B virus transcript produced by RNA splicing. J Virol 1989;63:8.
[65] Suzuki T, Masui N, Kajino K, Saito I, Miyamura T. Detection and mapping of spliced RNA from a human hepatoma cell line transfected with the hepatitis B virus genome. Proc Natl Acad Sci 1989;86:8422–8426.
[66] Terré S, Petit MA, Bréchot C. Defective hepatitis B virus particles are generated by packaging and reverse transcription of spliced viral RNAs in vivo. J Virol 1991;65:5539–5543.
[67] Kock J, Nassal M, Deres K, Blum HE, von Weizsacker F. Hepatitis B virus nucleocapsids formed by carboxy-terminally mutated core proteins contain spliced viral genomes but lack full-size DNA. J Virol 2004;78:13812–13818.
[68] Lam AM, Ren S, Espiritu C, Kelly M, Lau V, Zheng L, et al. Hepatitis B virus capsid assembly modulators, but not nucleoside analogs, inhibit the production of extracellular pregenomic RNA and spliced RNA variants. Antimicrob Agents Chemother 2017;61. https://doi.org/10.1128/AAC.00680-17.
[69] Rosmorduc O, Petit M-A, Pol S, Capel F, Bortolotti F, Berthelot P, et al. In vivo and in vitro expression of defective hepatitis B virus particles generated by spliced hepatitis B virus RNA. Hepatology 1995;22:10–19.
[70] Huang H-L, Jeng K-S, Hu C-P, Tsai C-H, Lo SJ, Chang C. Identification and characterization of a structural protein of hepatitis B virus: a polymerase and surface fusion protein encoded by a spliced RNA. Virology 2000;275:398–410.
[71] Soussan P, Garreau F, Zylberberg H, Ferray C, Brechot C, Kremsdorf D. In vivo expression of a new hepatitis B virus protein encoded by a spliced RNA 2000. J Clin Invest 2000;105(1):55–60.
[72] Bartenschlager R, Schaller H. Hepadnaviral assembly is initiated by polymerase binding to the encapsidation signal in the viral RNA genome. EMBO J 1992;11:3413–3420.
[73] Hirsch RC, Loeb DD, Pollack JR, Ganem D. cis-acting sequences required for encapsidation of duck hepatitis B virus pregenomic RNA. J Virol 1991;65:8.
[74] Junker-Niepmann M, Bartenschlager R, Schaller H. A short cis-acting sequence is required for hepatitis B virus pregenome encapsidation and sufficient for packaging of foreign RNA. EMBO J 1990;9:3389–3396.
[75] Knaus T, Nassal M. The encapsidation signal on the hepatitis B virus RNA pregenome forms a stem-loop structure that is critical for its function. Nucleic Acids Res 1993;21:3967–3975.
[76] DeStefano JJ, Buiser RG, Mallaber LM, Myers TW, Bambara RA, Fay PJ. Polymerization and RNase H activities of the reverse transcriptases from avian myeloblastosis, human immunodeficiency, and Moloney murine leukemia viruses are functionally uncoupled. J Biol Chem 1991;266:7423–7431.
[77] Sokoloski KJ, Haist KC, Morrison TE, Mukhopadhyay S, Hardy RW. Noncapped alphavirus genomic RNAs and their role during infection. J Virol 2015;89:6080–6092.
[78] Wang J, Yu Y, Li G, Shen C, Meng Z, Zheng J, et al. Relationship between serum HBV-RNA levels and intrahepatic viral as well as histologic activity markers in entecavir-treated patients. J Hepatol 2018;68:16–24.

III. "Science en bulles"

The "Fête de la Science » is a French national event that promotes science dissemination to a large public with conferences, exhibitions, educational workshops, etc.

Since 2016, "Fête de la Science" has started a project joining PhD students and cartoonists to create a book reporting different thesis subjects about diverse scientific fields, from mathematics to social sciences. This project is co-founded by the Ministry of Culture and the Ministry of Higher Education, Research and Innovation. I participated to this public outreach project in 2019 for the publishing of the book entitled "Science en bulles" at 80 000 copies that were distributed to French schools, high-schools, and Universities, and also cultural partner shops.

Cartoons were realized by Peb&Fox, a duo of cartoonists involved in this project since its creation. We worked on representing HBV as a well-hidden virus that hijacks cellular functions for its replication. We conveyed the most important information about HBV infection and dangerousness of the related diseases. We also illustrated the working environment in biology laboratories, notably security rules.

HBV, UN VIRUS BIEN CACHÉ

Plus de 257 millions de personnes souffrent de l'hépatite B dans le monde. Cette infection cause des maladies du foie sévères, pouvant aller jusqu'au développement d'un cancer, le carcinome hépatocellulaire, troisième cause de mortalité liée au cancer.

Le but de ma recherche est de mieux appréhender la manière dont fonctionne HBV, le virus à l'origine de cette infection, afin que de nouvelles thérapies puissent être développées. Car, actuellement, il n'existe pas de thérapeutique permettant d'éradiquer la maladie.

Lorsqu'il infecte une cellule, HBV forme, dans le noyau de celle-ci, un minichromosome qui ressemble au chromosome cellulaire. Le virus leurre ainsi la cellule, ce qui lui permet de se reproduire en masse, provoquant à long terme une maladie du foie.

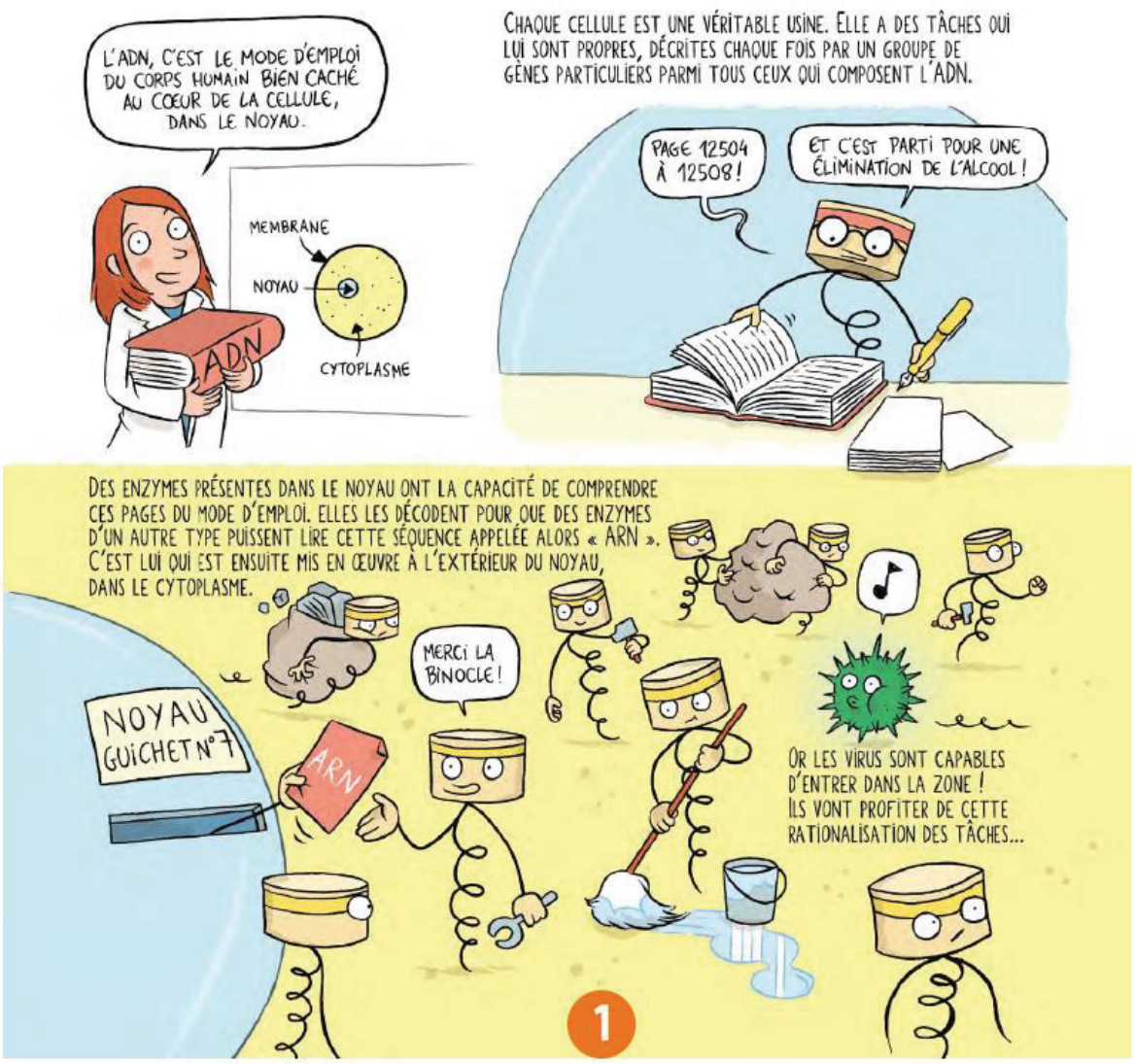
C'est donc ce minichromosome viral et son fonctionnement que j'étudie, pour comprendre les mécanismes de la vie du virus. En espérant trouver son maillon faible !

POUR EN SAVOIR PLUS

TITRE ORIGINAL DE LA THÈSE

Rôle des hélicases DDX5 et DDX17 et du complexe protéique associé dans la régulation transcriptionnelle du minichromosome du virus de l'hépatite B

Centre de recherche en cancérologie de Lyon (CRCL), université de Lyon


L'ADN, C'EST LE MODE D'EMPLOI DU CORPS HUMAIN BIEN CACHÉ AU CŒUR DE LA CELLULE, DANS LE NOYAU.
MEMBRANE
NOYAU
CYTOPLASME
ADN
CHAQUE CELLULE EST UNE VÉRITABLE USINE. ELLE A DES TÂCHES QUI LUI SONT PROPRES, DÉCRITES CHAQUE FOIS PAR UN GROUPE DE GÈNES PARTICULIERS PARMI TOUS CEUX QUI COMPOSENT L'ADN.
PAGE 12504 À 12508 !
ET C'EST PARTI POUR UNE ÉLIMINATION DE L'ALCOOL !
DES ENZYMES PRÉSENTES DANS LE NOYAU ONT LA CAPACITÉ DE COMPRENDRE CES PAGES DU MODE D'EMPLOI. ELLES LES DÉCODENT POUR QUE DES ENZYMES D'UN AUTRE TYPE PUISSENT LIRE CETTE SÉQUENCE APPELÉE ALORS « ARN ». C'EST LUI QUI EST ENSUITE MIS EN ŒUVRE À L'EXTÉRIEUR DU NOYAU, DANS LE CYTOPLASME.
NOYAU GUICHET N°7
ARN
MERCI LA BINOCLE !
OR LES VIRUS SONT CAPABLES D'ENTRER DANS LA ZONE ! ILS VONT PROFITER DE CETTE RATIONALISATION DES TÂCHES...
1


ILS N'ONT PAS LA CAPACITÉ DE SE REPRODUIRE TOUT SEULS. ALORS ILS ÉCRIVENT UN FAUX MODE D'EMPLOI POUR QUE LES ENZYMES LEUR DONNENT UN COUP DE MAIN...
TIENS, C'EST LE TYPE DU NOYAU QUI M'A DONNÉ ÇA POUR TOI.
AH OUI ?
J'ÉTAIS SUR AUTRE CHOSE, LÀ... MAIS OK, ÇA MARCHE !
ARN
CAR LE VIRUS SAIT ÉCRIRE DANS LE MÊME LANGAGE QU'ELLES...

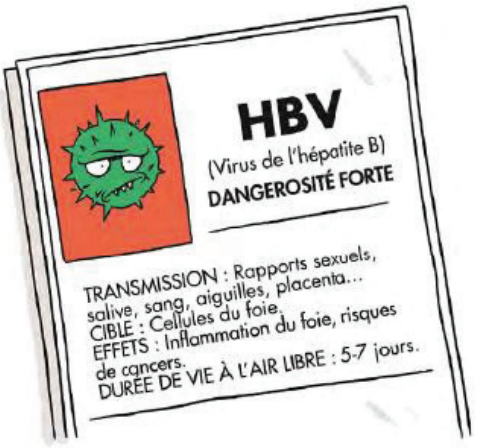

À CAUSE DE CERTAINES ENZYMES QUI SE SONT FAIT AVOIR ET SE SONT DÉTOURNÉES DE LEUR TÂCHE EN FABRIQUANT DES VIRUS À LA CHAÎNE, L'ORGANISME SE FATIGUE EN ESSAYANT DE LES COMBATTRE.


MOI, DANS MON LABO AVEC SAS D'ENTRÉE, JE ME SENS COMME L'ADN CONFINÉ DANS LE NOYAU.
LUNETTES DE SÉCURITÉ
MASQUE
DOUBLE PAIRE DE GANTS
SURCHAUSSURES
COMBINAISON INTÉGRALE

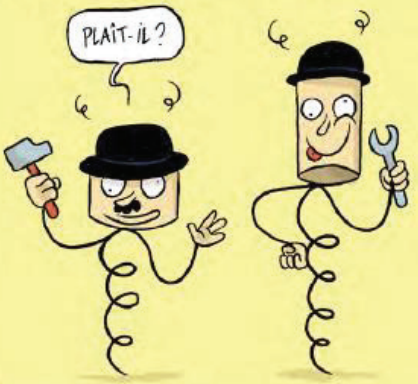

C'EST PARCE QUE MON ENVIRONNEMENT DE TRAVAIL EST ASSEZ DANGEREUX, DE PAR LES PRODUITS ET OUTILS QUE J'UTILISE...
EXPLOSIF
INFLAMMABLE
COMBURANT
GAZ SOUS PRESSION
TOXIQUE, IRRITANT
DANGER POUR LE MILIEU AQUATIQUE
CANCÉROGÈNE
CORROSIF
TOXIQUE, IRRITANT

2

... ET AUSSI PARCE QUE LE VIRUS QUE J'ÉTUDIE, C'EST PAS DU PETIT RHUME...



ON A PU IDENTIFIER PLUSIEURS ENZYMES, COMPLICES DU VIRUS ET RESPONSABLES DE SA FABRICATION DANS LA CELLULE. J'ÉTUDIE DEUX D'ENTRE ELLES, QUI TRAVAILLENT EN DUO.



JE LES PRIVE DE LEURS OUTILS, CE QUI LES EMPÊCHE DE METTRE EN ŒUVRE CORRECTEMENT L'ARN QUE LES VIRUS ONT FABRIQUÉ.

MAIS IL ME FAUT ENCORE M'ASSURER QUE SI JE NEUTRALISE CES ENZYMES LA CELLULE PEUT TOUJOURS FONCTIONNER NORMALEMENT...

3

IV. Communications

1. Oral Presentations

- Oral presentation **International HBV meeting.** Taormina, October 2018
  - *HBV meeting Travel Grant*
- Oral presentation **18th meeting of national hepatitis network**. Paris, April 2018

2. Poster Presentations

- Poster presentation **4th International Cancer Symposium** –CRCL. Lyon, October 2019
- Poster presentation **Chromatin & Epigenetics** –EMBL. Heidelberg, May 2019
- Poster presentation **4th international course of "Post-transcriptional gene regulation"** –Institut Curie. Paris, April 2019
  - *Doctoral school Travel Grant*
- Poster presentation **The Liver meeting –AASLD**. San Francisco, November 2018
- Poster presentation **EASL international liver congress**. Paris, May 2018

3. Publications

Full-length 5'RACE identifies all major Hepatitis B Virus transcripts during infection in cultured hepatocytes and in patients

Bernd Stadelmayer, Audrey Diederichs, Fleur Chapus, Michel Rivoire, Gregory Neveu, Antoine Alam, Laurent Fraisse, Kara Carter, Barbara Testoni, Fabien Zoulim
Journal of Hepatology, 2020

HiHiMap: single-cell quantitation of histones and histone posttranslational modifications across the cell cycle by high-throughput imaging.

Zane L, Chapus F, Pegoraro G, Misteli T
Mol Biol Cell, 2017