



Université de Sherbrooke

**Nouveau rôle oncogénique pour les virus de l'hépatite B et  
C: l'altération des événements d'épissage alternatif**

Par  
Marie-Pier Tremblay  
Programme de Biochimie

Mémoire présenté à la Faculté de médecine et des sciences de la santé  
en vue de l'obtention du grade de maître ès sciences (M. Sc.)  
en Biochimie

Sherbrooke, Québec, Canada  
Janvier 2016

Membres du jury d'évaluation  
Pr. Martin Bisaillon, Département de Biochimie  
Pr. Michelle Scott, Département de Biochimie  
Pr. Brendan Bell, Département de Microbiologie

© Marie-Pier Tremblay, 2016

## RÉSUMÉ

### **Nouveau rôle oncogénique pour les virus de l'hépatite B et C: l'altération des événements d'épissage alternatif**

Par  
Marie-Pier Tremblay  
Programme de Biochimie

Mémoire présenté à la Faculté de médecine et des sciences de la santé en vue de l'obtention du diplôme de maître ès sciences (M.Sc.) en biochimie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec, Canada, J1H 5N4

Le carcinome hépatocellulaire figure parmi les cancers les plus meurtriers et, encore en 2015, les infections aux virus de l'hépatite B et C sont en tête de liste des principales causes d'apparition de cancer du foie. En effet, 80% des cas de carcinomes hépatocellulaires sont attribuables à une infection virale par un de ces deux agents pathogènes. D'un point de vue épidémiologique, le carcinome hépatocellulaire n'est pas celui qui frappe le plus d'individus en Amérique du Nord. En effet, seulement une personne sur 264 sera affectée par ce cancer. Toutefois, il est important de l'étudier en raison de son faible taux de survie, ce qui le classe en cinquième place des cancers les plus meurtriers.

Dernièrement, les recherches font mention des modifications dans l'épissage alternatif comme mécanisme d'apparition de plusieurs types de cancer, notamment les cancers du sein, de l'ovaire et de la prostate. C'est pour cette raison que nous avons investigué l'implication des virus de l'hépatite B et C dans l'épissage alternatif de gènes cellulaires, afin de mettre à jour un nouveau mécanisme de carcinogénèse utilisé par les oncovirus pour transformer leurs cellules hôtes.

L'étude présentée dans ce mémoire rapporte des changements dans l'épissage alternatif de plusieurs gènes cellulaires, lorsque l'on évalue les patrons d'épissage alternatif des transcrits provenant d'analyse de séquençage d'ARN à haut débit de tumeurs du foie comparativement à des tissus hépatiques sains. Ces analyses ont permis de déterminer certains changements attribuables à la présence de chacun des deux virus hépatotropiques au niveau de l'épissage alternatif, et d'identifier certains changements qui pourraient impliquer la transformation du tissu sain en tissu cancéreux. Il serait intéressant de faire une validation des effets directs d'une protéine virale du virus de l'hépatite B, la protéine trans-activatrice HBx, et de vérifier si la présence de HBx en cellule serait responsable de changements au niveau de l'épissage alternatif de gènes impliqués dans le processus carcinogénique.

Mots clés : virus, hépatite B, hépatite C, épissage alternatif, cancer, carcinome hépatocellulaire, foie, protéine HBx

## TABLE DES MATIERES

<b>Résumé.....</b>	<b>ii</b>
<b>Table des matières.....</b>	<b>iii</b>
<b>Liste des figures .....</b>	<b>v</b>
<b>Liste des tableaux.....</b>	<b>vi</b>
<b>Introduction .....</b>	<b>1</b>
<b>1.1) VIRUS DE L'HÉPATITE B .....</b>	<b>1</b>
1.1.1) Caractéristiques générales .....	1
1.1.2) Transmission.....	1
1.1.3) Variabilité génotypique .....	2
1.1.4) Cycle de réplication virale .....	2
1.1.5) Structure et organisation génomique .....	4
1.1.5.1) Protéines de surface L, M et S .....	6
1.1.5.2) Core et pré-core .....	7
1.1.5.3) Polymérase virale.....	7
1.1.5.4) HBx.....	7
1.1.6) Maladies associées au VHB .....	11
1.1.6.1) Hépatite chronique.....	12
1.1.6.2) Cirrhose du foie.....	12
1.1.6.3) Carcinome hépatocellulaire.....	13
1.1.7) Modèles d'étude.....	13
1.1.8) Traitements et prévention .....	14
1.1.9) Virus de l'hépatite C .....	15
<b>1.2) LE CANCER .....</b>	<b>17</b>
1.2.1) Généralités et caractéristiques .....	17
1.2.1.1) Cellule cancéreuse.....	17
1.2.1.2) Tumeur.....	19
1.2.2) Épidémiologie du cancer.....	19
1.2.2.1) Les virus et le cancer.....	22
1.2.3) Dépistage et traitements des cancers .....	23
1.2.3.1) Dépistage.....	23
1.2.3.2) Traitements.....	24
<b>1.3) ÉPISSAGE ALTERNATIF .....</b>	<b>27</b>
1.3.1) Définition et mécanisme général.....	27

1.3.1.1) Mécanisme d'action .....	28
1.3.1.2) Mécanismes de régulation.....	30
1.3.2) Les types d'épissage alternatif.....	31
1.3.3) Implication dans différentes maladies .....	33
1.3.3.1) L'importance de l'épissage alternatif dans le cancer.....	35
<b>1.4) HYPOTHÈSE ET OBJECTIF DE RECHERCHE .....</b>	<b>38</b>
Objectif .....	38
<b>Article .....</b>	<b>39</b>
<b>2.1) AVANT-PROPOS DE L'ARTICLE.....</b>	<b>39</b>
<b>2.2) RÉSUMÉ DE L'ARTICLE ET IMPLICATION .....</b>	<b>40</b>
<b>2.3) ARTICLE.....</b>	<b>42</b>
<b>Discussion .....</b>	<b>90</b>
<b>3.1 MODIFICATION DES ÉVÉNEMENTS D'ÉPISSAGE ALTERNATIF PAR UN ONCOGÈNE</b>	
<b>VIRAL.....</b>	<b>91</b>
3.1.1 Identification de modifications d'épissage alternatif par la protéine HBx.....	92
3.1.2 Comparaison avec les résultats obtenus pour les tissus positifs pour le VHB.....	95
3.1.3 Comparaison avec une autre protéine virale: EBNA1.....	96
3.1.5 Recherche de séquence consensus .....	99
<b>3.2 PERSPECTIVES .....</b>	<b>104</b>
<b>Remerciements .....</b>	<b>106</b>
<b>Liste des références.....</b>	<b>107</b>

## LISTE DES FIGURES

Figure 1 - Cycle de réplication virale du virus de l'hépatite B .....	3
Figure 2 - Structure génomique du virus de l'hépatite B .....	5
Figure 3 - Transcrits produits par le génome du VHB .....	6
Figure 4 - Les trois antigènes de surface du VHB.....	6
Figure 5 - Quelques effets de HBx sur le développement d'une cellule cancéreuse .....	9
Figure 6 - Schématisation de la progression de l'infection par le VHB.....	12
Figure 7 – Organisation génomique du virus de l'hépatite C .....	16
Figure 8 - Les caractéristiques d'une cellule cancéreuse .....	19
Figure 9 - Compilation des causes de décès au Canada .....	20
Figure 10 - Statistiques sur le cancer .....	21
Figure 11 - Les virus causant des cancers chez l'humain .....	22
Figure 12 - Réaction d'épissage .....	29
Figure 13 - Les différents types d'épissage alternatif d'un pré-ARNm.....	32
Figure 14 – Méthodologie pour l'analyse des modifications d'épissage alternatif.....	93
Figure 15 – Candidats obtenus pour l'épissage alternatif modifié lorsque la protéine HBx est exprimée en cellules eucaryotes .....	94
Figure 16 – Distribution des candidats au sein de différents processus biologiques impliqués dans l'apparition de cancers.....	95
Figure 17 – Comparaison avec les résultats obtenus dans les tissus cancéreux positifs pour le virus de l'hépatite B.....	96
Figure 18 – Comparaison avec la protéine EBNA1 du virus epstein-barr .....	97
Figure 19 – Présence d'une séquence consensus de 57 nucléotides dans 129 événements d'épissage alternatif modifiés en présence de la protéine HBx .....	99
Figure 20 – Possibilité de liaison de HBx à la séquence consensus .....	101
Figure 21 – Possibilité de liaison de la protéine HBx à des facteurs d'épissage qui lieraient la séquence consensus.....	103

## LISTE DES TABLEAUX

Tableau 1 – Modifications communes entre les cellules exprimant EBNA1 et celles exprimant HBx et l'isoforme favorisée .....	98
Tableau 2 – Présence de la séquence consensus de 57 nucléotides dans 121 gènes dont l'épissage alternatif est modifié en présence de la protéine HBx .....	100

## INTRODUCTION

### 1.1) VIRUS DE L'HÉPATITE B

#### *1.1.1) Caractéristiques générales*

Le virus de l'hépatite B (VHB) est un virus de la famille des *Hepadnaviridae* appartenant au genre des *Orthohepadnavirus* (Francki et al, 1991). Il est un des pathogènes les plus importants infectant l'Homme. La découverte de ce virus remonte aux années 1940 (MacCallum, 1946). Déjà à cette époque, il était connu qu'un virus était présent dans le sang et causait des hépatites et des cancers du foie chez les individus infectés. Par contre, ce n'est que dans les années 1950 que le VHB a été nommé ainsi, de par sa capacité à infecter les cellules du foie, plus particulièrement les hépatocytes, et causer une inflammation du foie que l'on appelle hépatite (Edmondson et Steiner, 1954; Ganem et Prince, 2004). Bien qu'un vaccin existe depuis les années 1980 afin de prévenir les infections par le virus de l'hépatite B, plusieurs décès y sont encore associés, de manière plus prononcée dans les pays en développement (Szmunn et al, 1980). En effet, sur l'ensemble du globe, plus de 2 milliards de personnes ont contracté le VHB au cours de leur vie. De ces 2 milliards, on compte environ 350 millions d'individus chroniquement infectés. Environ 1 million d'individus infectés meurent chaque année des suites d'une infection au VHB, que ce soit des décès dus à la persistance d'une infection, par cirrhose du foie ou par cancer. Enfin, chaque année, on compte environ 4 millions de nouveaux cas d'hépatite chronique (WHO, 2015).

#### *1.1.2) Transmission*

Ce virus se transmet principalement par contact entre fluides corporels, plus précisément par contact périnatal de la mère à l'enfant ou lors de relations sexuelles non protégées, ou encore par contact direct avec du sang contaminé, par exemple lors de l'utilisation de drogues injectables. De plus, il est intéressant de mentionner que le virus peut rester actif hors de son hôte pendant au moins sept jours, et donc infecter une personne qui ne serait pas protégée contre le VHB. Bien qu'une période d'incubation variant de 30 à 80 jours est nécessaire au



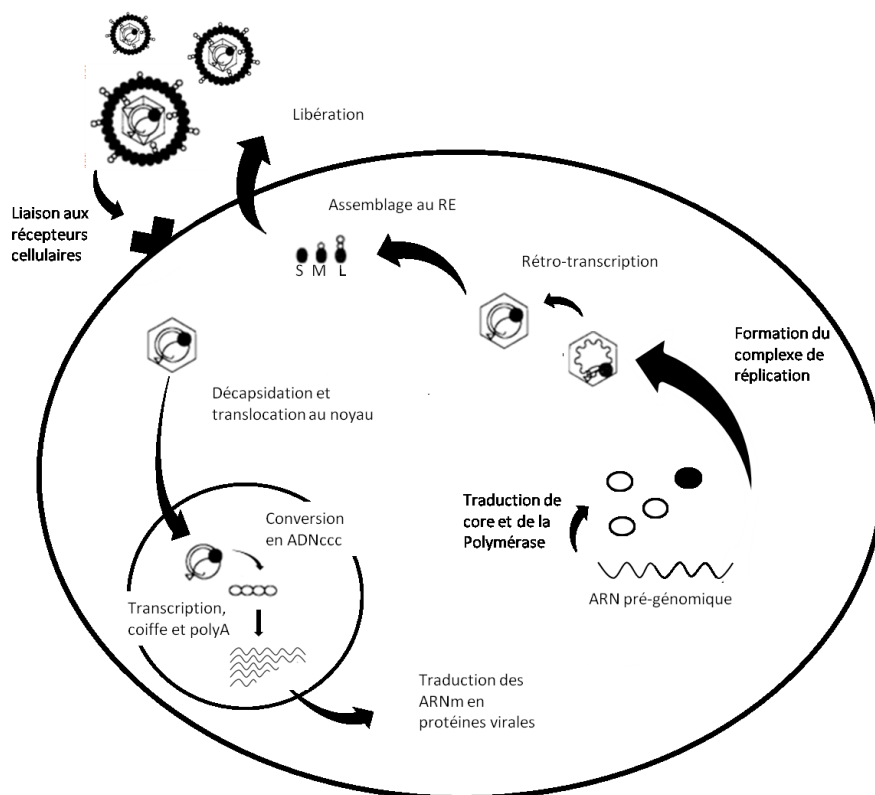
virus, il peut être détectable dans le sang à partir du trentième jour suivant l'infection (Villarejos et al, 1974; Karayiannis et al, 1985; Chlumsky et Krtek, 1982).

### ***1.1.3) Variabilité géotypique***

Huit différents géotypes du virus de l'hépatite B ont été identifiés, nommés de A à H (Sun et Chu, 1984). Ces géotypes se distinguent les uns des autres de par la variation dans les épitopes de l'antigène de surface (HBsAg). Les différents géotypes sont pour la plupart associés à une région géographique où ils se retrouvent majoritairement. À titre d'exemple, sur le continent nord-américain, ceux qui prédominent sont les géotypes A et C, alors qu'en Europe et au Moyen-Orient, ce sont les géotypes A et D qui sont en plus grande importance (Kumar et al, 2011). Le géotype C est celui associé à un risque accru de carcinome hépatocellulaire (Wang et al, 2007).

### ***1.1.4) Cycle de réplication virale***

Comme tout virus, le VHB va infecter une cellule afin de se répliquer. En effet, de par sa nature, un virus est un organisme biologique nécessitant la machinerie cellulaire d'un autre organisme pour se multiplier. Il est donc parasite d'une cellule hôte qui lui fournira le métabolisme et les constituants nécessaires à cette réplication. Celle-ci suivra un dessein bien défini: (Tortora et al, 2003)



**Figure 1 - Cycle de réplication virale du virus de l'hépatite B**

Le cycle de réplication virale du VHB passe par les étapes classiques de réplication de tous les virus: attachement, pénétration, décapsidation, réplication virale, assemblage et libération. Adapté de (Nassal et Schaller, 1996).

La première étape de l'infection virale est caractérisée par l'attachement du virus aux cellules cibles, dans ce cas-ci les hépatocytes. Lorsque le VHB rencontrera les cellules hépatiques, il devra pénétrer celles-ci pour s'y répliquer. Pour ce faire, il a été proposé que l'antigène pré-S1 de la protéine de surface L soit responsable de l'attachement de la particule virale aux récepteurs membranaires de l'hépatocyte (Locarnini et Zoulim, 2010; Urban et Gripon, 2002). Les récepteurs cellulaires responsables de l'attachement n'ont toutefois pas encore été définis. Subséquemment, le virion pourra pénétrer les cellules par endocytose et se rendre au cytoplasme (Glebe et Urban, 2007). Il y aura alors décapsidation du virus et l'ADN viral sera transporté jusqu'au noyau par les pores nucléaires, lieu où se produit la réplication virale (Rabe et al, 2003; Kann et al, 1997). À son arrivée au noyau, l'ADN partiellement double brin du VHB sera converti en ADN circulaire double brin épisomal (ADNccc) qui pourra se maintenir au noyau (Bock et al, 2001; Bock et al, 1994; Newbold et

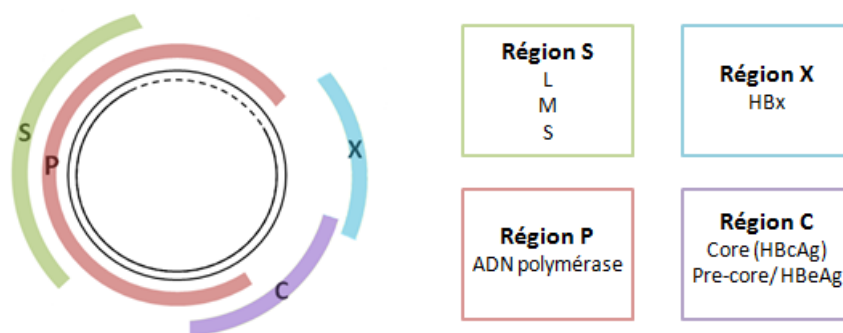
al, 1995). C'est à partir de cet ADNccc que la transcription des différents ARN sera effectuée par l'ARN polymérase II humaine. Suite à la transcription, les ARNm ainsi que l'ARN pré-génomique servant de matrice pour la réplication du génome seront coiffés et polyadénylés au noyau, pour ensuite être transportés au cytoplasme où la traduction aura lieu (Nassal et Schaller, 1996).

Bien que le virus de l'hépatite B soit un virus à ADN, la réplication du génome se fait par transcription inverse, de manière semblable aux rétrovirus (l'ADN étant rapidement transcrit en ARN pré-génomique qui servira à la réplication) (Newbold et al, 1995; Bock et al, 2001). Pour y arriver, la polymérase et la protéine core seront rapidement traduites à partir de l'ARN pré-génomique dû au fait qu'elles sont essentielles à la rétro-transcription (Mason et al, 1982; Summers et Mason, 1982). La polymérase ainsi synthétisée va s'attacher à l'ARN pré-génomique à partir duquel elle a été traduite pour engendrer la transcription inverse qui générera l'ADN partiellement double brin du VHB (Bartenschlager et Schaller, 1992; Hirsch et al, 1991). Cette liaison va engendrer la formation d'une capsid virale immature dans laquelle la transcription inverse de l'ARN pré-génomique aura lieu (Coppolino et al, 2015; Nassal et Schaller, 1996). Une fois l'ADN partiellement double brin du virus synthétisé, deux voies sont observées. D'une part, une partie du complexe contenant le nouvel ADN partiellement double brin se rendra au noyau, pour reformer une population d'ADNccc, et la réplication pourra recommencer (Summers et al, 1990). D'autre part, la nucléocapsid sera maturée, pour être ensuite enveloppée par les trois antigènes de surface (HBsAg) L, M et S, en plus de matériel provenant de la membrane du réticulum endoplasmique où ce processus a lieu. Les virions formés pourront finalement être sécrétés hors de la cellule (Bruss, 2007; Lambert et al, 2007).

### ***1.1.5) Structure et organisation génomique***

Le virus de l'hépatite B est un petit virus de 42 nm à ADN circulaire partiellement double brin de 3.2 kb. Son organisation génétique est très compacte (Seeger et Mason, 2000). En effet, le génome du VHB comporte quatre régions codantes différentes qui se chevauchent, soit les régions S, C, P et X (Glebe et Bremer, 2013). La région S sert à l'expression de toutes les protéines de surface (S, M et L) (Bayer et al, 1968; Dane et al, 1970; Glebe et Urban,

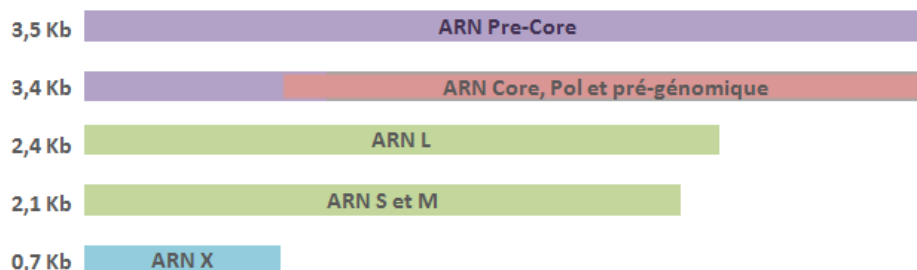
2007). La région C, quant à elle, code pour les protéines pré-core et core de la capsid virale. Enfin, la région codante P exprime la polymérase virale alors que la région X encode pour la protéine transactivatrice HBx (Nassal et Schaller, 1993; Tiollais et al, 1985; Seeger et Mason, 2015).



**Figure 2 - Structure génomique du virus de l'hépatite B**

Le génome du virus de l'hépatite B est un génome d'ADN partiellement double brin contenant quatre régions codantes (S, X, P et C) donnant lieu à sept protéines virales.

Ces quatre régions codantes sont transcrites en cinq ARN responsables de l'expression de sept protéines. La région S est transcrite en deux ARNm, soit un de 2.4 kb exprimant la protéine L, et un de 2.1 kb traduit en protéines M et S. Un petit ARNm de 0.7 kb est transcrit à partir de la région X. Finalement, deux grands ARN de 3.4 et 3.5 kb servent respectivement d'ARN pré-génomique (ARNpg) et d'ARNm pour la traduction de la protéine pré-core (Mast et al, 1999; Chu et al, 2004; Quasdorff et Protzer, 2010; Seeger et Mason, 2000; Pollicino et al, 2014; Bruss, 2007).

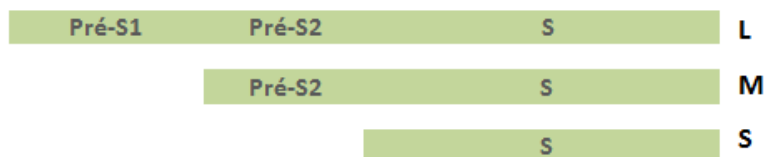


**Figure 3 - Transcrits produits par le génome du VHB**

Suite à la conversion de l'ADN partiellement double brin en ADNccc, l'ARN polymérase II cellulaire transcrita cinq ARN qui seront coiffés et polyadénylés avant d'être exportés au cytoplasme.

#### *1.1.5.1) Protéines de surface L, M et S*

La région S code pour les trois protéines formant l'enveloppe virale, les protéines de surface L, M et S. Les trois protéines se distinguent de par leur taille, S étant la plus courte et L la plus longue. Elles sont toutes traduites à partir du même cadre de lecture ouvert (ORF), à partir de codons d'initiation différents. Ainsi, les trois protéines contiennent le domaine S, les protéines L et M contiennent le domaine pré-S2 et L est la seule qui contient le domaine pré-S1 (Glebe et Urban, 2007; Churin et al, 2015).



**Figure 4 - Les trois antigènes de surface du VHB**

Les trois protéines de surface possèdent le domaine S en commun, les protéines L et M partagent le domaine pré-S2, et la protéine L se voit rajouter le domaine pré-S1 par rapport à ses deux partenaires.

Par contre, elles sont traduites à partir de deux ARNm différents, un pour la protéine L et un deuxième pour M et S. Suite à la traduction, les trois protéines vont se loger au niveau

de la membrane du réticulum endoplasmique. De cette façon, lorsque la nucléocapside mature arrive au réticulum endoplasmique, elle peut être enveloppée par les protéines de surface et le virion mature peut être sécrété (Glebe et Urban, 2007; Churin et al, 2015).

#### *1.1.5.2) Core et pré-core*

La région codante C encode deux protéines: core et pré-core. En ce qui a trait à pré-core, sa fonction est nébuleuse. Suite à la traduction de la protéine, pré-core est processée à l'intérieur de la cellule pour former l'antigène HBe, lequel sera par la suite sécrété dans le milieu extracellulaire. L'antigène HBe ne fait pas partie de la particule virale et n'est pas essentiel à la réplication virale (Chang et al, 1987; Chen et al, 1992). HBeAg pourrait avoir un effet sur la persistance de l'infection virale et sur l'évasion au système immunitaire (Milich et al, 1998; Milich et al, 1990). Par ailleurs, il sert de marqueur pour la détection d'une infection virale par le virus de l'hépatite B. Pour sa part, core a un rôle plutôt simple mais somme toute important: il forme la nucléocapside virale. Cette nucléocapside sert de capsule pour la synthèse de l'ADN viral à partir de l'ARNpg. La capsid sert également de structure protectrice pour le génome alors synthétisé (Chang et al, 1994).

#### *1.1.5.3) Polymérase virale*


La région codante P du génome encode pour une polymérase virale de 94 kDa ayant plusieurs fonctions au niveau de la réplication du virus. La polymérase est impliquée dans les activités de synthèse de l'ADN double brin par transcriptase inverse en plus d'être une nucléase. La polymérase virale commence d'abord par rétro-transcrire l'ADN de polarité négative à partir de la matrice d'ARNpg. Par la suite, l'ADN positif, lequel est incomplet, est synthétisé. Aussi, lors de la synthèse de cet ADN, le gabarit d'ARN doit être détruit puisqu'il n'est pas encapsidé avec le génome du virus. Ceci est fait par l'activité nucléase de cette même polymérase (Bartenschlager et Schaller, 1992; Bruss, 2007; Locarnini et Zoulim, 2010; C Seeger et Mason, 2000).

#### *1.1.5.4) HBx*

La protéine de 17 kDa encodée par la région X du génome du VHB est nommée HBx, car sa fonction précise reste inconnue. HBx serait essentielle à l'établissement de l'infection virale

mais dispensable à la réplication du virus et semble être très conservée parmi les *Hepadnaviridae* parasitant les cellules de mammifères (Chen et al, 1993; Zoulim et al, 1994). Les fonctions qu'on lui attribue consistent principalement à être un régulateur à multiples effets sur la cellule hôte. Par exemple, HBx est un régulateur transcriptionnel important. Bien qu'il ne soit pas capable de lier directement l'ADN, il peut moduler la transcription cellulaire par interaction avec d'autres protéines cellulaires. Ceci lui permet de contrôler l'expression de gènes importants pour la réplication virale, entre autres les gènes contrôlant la survie de l'hépatocyte infecté (Zahm et al, 1988; Twu et Schloemer, 1987). De la même façon, HBx peut aussi moduler les voies de transduction de signal (Vivanco et Sawyers, 2002; Khattar et al, 2012; Cha et al, 2004), le cycle cellulaire (Benn et Schneider, 1995), la réparation de l'ADN (Qadri et al, 2011) et l'apoptose des cellules (Hu et al, 2011; Chen et Siddiqui, 2007). HBx est également responsable de modifications épigénétiques sur le génome de la cellule hôte (Tong et al, 2009). Enfin, la protéine HBx est capable de moduler les effets de certains micro-ARN cellulaires, ceci ayant comme conséquence de multiples changements importants dans le fonctionnement de la cellule (Xie et al, 2014).

Tous ces changements au niveau de la cellule favorisent la transformation vers un phénotype cancéreux. En effet, le VHB induit le cancer du foie dans les hépatocytes qu'il infecte (cet aspect du virus sera traité plus loin dans ce mémoire) (Zheng et al, 2007). Étant donné les rôles variés que HBx semble avoir au niveau de la régulation des mécanismes d'apoptose et de survie cellulaire, cette protéine serait l'acteur principal dans le développement de tels cancers. De plus, plusieurs évidences directes démontrent que HBx induirait la carcinogénèse lorsqu'injectée dans des souris transgéniques (Kim et al, 1991; Koike et al, 1994). Il est donc stipulé que la fonction principale de HBx serait de modifier la cellule hôte afin d'optimiser la réplication virale, et ce, en contrôlant les processus de survie et d'apoptose des cellules infectées par le VHB. HBx serait donc un oncogène viral.



Régulation transcriptionnelle	Modifications épigénétiques	Effets anti-apoptotiques	Cycle cellulaire	Réparation de l'ADN	Voie de transduction de signal	miARN
AP-1 AP-2 NF-κB TBP IL-8 TGF-1β/Smad3/ c-myc TNF hTERT HNF4α	DNMTs /méthylation îlots CpG (hyperméthylation des promoteurs: diminution de l'expression de gènes suppresseurs de tumeurs)  HDAC1 (acétylation de gènes promouvant la survie cellulaire)	p53 Survivin CASP3 CASP9 PDCD4	Entrée en phase S (CDK2-cyclin-E, CDK2-cyclin-A)  Déstabilisation des fuseaux mitotiques (instabilité chromosomique)  Inhibition de p18 par dérégulation de HULC (lncRNA)  Points de contrôle (p21, p15, p16, p27, CDK4)	DDB1 UGT1A9 UNG XRCC1 XRCC3 XRCC4 RAD17	Wnt/β-caténine EGRF Notch1 PI3K/ AKT/HNF4α Ras/Raf /MAPK JAK /STAT VEGF/HIF-1 ANG2 MMP/COX-2 (migration et invasion cellulaire)	IL-6/STAT3/miR-21 miR-34a miR-29a (PTEN) miR-155 miR-548 miR-15a miR-17-92 miR-122 (cyclin G1) miR-199a-3p (mTOR) miR-205 miR-210 miR-1 (HDAC4) miR-22 Let-7 (STAT3, Bcl-xL, Ras)

**Figure 5 - Quelques effets de HBx sur le développement d'une cellule cancéreuse**

HBx a de multiples effets et ce, sur différents métabolismes cellulaires. Elle peut réguler la transcription par interaction avec des facteurs de transcription et par modifications épigénétiques. Elle peut modifier le taux de survie cellulaire par inhibition de l'apoptose, par dérégulation du cycle cellulaire, par activation de voie de survie et par modulation de micro-ARN cellulaires. Elle peut aussi inhiber les processus de réparation de l'ADN, favorisant l'apparition de mutations dans le génome de l'hôte.

La figure 5 ne fait que dresser un portrait de quelques effets de la protéine virale HBx. Parmi ceux-ci, on note que HBx peut moduler de nombreux acteurs au niveau de plusieurs mécanismes cellulaires. Toutefois, la manière précise d'apparition du cancer par la protéine HBx reste inconnue. Cette figure classe les effets de HBx en différentes catégories de manière plutôt grossière. En effet, il est difficile de classer un effet dans une catégorie spécifique car tous ces mécanismes cellulaires se chevauchent et s'entremêlent. Par exemple, les effets au niveau de la transcription vont moduler les acteurs dans les autres catégories, et les modifications épigénétiques vont également moduler l'expression des gènes affectés, ce qui va causer un effet au niveau de la survie ou de l'apoptose.



Globalement, l'effet le plus important se situe au niveau de la régulation transcriptionnelle. Il est bien connu que HBx peut réguler l'expression de plus de 1200 gènes, que ce soit de manière positive ou négative (Qiu et al, 2013). La régulation de l'expression des gènes au niveau transcriptionnel passe par l'interaction de HBx avec certains facteurs de transcription de l'ARN polymérase II cellulaire, comme les facteurs AP-1, AP-2 et NF- $\kappa$ B (Zhang et al, 2009; Seto et al, 1990). HBx modifiera aussi la transcription de l'oncogène c-myc (Terradillos et al, 1997; Li et al, 2010) et de hTERT, (Zou et al, 2004; Liu et al, 2010) ce qui aura de grandes conséquences au sein de la cellule.

HBx est également capable de réguler la transcription au niveau épigénétique (Tong et al, 2009). En effet, l'oncoprotéine virale recrute des méthyltransférases pour rendre certaines régions du génome moins accessibles, ou à l'inverse des acétyltransférases pour ouvrir certaines régions chromosomiques. Les gènes principalement ciblés lors de telles modifications sont les gènes suppresseurs de tumeurs (inhibés par la méthylation des îlots CpG) et les gènes de survie cellulaire (activés grâce à l'acétylation) (Zhu et al, 2010; Lee et al, 2005; Park et al, 2011; Rawat et Bouchard, 2015).

Au niveau de l'apoptose, la protéine HBx réduit la concentration nucléaire du suppresseur de tumeurs p53 par sa séquestration au cytoplasme des cellules infectées par le VHB (Wang et al, 1995). Elle peut également inactiver la caspase 3, ce qui va avoir un effet protecteur sur la cellule contre l'apoptose (Gottlob et al, 1998). Par contre, il faut savoir que la protéine HBx a des effets contradictoires au niveau de la régulation de l'apoptose. Effectivement, dans certains cas, HBx favoriserait l'apoptose des cellules. Les hypothèses qui ont été émises pour expliquer ces effets opposés font mention de différentes mutations dans le gène de la protéine HBx (Lee et al, 2012), ou alors de différence dans la quantité de protéines exprimées (Ye et al, 2008).

Une autre façon d'avantager la survie des cellules infectées par le virus passe par le contrôle du cycle cellulaire, et donc par l'augmentation du taux de division cellulaire. La protéine HBx peut accélérer le cycle cellulaire en raccourcissant la phase G0 et donc en accélérant l'entrée en phase de synthèse de l'ADN (S). Ceci est réalisé par l'action de HBx

sur les complexes formés entre CDK2 et les cyclines E et A. De plus, HBx est capable d'inhiber les régulateurs de points de contrôle de cycle cellulaire, p16, p21 et p27, lesquels sont responsables de détecter les erreurs qui seraient survenus lors que la réplication de l'ADN (Benn et Schneider, 1995; Kim et al, 2008). Enfin, HBx va causer une instabilité chromosomique en déstabilisant les fuseaux mitotiques lors de la division cellulaire. Les mécanismes de réparation de l'ADN sont aussi grandement affaiblis par la protéine HBx (Forgues et al, 2003; Wen et al, 2008; Martin-Lluesma et al, 2008).

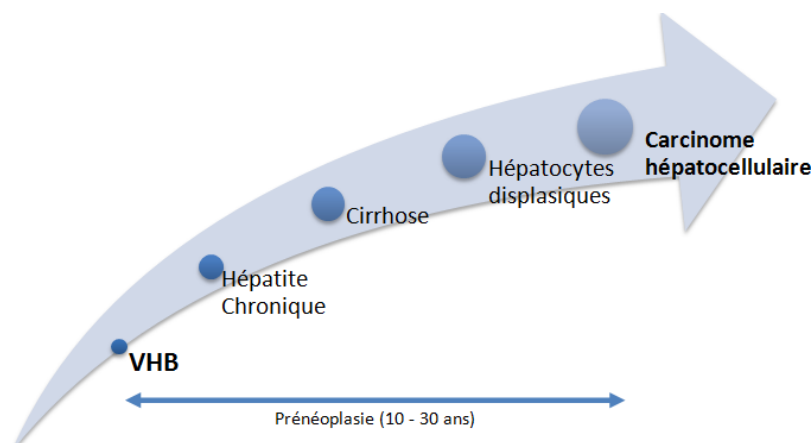
Ensuite, la protéine HBx peut avoir des effets positifs au niveau de la survie cellulaire par l'activation de certaines voies de survie, comme les voies Pi3k/Akt, JAK/STAT (Lee et Yun, 1998) et Ras/Raf/MAPK (Benn et Schneider, 1994). Par exemple, il est connu que dans les cancers du foie, l'activation de la voie Pi3k/Akt inhibe l'expression d'un facteur de transcription important pour la différenciation des cellules hépatiques, HNF4 $\alpha$ . La perte de ce facteur entraîne la dédifférenciation des cellules et ainsi une prolifération accrue des cellules (Rawat et Bouchard, 2015). L'angiogénèse des cellules est aussi augmentée par l'action de la protéine HBx. En effet, HBx active les voies VEGF et ANG2, favorisant ainsi la formation de nouveaux vaisseaux sanguins au niveau de la tumeur (Yoo et al, 2008; Han et al, 2007; Sanz-Cameno et al, 2006). Enfin, HBx favorise la migration et l'invasion des cellules en activant les métalloprotéases via l'activation préalable des cyclooxygénases 2 (COX-2) (Lara-Pezzi et al, 2002).

Finalement, la protéine HBx peut réguler de multiples micro-ARN. La figure 5 dresse un aperçu de ceux modulés par la protéine. En variant les miARN, HBx peut aller influencer différentes fonctions cellulaires énoncées plus haut. Somme toute, il est facile de constater les nombreux effets de HBx au niveau de l'apparition du cancer (Xie et al, 2014). Tous ces rôles laissent effectivement croire que la protéine HBx est capable de transformer une cellule saine en cellule cancéreuse.

#### ***1.1.6) Maladies associées au VHB***

Comme il a été mentionné, le virus de l'hépatite B tient son nom du fait qu'il infecte les hépatocytes, lesquels constituent 80% de la population cellulaire du foie. Ce faisant, une

inflammation au niveau de l'organe est engendrée, nommée hépatite, causant des symptômes tels que la jaunisse, une fatigue extrême, des nausées et des vomissements. Si l'organisme ne parvient pas à combattre le virus, l'atteinte inflammatoire persiste et elle devient chronique; la pathologie peut alors se transformer en cirrhose du foie et parfois même en cancer du foie (Motavaf et al, 2013; WHO, 2015).



**Figure 6 - Schématisation de la progression de l'infection par le VHB**

L'infection d'une personne par le virus de l'hépatite B mène dans certains cas à une hépatite chronique. Celle-ci pourra évoluer vers la cirrhose du foie, et éventuellement vers un cancer du foie nommé carcinome hépatocellulaire.

#### *1.1.6.1) Hépatite chronique*

L'hépatite chronique se définit par la présence constante d'inflammation au niveau du foie, entraînant à long terme une perte d'élasticité de l'organe, développée à cause des attaques inflammatoires constantes que subit le tissu. On considère qu'une hépatite est chronique lorsque sa durée dépasse les six mois de maladie. L'inflammation étant très intense, la fatigue extrême est le principal symptôme d'une hépatite chronique. Chez certains patients, toutefois, la maladie passe inaperçue puisque le patient ne perçoit aucun symptôme majeur (Shepard et al, 2006; Chang, 2007; Mahoney, 1999).

#### *1.1.6.2) Cirrhose du foie*

La cirrhose du foie est une maladie grave du foie, souvent causée par une consommation excessive d'alcool, de drogues, ou bien qui survient suite à une infection virale à un virus d'hépatite. Elle peut aussi être d'origine génétique. Elle est caractérisée par une régénération

des hépatocytes en nodules causant une fibrose diffuse au niveau du foie, formant des cicatrices. L'organe est alors obstrué, et ceci l'empêche de remplir ses fonctions métaboliques correctement. La diminution de la fonction hépatique va provoquer des œdèmes au niveau des jambes et de l'abdomen, des jaunisses et des problèmes dans l'élimination des toxines présentes dans le corps lors d'absorption de médicaments, par exemple. La cirrhose est un grave problème de santé qui doit être pris au sérieux car plusieurs complications peuvent survenir, comme le coma et des hémorragies internes. La maladie peut être contrôlée en minimisant l'exposition avec les agents qui ont été responsables de l'apparition de la cirrhose (WHO, 2015; Canadian Liver Foundation, 2015).

#### *1.1.6.3) Carcinome hépatocellulaire*

Environ 25% des infections chroniques au virus de l'hépatite B progressent vers un cancer du foie. Lorsque le cancer origine d'un hépatocyte, il porte le nom de carcinome hépatocellulaire (HCC). En effet, plus de 50% de tous les cas de HCC sont attribuables à une infection par le virus de l'hépatite B. Les autres principales causes connues de HCC sont le virus de l'hépatite C, responsables de plus ou moins 30% des HCC, l'alcool et le tabac. C'est le génotype C du virus de l'hépatite B qui montre un risque accru quant au développement d'un HCC. Bien que le cancer du foie ne soit pas la forme de cancer la plus répandue en Amérique du Nord (1 personne sur 264), il reste un des cancers les plus meurtriers, avec un taux de survie à 5 ans de moins de 20%. Plus d'hommes que de femmes sont atteints par ce type de cancer. Il y a peu de symptômes apparents du HCC (c'est pourquoi il est souvent diagnostiqué trop tard) et sa progression est souvent trop avancée pour sauver la vie du patient. De nouveaux biomarqueurs doivent donc être développés afin de détecter plus rapidement le cancer du foie dans le parcours de vie des patients (Parkin et al, 2001; Lok et McMahon, 2007; Lavanchy, 2004; Canadian Liver Foundation, 2015; WHO, 2015).

#### *1.1.7) Modèles d'étude*

Outre le fait que l'humain est le seul hôte naturel du virus de l'hépatite B, il est possible d'infecter certains primates pour l'étude du VHB. Il est par contre très difficile d'infecter des cellules en culture avec le virus, c'est pourquoi certains modèles bien particuliers de culture

cellulaire ont été établis afin de palier à ce problème. Deux autres virus de la famille des *Hepadnaviridae*, le virus de l'hépatite de la marmotte (WHV) et le virus de l'hépatite du canard (DHBV) sont donc exploités pour fournir des réponses quant à la réplication virale, le pouvoir pathogène et l'évaluation de potentielles drogues antivirales. Ces virus ne sont toutefois pas capables d'infecter l'humain, ils sont donc sécuritaires pour une utilisation en laboratoire (Menne et Tennant, 1999; Guo et al, 2005; Mason et al, 1980; Sprengel et al, 1988).

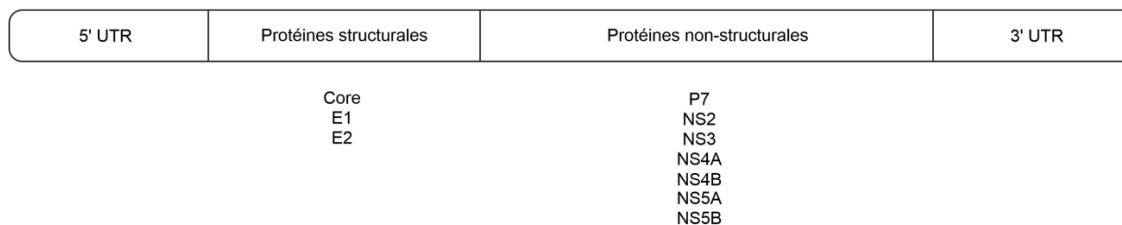
### ***1.1.8) Traitements et prévention***

Il est possible de prévenir l'hépatite chronique, la cirrhose du foie et le carcinome hépatocellulaire engendré par une infection virale au virus de l'hépatite B en se faisant vacciner contre ce pathogène. En effet, depuis 1982, un vaccin existe contre le VHB, et depuis, plusieurs pays l'ont adopté et l'ont même placé dans la routine de vaccination des nouveaux nés, enfants et adolescents. Dans les pays qui ont justement adopté cette cédule de vaccination, la prévalence de l'infection au VHB a fortement diminuée, tombant parfois même sous la barre des 1% de cas de porteurs chroniques. Dans le cas de la vaccination, on parle de prévention primaire puisqu'il est encore possible de prévenir l'infection avant qu'elle n'apparaisse. Toutefois, lorsqu'un individu est infecté, il n'est plus possible de le traiter par la vaccination; on doit donc faire appel à un programme de prévention secondaire au HCC, et donc le but est de freiner la progression de la maladie. On veut alors prévenir l'apparition d'hépatite chronique, de cirrhose ou de cancer du foie, en soignant l'infection virale. Pour y arriver, il existe plusieurs antiviraux sur le marché. Les traitements principaux consistent en un traitement constitué de prise d'interféron et d'analogues de nucléosides ou de nucléotides, lesquels servent d'inhibiteurs pharmacologiques à la polymérase virale. Bien qu'il existe certains traitements, ceux-ci ne sont pas parfaits; c'est pourquoi la recherche s'évertue à développer de meilleurs remèdes. Lorsque toutes ces mesures s'avèrent inefficaces et que le patient développe un carcinome hépatocellulaire, on doit alors passer aux traitements classiques de tous les cancers, c'est-à-dire la chirurgie, la chimiothérapie et la radiothérapie (Zanetti et al, 2008; Beasley, 2009; Lai et Yuen, 2013; Kim et al, 2015).

### ***1.1.9) Virus de l'hépatite C***

Comme il a été mentionné, le virus de l'hépatite C (VHC) est l'autre acteur principal dans le développement de carcinomes hépatocellulaires. En effet, le virus de l'hépatite C est un virus qui, comme le VHB, infecte les cellules du foie. En plus de partager avec le VHB la similarité au niveau du cancer engendré, le VHC induit également des hépatites chroniques, ainsi que des cirrhoses du foie, puisque ces deux virus ont le même tropisme (WHO, 2015). Toutefois, ces deux virus sont loin d'être identiques. Malgré la similarité de nomenclature avec le virus de l'hépatite B, ces deux espèces ne se retrouvent pas dans la même famille de virus. En effet, le virus de l'hépatite C est membre de la famille des *Flaviviridae*. Il est également le seul virus faisant partie du genre des *Hepacivirus* (Francki et al, 1991). Cette différence entre la classification de ces deux virus qui ont les mêmes effets au niveau pathologique est due à leur grande différence au niveau génomique. En effet, alors que le génome du VHB est composé d'ADN double-brin, le VHC préserve son information génétique sur un brin d'ARN de polarité positive. Cet ARN est dit polycistronique, c'est-à-dire qu'il est traduit en une seule polyprotéine qui sera par la suite clivée afin de former les différentes protéines du virus.

Le cycle de réplication virale du VHC implique d'abord l'entrée du virion dans la cellule hépatique. Suite à l'internalisation, le virus est décapsidé et le matériel génomique est relâché. C'est alors que l'ARN de polarité positive sera traduit par la machinerie cellulaire afin de produire les différentes protéines qui serviront à la réplication virale. Celle-ci a lieu au niveau du réticulum endoplasmique alors que les protéines virales vont former une structure que l'on appelle toile membranaire. Cette toile sert à l'ancrage des différentes protéines virales. L'ARN de polarité positive va par la suite servir de matrice pour la synthèse d'un deuxième brin d'ARN qui est cette fois-ci de polarité négative. L'utilité de ce second brin est de servir de matrice pour la synthèse des nombreux brins d'ARN positifs, lesquels seront subséquentement encapsidés lors de la formation de virions matures. Lorsque les virions sont formés, le virus peut alors s'échapper de la cellule hôte par exocytose (Dubuisson et Cosset, 2014; Poenisch et Bartenschlager, 2010).



**Figure 7 – Organisation génomique du virus de l'hépatite C**

Le génome du VHC code pour différentes protéines, dont certaines sont structurales et d'autres ont plutôt un effet sur la réplication virale, et est encadré de régions 5' et 3' non-traduites (UTR).

Le génome du virus de l'hépatite C code pour un total de dix protéines virales. Chacune d'elles a son rôle. D'abord, le cadre de lecture commence avec les protéines structurales. Comme le nom l'indique, ces protéines servent à former la structure du virion. La protéine core sert de nucléocapside pour la protection des nouveaux brins d'ARN synthétisés. Les protéines E1 et E2 sont celles qui vont envelopper cette capsid avant le relâchement des nouveaux virions matures formés au cœur de la cellule hôte. Par la suite viennent les protéines non-structurales qui sont celles qui ont des effets enzymatiques et qui vont donc servir à la bonne réplication du virus. De celles-ci, certaines ont un rôle bien connu. Par exemple, la protéine NS3 possède une double fonction. Elle est d'une part une hélicase qui sert au niveau de la réplication virale. D'autre part, NS3 est la protéase qui sert au clivage de la polyprotéine. L'activité protéase de NS3 est aidée par la protéine NS2 lors du clivage en amont de NS3 de la polyprotéine. Elle est aussi fortement aidée par le co-facteur NS4A qui va favoriser le clivage de la partie en aval de NS3 sur la polyprotéine. Une autre protéine non-structurale, la protéine NS4B, a comme fonction de former la toile membranaire, laquelle sert de lieu de réplication virale. La protéines NS5B a également un rôle très important, celui de répliquer le brin d'ARN de HCV afin de former plusieurs copies du génome. La protéine NS5B est donc la polymérase à ARN dépendante de l'ARN (Suzuki et al, 2007).

Par contre, certaines protéines du virus de l'hépatite C ont encore des fonctions plutôt méconnues, comme c'est le cas de la protéine P7 ainsi que de la protéine NS5A. Certaines hypothèses ont été mises de l'avant quant à leur rôle, mais les preuves manquent encore pour justifier celles-ci.

## 1.2) LE CANCER

### 1.2.1) Généralités et caractéristiques

Le cancer est un ensemble de maladies qui touchent différents organes ou tissus du corps humain. Selon le type cellulaire qui est atteint, on peut parler de carcinome ou de sarcome. D'un côté, le carcinome touche les cellules épithéliales alors que de l'autre côté, le sarcome est une atteinte des cellules mésenchymateuses du tissu conjonctif. On peut aussi parler d'astrocytome lorsque les cellules cancéreuses sont d'origine neuronale. De la même manière, le tissu touché va servir pour l'appellation du cancer. Par exemple, un ostéosarcome est un sarcome des cellules osseuses et un lymphome est un cancer qui tient son origine des lymphocytes, et donc un cancer du système lymphatique. Ainsi, selon le type de cellules et le tissu atteint, on aura différentes maladies (WHO, 2015). Par contre, toutes les cellules cancéreuses auront les mêmes caractéristiques, et c'est ce qui fait en sorte que l'on généralise souvent en une seule pathologie.

#### 1.2.1.1) Cellule cancéreuse

Le cancer est un état dans lequel les cellules dites cancéreuses vont perdre leur phénotype normal, et tendre vers un phénotype de cellules immortelles. Ceci est caractérisé de plusieurs façons. Pour commencer, la cellule va perdre la balance normalement très régulée entre la prolifération (division cellulaire) et la mort cellulaire par apoptose. Le débalancement se fait au profit de la prolifération. La cellule va croître sans limite alors qu'il n'y aura pas de régulation effectuée par les processus apoptotiques (Lowe et al, 2004; Adams et Cory, 2007). Normalement, une cellule primaire cesse de proliférer après un nombre prédéterminé de divisions cellulaires. L'instabilité génomique et la présence de plusieurs mutations consécutives va aider à acquérir ce phénotype. De plus, pour que la cellule puisse se diviser à un rythme incontrôlé, elle devra modifier son métabolisme énergétique (Warburg, 1956; Vander Heiden et al, 2009; Jones et Thompson, 2009; Christofk et al, 2008).

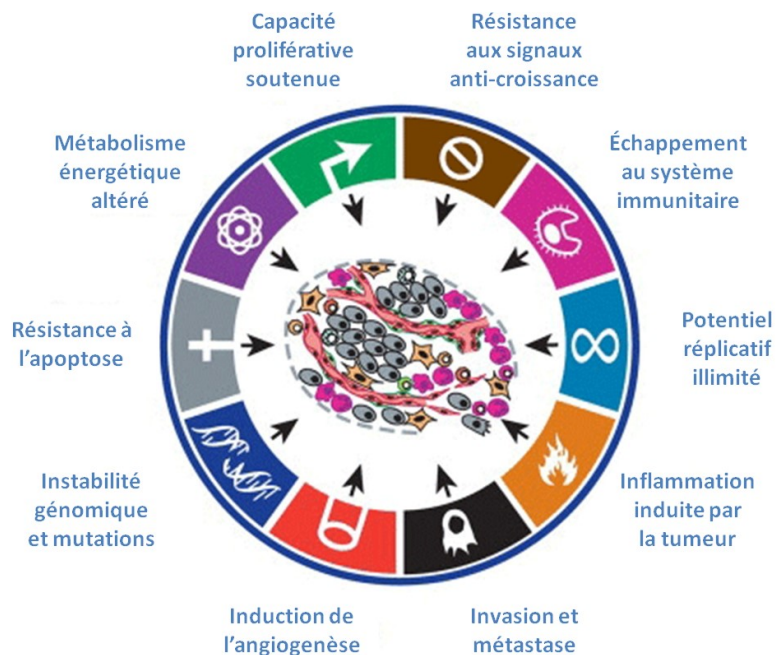
Une autre caractéristique des cellules cancéreuses est qu'elles perdent leur inhibition de contact. Normalement, une cellule va arrêter de croître lorsqu'elle sera en contact avec d'autres cellules, signe que le tissu est formé. En perdant l'inhibition de contact, la cellule va



se diviser sans limite, et de cette façon, les cellules formeront une masse (ou tumeur) (Hanahan et Folkman, 1996). De la même manière, les cellules cancéreuses deviennent autosuffisantes en signaux de croissance (Witsch et al, 2010; Perona, 2006; Hynes et MacDonald, 2009). Donc, en temps normal, la cellule aura besoin de signaux de survie provenant de l'extérieur pour lui dire de se diviser. Avec le phénotype cancéreux, la cellule sera capable de se fournir ses propres signaux et donc de se diviser sans aide extérieure. Inversement, la cellule deviendra également résistante aux signaux provenant de l'extérieur qui lui indiqueront qu'elle doit arrêter de croître. Donc, encore une fois, la cellule entrera en prolifération incontrôlée.

Les cellules cancéreuses créent un environnement inflammatoire dans lequel elles s'enrichissent. Ce milieu est favorable à leur réplique. Aussi, quand la cellule aura acquis toutes ces caractéristiques de cellule immortelle, le système immunitaire voudra essayer d'éliminer les cellules problématiques. En effet, en temps normal, lorsque les cellules de l'immunité perçoivent d'autres cellules présentant un phénotype inhabituel, elles les éliminent pour préserver l'état normal de l'organisme. Par contre, dans le cas de ces cellules cancéreuses, le système immunitaire sera incapable de les détruire (Swann et Smyth, 2007; Yang et al, 2010; Mougiakakos et al, 2010; Ostrand-Rosenberg et Sinha, 2009).

Finalement, les cellules cancéreuses acquièrent des propriétés invasives et migratoires, qui font en sorte qu'elles peuvent prendre de l'expansion dans le tissu d'origine, mais aussi qu'elles peuvent migrer dans l'organisme (par la lame basale) pour aller coloniser d'autres organes et former ce que l'on appelle des métastases (Talmadge et Fidler, 2010). Les cellules cancéreuses vont induire l'angiogenèse qui est le processus de formation de nouveaux vaisseaux sanguins. Ces vaisseaux vont servir à la fois pour l'apport de nutriments à la tumeur et pour la migration pour former des métastases (Hanahan et Folkman, 1996).



**Figure 8 - Les caractéristiques d'une cellule cancéreuse**

Une cellule acquiert plusieurs caractéristiques particulières qui lui confèrent le titre de cellule cancéreuse. Toutes ces caractéristiques favorisent la survie de la cellule. Adapté de (Hanahan et Weinberg, 2000).

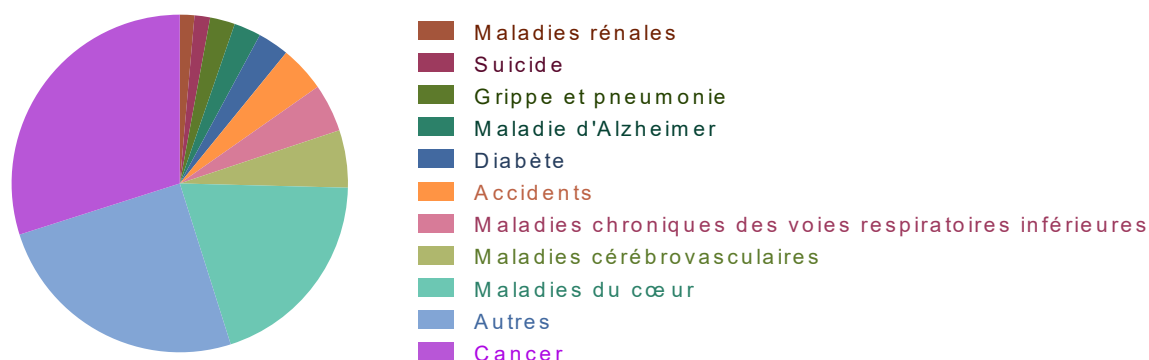
#### *1.2.1.2) Tumeur*

Un cancer, ou tumeur maligne, est composé de plusieurs types cellulaires distincts qui vont s'entraider pour favoriser la croissance et la progression tumorale. Le cancer va apparaître après plusieurs mutations et modifications successives d'une seule cellule (croissance monoclonale). Le fait que l'on retrouve plusieurs cellules distinctes dans une tumeur est dû à la grande instabilité génomique des cellules. Chacune va acquérir des mutations spécifiques qui peuvent être différentes des autres cellules, mais toutes proviennent de la cellule cancéreuse de départ. Ceci va favoriser la coexistence de cellules transformées et donc l'hétérogénéité de la tumeur (Hahn et Weinberg, 2002).

#### *1.2.2) Épidémiologie du cancer*

Le cancer est un terme général qui caractérise un grand nombre de maladies. Il représente la principale cause de décès et de complications médicales dans le monde. Le nombre de nouveaux cas dans une année se chiffre à une quinzaine de millions d'individus. Environ 8 à

9 millions de patients cancéreux décèderont par année des suites de la maladie (Stewart et Wild, 2014). On considère que dans les prochaines décennies, l'incidence devrait augmenter d'environ 70%.



**Figure 9 - Compilation des causes de décès au Canada**

Le cancer figure en tête de liste des causes de mortalité des Canadiens. Différentes autres maladies chroniques sont responsables de décès, en plus des causes accidentelles et des suicides. Adapté de Statistique Canada.

Plus d'hommes que de femmes sont touchés par le cancer. La figure 9 dresse un portrait des cancers les plus fréquents pour chaque sexe, ainsi que les cancers les plus meurtriers. Le cancer apparaît plus fréquemment chez les personnes en haut de 50 ans, mais il peut survenir à tout âge.

A	Hommes	Femmes
	Poumon	Poumon
Prostate	Sein	
Colorectal	Colorectal	
Estomac	Col de l'utérus	
Foie	Estomac	

B	Cancer	Décès
	Poumon	1 590 000
Foie	745 000	
Estomac	723 000	
Colorectal	694 000	
Sein	521 000	
Œsophage	400 000	

**Figure 10 - Statistiques sur le cancer**

**A** Les cancers les plus fréquemment observés chez les hommes et chez les femmes (mondialement). **B** Les cancers les plus meurtriers et le nombre de décès associés. (Stewart et Wild, 2014)

Outre le processus normal de vieillissement, on classe les agents qui causent le cancer en trois grandes catégories. Ces agents sont des facteurs endogènes qui contribuent au cancer, en combinaison avec les facteurs génétiques. Ils sont les cancérrogènes physiques (rayonnements ultraviolet et radiations ionisantes), les cancérrogènes chimiques (amiante, tabac, aflatoxine et arsenic) et les cancérrogènes biologiques (bactéries, virus ou parasites).

Une grande majorité de décès par cancer sont attribuables à certains facteurs de risque de nature comportementale ou alimentaire. En effet, 30% des décès seraient provoqués par un cancer engendré par un indice de masse corporelle au-dessus des valeurs normales, une alimentation pauvre en fruits et légumes, un manque d'exercice physique ainsi qu'une consommation excessive de tabac et d'alcool. De ceux-ci, le tabac reste le principal meurtrier. Il est responsable de plus de 20% de décès par cancer (tous types confondus) et de manière plus importante de 70% de mortalité par cancer du poumon. Certaines infections virales ou même bactériennes vont également causer des cancers. En effet, il existe sept virus oncogènes

chez l'humain, ainsi que des bactéries comme *Helicobacter Pylori*, qui sont responsables d'environ 20% des cancers. Les agents infectieux les plus fréquemment rencontrés sont les virus du papillome humain (VPH), de l'hépatite B et de l'hépatite C (VHB et VHC). Dans les pays à revenu faible ou intermédiaire, les infections sont responsables de plus de 20% des décès par cancer (De Martel et al, 2012).

Virus	Cancer (s)
Virus du papillome humain (VPH)	Cancer de la tête et du cou et cancer du col de l'utérus
Virus de l'hépatite B (VHB)	Carcinome hépatocellulaire
Virus de l'hépatite C (VHC)	Carcinome hépatocellulaire
Virus Epstein-Barr (EBV)	Lymphomes
Virus T-lymphotropique humain (HTLV-1)	Leucémies
Herpesvirus humain 6 (HHV6)	Sarcome de Kaposi
Polyomavirus des cellules de Merckel (MCPV)	Carcinome des cellules de Merckel

**Figure 11 - Les virus causant des cancers chez l'humain**

Les sept virus humains qui causent le cancer chez l'humain et les cancers qui leur sont associés.

#### 1.2.2.1) Les virus et le cancer

Plusieurs mécanismes connus expliquent par quelles façons les virus sont capables de transformer une cellule saine qu'ils infectent en une cellule cancéreuse. En premier lieu, il y a l'activation de voie de survie cellulaire ou encore l'inhibition de suppresseurs de tumeur. Ces processus agissent directement au niveau de la prolifération de la cellule. Ces mécanismes sont employés par les protéines E6 et E7 du virus du papillome humain. Par exemple, la protéine E7 est capable de lier la protéine cellulaire pRb, laquelle est responsable du contrôle du cycle cellulaire par inhibition de la réplication de l'ADN. Lorsque liée par E7, pRb n'est plus en mesure d'accomplir ses fonctions. La croissance des cellules devient donc incontrôlée. De manière semblable, la protéine E6 du VPH est capable d'aller activer des voies de survie cellulaire comme les voies Ras/Raf/MAPK et EGFR (Cornet et al, 2012). Aussi, E6 lie le suppresseur de tumeur p53 et inhibe ses fonctions (Bernard et al, 2011). Ensuite, certains virus peuvent ajouter un promoteur fort en amont d'un oncogène ou agir

positivement sur une région activatrice (enhancer) d'un oncogène (Swann et Smyth, 2007; Yang et al, 2010; Mougiakakos et al, 2010; Ostrand-Rosenberg et Sinha, 2009). Ces mécanismes sont observés chez un virus causant des cancers chez les oiseaux. Le virus est capable d'activer la transcription de l'oncogène c-myc directement au niveau du promoteur (Hayward et al, 1981). Plusieurs mécanismes restent toutefois à élucider.

### ***1.2.3) Dépistage et traitements des cancers***

#### *1.2.3.1) Dépistage*

Selon le cancer que l'on souhaite dépister, différentes interventions existent afin d'obtenir un diagnostic. Premièrement, il est possible de voir certaines tumeurs par imagerie médicale. Plusieurs types de techniques d'imagerie existent pour le dépistage des tumeurs. D'abord, les techniques d'imagerie nucléaires peuvent être efficaces pour trouver une tumeur à l'aide de réactifs radiopharmaceutiques. Ces réactifs peuvent déceler une tumeur de deux façons: soit la tumeur va absorber une plus grande quantité du réactif que le tissu environnant, soit c'est le tissu qui va absorber plus fortement le réactif. Différents réactifs existent, selon le tissu que l'on souhaite observer et le type de test effectué. Les plus communs sont l'iode radioactif, le technétium 99m, le strontium, le gallium et le thallium. Ensuite, certaines tumeurs peuvent être diagnostiquées à l'aide d'IRM (imagerie par résonance magnétique). L'IRM est efficace pour identifier des tumeurs du cerveau, des os, des seins et plusieurs autres. Pour y arriver, on utilise des forces puissantes de magnétisme et des ondes radio-électriques, souvent jumelées à des agents de contrastes afin de bien voir la plupart des structures du corps humain en trois dimensions. Le principe est que les différents tissus réagissent inégalement aux ondes magnétiques. On peut aussi utiliser les différentes techniques d'imagerie classiques, comme l'endoscopie et l'échographie afin d'observer certaines structures internes (National Cancer Institute, 2015; Medical Imaging & Technology Alliance, 2015).

Deuxièmement, les biopsies sont aussi utilisées afin de diagnostiquer un cancer. La biopsie est un examen médical invasif, qui consiste à prélever un échantillon du tissu pour l'examiner au microscope, et ainsi voir si on y retrouve des cellules cancéreuses. Il est possible d'effectuer une biopsie sur n'importe quel tissu du corps. Par contre, selon l'endroit,

des types de biopsie différents seront effectués (WHO, 2015; Société Canadienne du Cancer, 2015).

Finalement, une des méthodes les plus simples et non-invasives de détecter un cancer est de faire la recherche de biomarqueurs spécifiques pour la maladie, soit par analyse sanguine ou encore par analyse d'urine. Ces biomarqueurs sont des molécules biochimiques détectables, quantifiables et qui sont associées à une maladie spécifique. Un bon biomarqueur doit être facilement quantifiable et corrélé avec le stade de la maladie. Plusieurs biomarqueurs sont utilisés en clinique. Par exemple, l'antigène prostatique spécifique (APS) est utilisé pour la détection du cancer de la prostate. L'antigène tumoral 125 (CA125) est aussi utilisé depuis les années 1980 comme marqueur du cancer de l'ovaire. Plusieurs recherches en cours visent à développer de nouveaux biomarqueurs pour certains cancers, mais également pour d'autres maladies (WHO, 2015; Société Canadienne du Cancer, 2015).

#### *1.2.3.2) Traitements*

Actuellement, la plupart des traitements contre le cancer ne ciblent pas spécifiquement le type de cancer du patient, ni la cause à l'origine du cancer. Les trois principaux types de traitements sont la chirurgie, la chimiothérapie et la radiothérapie. Il existe aussi l'hormonothérapie, la greffe de cellules souches et la thérapie biologique.

Tout d'abord, la chirurgie consiste en l'ablation de la tumeur, mais parfois aussi de l'organe au complet. C'est la forme de traitement la plus ancienne du cancer, mais encore à ce jour, elle est utilisée dans plus de la moitié des cas. Le retrait de la tumeur est souvent accompagné d'une autre intervention (WHO, 2015; Société Canadienne du Cancer, 2015).

Pour les traitements de radiothérapie, le patient recevra une forte dose de radiation afin d'éliminer les cellules cancéreuses. Des séances successives de radiothérapie endommageront les cellules cancéreuses, mais aussi les cellules saines. Par contre, à la différence des cellules normales, les cellules cancéreuses ne pourront pas se réparer entre les séances, car elles sont en défaillance de système de réparation de l'ADN. C'est pourquoi ces cellules seront plus affectées que les cellules saines, et donc il sera possible que le patient

guérissent grâce à un tel traitement (Société Canadienne du Cancer, 2015; Cancer Research UK, 2014).

La chimiothérapie est un traitement systémique, et donc qui provoque plusieurs effets secondaires chez les patients qui en reçoivent. Elle consiste en l'administration de substances toxiques qui détruiront les cellules cancéreuses. De la même manière que la radiothérapie, elle causera aussi des dommages au niveau des cellules saines; c'est ce qui explique les nombreux effets secondaires ressentis (OMS, 2015). Par exemple, en clinique, le méthotrexate est couramment utilisé. Il est un antimétabolite, donc il s'attaque au métabolisme accru des cellules cancéreuses. Plus précisément, il va intervenir au niveau de la synthèse des bases azotées et donc au niveau de la synthèse de l'ADN. Comme plusieurs agents chimiothérapeutiques, elle est souvent utilisée en combinaison avec d'autres molécules, ou avec une autre intervention comme la chirurgie, la radiothérapie ou la thérapie biologique (NIH, 2014).

La thérapie biologique vise à attaquer les cellules cancéreuses à l'aide du principe selon lequel le système immunitaire aide à détruire les cellules défaillantes. Elle était auparavant utilisée pour traiter d'autres pathologies, principalement des maladies auto-immunes, mais depuis quelques années, on voit apparaître cette solution comme traitement du cancer. Les thérapies à l'interféron sont souvent utilisées afin de stimuler les cellules immunitaires à combattre les cellules cancéreuses (NIH, 2014).

Bien que la majorité des traitements actuels soient non-spécifiques au type de cancer de l'individu à traiter, quelques traitements ciblés apparaissent peu à peu sur le marché. En effet, ces traitements sont constitués de molécules spécifiques à la cellule cancéreuse qui doit être détruite. Généralement, ces molécules vont affecter les voies de transduction de signal de vie ou de mort cellulaire. Les molécules les plus populaires qui existent jusqu'à présent sont les inhibiteurs de tyrosine kinases. Les récepteurs à tyrosine kinases envoient les signaux de réplication à la cellule et sont dérégulés dans plusieurs cancers. Aussi, certaines molécules biologiques sont utilisées comme thérapie ciblée. Par exemple, plusieurs anticorps monoclonaux spécifiques à un type de cellules cancéreuses sont sur le marché. Ils visent à



détruire ces cellules à l'aide de l'immunité du patient. Ces médicaments seront d'usage plus courant puisqu'ils induisent généralement moins d'effets secondaires que les soins chimiothérapeutiques largement utilisés (NIH, 2014; Société Canadienne du Cancer, 2015).

### 1.3) ÉPISSAGE ALTERNATIF

#### *1.3.1) Définition et mécanisme général*

Lors de la transcription de l'ADN en ARNm, le transcrit produit n'est pas encore mature. Ce pré-ARNm doit être coiffé, polyadénylé et épissé afin de pouvoir être traduit en protéine. L'épissage est l'action de retirer les introns (régions non-codantes) dans un pré-ARNm afin de liguer les exons (régions codantes) les uns avec les autres. L'épissage alternatif, lui, est le processus qui, en plus de retirer les introns, réarrange le pré-ARNm de sorte à produire plus d'un ARNm à partir d'un même pré-ARNm. Certains exons pourront être retirés, par exemple, afin de produire une plus grande variété de transcrits (Brett et al, 2002). Les exons constitutifs sont ceux qui seront toujours présents dans l'ARNm final, alors que les exons cassettes sont ceux qui peuvent être inclus ou exclus de l'ARNm mature par épissage alternatif. Certains mécanismes vont favoriser l'inclusion ou l'exclusion de ces exons cassettes (Smith et Nadal-Ginard, 1989).

L'épissage alternatif est d'une grande importance au niveau de la cellule, car plus de 95% de tous les gènes humains subissent l'épissage alternatif (Wang et al, 2008; Pan et al, 2008). Ceci confère à la cellule une grande variété de fonctions au niveau protéique, tout en gardant le matériel génétique le plus simple possible. En effet, s'il fallait avoir un gène pour chaque protéine produite, le génome humain serait beaucoup plus gros qu'il ne l'est. Environ 1.1% du génome encode pour les exons, et près de 24% correspond aux introns. En effet, les portions introniques des pré-ARNm sont beaucoup plus longues que les régions exoniques. Dans les 22 000 gènes humains, il y aurait en moyenne huit à douze exons par gènes, avec une longueur moyenne de 130 paires de bases. Les introns, eux, sont longs d'environ mille à trois mille paires de bases. Un transcrit d'un gène peut représenter plus d'un millier de patrons d'épissage alternatif différent. Il est donc aisé de constater l'énorme variabilité protéique qui découle de l'épissage alternatif (Black, 2003; Barash et al, 2010).

Parmi toutes les cellules, ce sont les cellules eucaryotes des espèces supérieures qui ont le plus grand nombre de gènes subissant l'épissage alternatif. Ceci laisse croire que l'épissage alternatif est un mécanisme moléculaire important au niveau de l'évolution des espèces. En effet, bien que les différences génétiques entre l'Homme et le chimpanzé soient

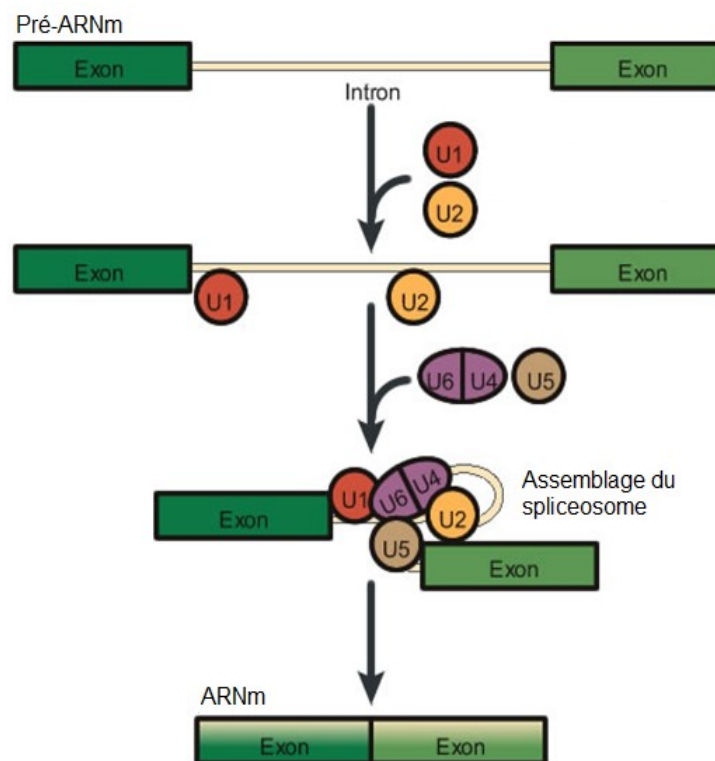
de seulement 1%, on estime que les différences d'épissage sont de 7%, ajoutant ainsi un niveau de complexité à l'organisme (Irimia et al, 2007). L'épissage alternatif est un élément central dans l'expression génique puisqu'il peut contrôler tous les mécanismes cellulaires en produisant des protéines variées à partir d'un seul et même gène (Kelemen et al, 2013).

#### *1.3.1.1) Mécanisme d'action*

L'épissage est une réaction biochimique qui se produit à l'intérieur du noyau de la cellule. Elle se fait en deux étapes et n'implique pas d'endonucléase ni de ligase pour accomplir l'assemblage des régions épissées. En effet, l'épissage se produit à l'aide d'une réaction chimique de transestérification qui comprend deux attaques nucléophiliques. Pour qu'il y ait épissage, il doit d'abord y avoir reconnaissance des régions nécessitant le retrait d'un intron. Ceci se produit à l'aide de séquences de reconnaissance nommées 3'ss et 5'ss. Cette reconnaissance se fait par des petits ARN nucléaires (snARN) et est aidée par des facteurs que l'on appelle facteurs d'épissage. Il existe plusieurs facteurs d'épissage qui reconnaissent différentes séquences. Lorsque ces séquences sont reconnues par ces éléments, il y a formation de la structure responsable de la réaction d'épissage, le spliceosome. Le spliceosome va réarranger l'ARN de manière bien précise afin de positionner les nucléotides pour les attaques nucléophiliques. La première étape consiste en l'attaque nucléophile du 2'OH de l'adénosine au site de branchement. La deuxième étape est l'attaque nucléophile du 3'OH du premier exon vers le deuxième exon. Lors de l'épissage, il y a formation d'un lasso, appelé lariat, formé de l'intron et de toutes autres séquences nucléotidiques retirées de l'ARN (Cieply et Carstens, 2015; Black, 2003; Chen et Cheng, 2012).

Le spliceosome est un gros complexe composé de petits ARN nucléaires (snARN) et de plusieurs protéines. Il est une structure dynamique, qui se transforme lorsque chacun de ses composants lie le pré-ARNm (Wahl et al, 2009). Les snARN combinés à des protéines forment des complexes plus grands que l'on appelle snRNP (petits ribonucléoprotéines nucléaires). Ceux-ci sont appelés U1, U2, U4, U5 et U6 et ils ont pour rôle de changer la structure secondaire adoptée par le pré-ARNm afin de favoriser les réactions d'attaques nucléophiliques (Jurica et Moore, 2003). L'assemblage du spliceosome suit généralement un ordre précis qui implique la liaison des différents snRNP. Tout d'abord, il y a reconnaissance

du snRNP U1 au site 5'ss du pré-ARNm, suivi par la liaison en 3'ss du snRNP U2. Ceux-ci vont alors se joindre ensemble pour rapprocher les deux sites. C'est à ce moment que les snRNP U4, U5 et U6 vont aller compléter le spliceosome. Il y aura mouvement de U1 et de U4 lorsque le spliceosome aura pris place et la réaction de transestérification aura lieu, formant l'ARNm mature et le lariat (Shcherbakova et al, 2013). Les snRNP pourront être récupérés pour un épissage ultérieur. La liaison du spliceosome au pré-ARNm, tout comme l'épissage, a lieu en même temps que la transcription (Chen et Cheng, 2012).



**Figure 12 - Réaction d'épissage**

L'intron du pré-ARNm à être épissé est d'abord reconnu en 5'ss par le snRNP U1. Ensuite, il y a reconnaissance par le snRNP U2 en 3'ss. Suite à cela, le tri-snRNP U4/U5-U6 est recruté. La liaison du tri-snRNP déloge U1 et U4. Ce faisant, les sites se positionnent pour la première attaque nucléophilique. La deuxième attaque nucléophilique lie les deux exons entre eux, et le résultat final est l'ARNm mature en plus du lariat, formé de l'intron épissé. Adapté de (Loria et al, 2003).

### *1.3.1.2) Mécanismes de régulation*

Les sites d'épissage en 5' et en 3' servent de reconnaissance pour les éléments du spliceosome. Ils sont situés aux jonctions entre les exons et les introns. Le 5'ss est situé à la première jonction, soit entre l'exon 1 et l'intron. Il est caractérisé par un dinucléotide GU positionné dans l'intron. Celui-ci est compris dans une plus grande séquence consensus. Cette séquence sert de reconnaissance pour le snRNP U1. Pour sa part, le 3'ss est situé à l'autre extrémité de l'intron, soit à son extrémité 3'. Il est donc situé dans la jonction entre l'intron et l'exon 2. Le 3'ss est un peu plus complexe que le 5'ss, car il comprend trois séquences conservées qui serviront à la reconnaissance du spliceosome. Le site de branchement est la première de ces séquences. Elle sert à la liaison du snRNP U2. Ce site de branchement est suivi d'une séquence de polypyrimidines qui se termine par un AG final, qui définit la frontière entre l'intron et l'exon. Le 3'ss est lié par un groupe de protéines régulatrices (facteurs d'épissage). Le snRNP U2 lie la séquence de polypyrimidines et l'AG terminal via sa sous-unité U2AF. Suite à la reconnaissance des toutes ces séquences bien précises par les snRNP U1 et U2, il y aura assemblage final du spliceosome tel que décrit plus haut (Cieply et Carstens, 2015; Black, 2003).

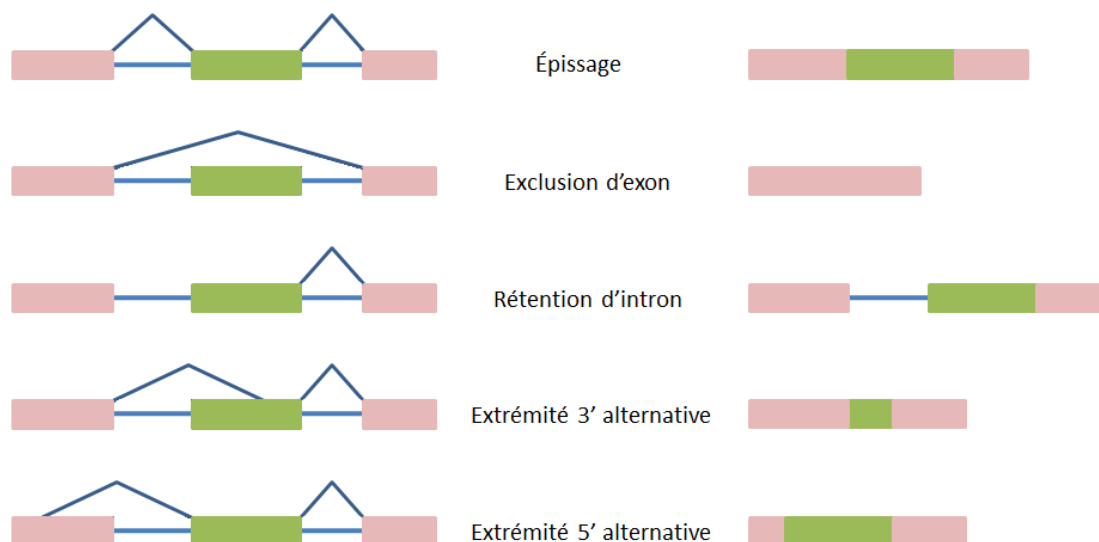
Outre ces séquences classiques nécessaires à l'épissage, on retrouve généralement des séquences activatrices ou inhibitrices qui régulent l'épissage de certains exons. Ces séquences sont reconnues par des facteurs d'épissage qui vont favoriser ou bien déstabiliser la formation du spliceosome. Ces séquences régulatrices peuvent être positionnées au niveau des introns ou des exons. Les protéines SR vont se lier aux activateurs d'épissage et vont agir en favorisant l'inclusion de l'exon (Busch et Hertel, 2012; Manley et Tacke, 1996; Graveley, 2000; Long et Caceres, 2009). À l'inverse, les protéines hnRNP sont des inhibiteurs, et donc favorisent l'exclusion de l'exon (Yoo et al, 2008; Han et al, 2007; Sanz-Cameno et al, 2006). De manière générale, les séquences activatrices de l'épissage sont utilisées dans l'épissage constitutif alors que les séquences inhibitrices vont servir au niveau de la sélection des exons retenus lors de l'épissage alternatif (Wang et al, 2006; Wang et Burge, 2008).

L'épissage alternatif est également régulé par l'état de la chromatine et la modification des histones, car ces éléments favorisent le regroupement des constituants du spliceosome

(Luco et al, 2011). Il semblerait également que les nucléosomes soient plus fortement enrichis au niveau des jonctions entre les exons et les introns, et donc qu'ils définissent les exons (Andersson et al, 2009; Chodavarapu et al, 2010; Dhimi et al, 2010; Schwartz et al, 2009). De la même manière, l'épissage est régulé par la transcription par l'ARN polymérase II. En effet, l'assemblage co-transcriptionnel du spliceosome sert d'élément régulateur à la détermination des choix des sites d'épissage utilisés. Les sites d'épissage qui seront transcrits rapidement auront plus tendance à être utilisés. Le domaine C-terminal de l'ARN polymérase II est responsable du recrutement de plusieurs éléments du spliceosome et de facteurs d'épissage jouant un rôle dans la régulation des sites d'épissage. De plus, il semble que les gènes qui ont le plus tendance à être épissés co-transcriptionnellement soient les transcrits avec les plus longs introns (Corden et Patturajan, 1997; Morris et Greenleaf, 2000; Robert et al, 2002).

### ***1.3.2) Les types d'épissage alternatif***

Il existe plusieurs types d'épissage alternatif. Le plus prédominant dans la cellule de mammifère est l'exclusion d'exon. Comme son nom l'indique, l'exclusion d'exon vise à retirer un exon du produit final. De cette manière, des domaines entiers d'une protéine peuvent être retirés. Aussi, l'exclusion d'exon peut causer un changement dans le cadre de lecture ouvert, si le nombre de paires de bases retirées n'est pas un multiple de trois. Ce faisant, la séquence de la protéine produite sera complètement différente de la protéine dont l'exon n'a pas été retiré (Black, 2003; Barash et al, 2010). L'exclusion d'exon compte pour 40% de tous les événements d'épissage d'une cellule (Keren et al, 2010; Alekseyenko et al, 2007; Sugnet et al, 2004).



**Figure 13 - Les différents types d'épissage alternatif d'un pré-ARNm**

L'épissage alternatif peut se produire de plusieurs façons afin de modifier l'ARNm mature et ainsi créer une diversité de transcrits plus importante (Black, 2003).

Un autre mode d'épissage alternatif est la rétention d'intron. La rétention d'intron consiste à garder un des introns dans la séquence finale de l'ARNm. Ce faisant, la protéine finale pourra être plus grande, et elle pourra même acquérir un nouveau domaine. Par contre, dans la plupart des cas, la rétention d'intron se produit au niveau des régions non-traduites (UTR) (Galante et al, 2004).

Dans certains cas, les exons pourront être gardés partiellement. Dans le cas d'un épissage de type extrémité 3' alternative, c'est l'exon en 3' qui perdra une portion de sa séquence. De cette façon, on conservera seulement une fraction de cet exon. De la même manière, lors d'un épissage alternatif de type extrémité 5' alternative, c'est l'exon en 5' qui sera fractionné. Dans ces deux cas, on aura encore une fois un changement dans la longueur du transcrit et donc de la protéine produite (Blencowe, 2006).

Dans tous les types d'épissage alternatif, le changement dans la séquence peut engendrer plusieurs modifications fonctionnelles. Par exemple, l'ajout ou le retrait d'une

séquence dans l'ARNm mature peut modifier le cadre de lecture ouvert. Ce faisant, la séquence en acides aminés de la protéine peut être complètement différente, mais aussi un codon stop peut apparaître ou disparaître et modifier la longueur de la protéine produite. L'apparition précoce d'un codon stop dans un cadre de lecture d'une protéine peut également mener à la voie de dégradation NMD qui vise à contrôler la qualité des ARNm en cellule (Maquat et Carmichael, 2001). Le tiers des événements d'épissage alternatif mènerait à un codon terminateur prématuré au niveau de l'ARNm. Celui-ci sera reconnu et donc dégradé par le NMD (non-mediated decay) (Lareau et al, 2007; Lewis et al, 2003). La protéine peut aussi acquérir ou perdre un domaine fonctionnel qui lui confèrera des fonctions différentes. Par contre, dans plusieurs cas, l'épissage alternatif va causer un changement dans la longueur des UTR. Ce faisant, l'épissage ne changera pas la protéine produite mais modulera son niveau d'expression en changeant l'efficacité de traduction de l'ARNm, sa stabilité et même sa localisation (Licatalosi et Darnell, 2010; Nilsen et Graveley, 2010; Isken et Maquat, 2008).

### ***1.3.3) Implication dans différentes maladies***

L'épissage alternatif va moduler un grand nombre de processus biologiques incluant la prolifération et l'apoptose, la différenciation des cellules, le développement, le cycle circadien et la réponse aux stress cellulaires (Barbosa-Morais et al, 2012; Kalsotra et Cooper, 2011). De par son grand impact sur le transcriptome, et donc sur le protéome cellulaire, l'épissage alternatif peut être responsable de plusieurs pathologies chez l'Homme lorsqu'il est dérégulé. En effet, un manque dans la régulation des processus d'épissage peut avoir de grandes conséquences sur la cellule.

Plusieurs maladies sont associées à un dérèglement dans les processus d'épissage alternatif. Notamment, certaines pathologies sont occasionnées par un défaut dans le spliceosome, rendant l'épissage moins fonctionnel. Par exemple, une maladie de l'œil, appelée rétinite pigmentaire, est causée par des mutations dans certains facteurs d'épissage (PRPF3, PRPF8, PRPF31). Aussi, un des exemples les plus importants de maladies causées par un défaut au niveau du spliceosome est l'amyotrophie spinale (SMA). La SMA est une grave maladie caractérisée par une atrophie musculaire et donc une mobilité réduite. La SMA est la première cause de décès par maladies génétiques chez les enfants. Elle se produit car il



Il y a une mutation dans le gène de SMN1 qui code pour une des protéines SMN impliquée dans l'assemblage des snRNP. En absence d'une SMN1 fonctionnelle, c'est la protéine SMN2 qui va compenser en étant plus fortement exprimée. Par contre, l'exon 7 de SMN2 est alternatif et donc seulement 10 à 20% des SMN2 produites seront réellement fonctionnelles (Pearn, 1980). Par ailleurs, certaines thérapies ciblant cet exon 7 de SMN2 ont été développées afin de traiter les individus atteints par cette maladie (Zhou et al, 2012). Également, des leucémies myéloïdes peuvent être causées par un défaut dans les constituants du spliceosome, entre autre le facteur d'épissage SF3B1 (Cazzola et al, 2013).

Plusieurs maladies neurodégénératives et musculaires sont associées à un défaut dans l'épissage alternatif. Par exemple, l'épissage a un grand rôle à jouer dans l'apparition des dystrophies myotoniques de types 1 et 2. Ces dystrophies sont caractérisées par une défaillance au niveau cardiaque, musculaire et neuronal. Ce sont principalement les facteurs d'épissage MBNL1 et CELF1 qui sont affectés, en étant plus fortement traduits dû à une stabilité accrue de leur ARNm (Cooper et al, 2009; Poulos et al, 2011; Mohan et al, 2014). Aussi, certaines formes de la maladie de Parkinson sont associées à une dérégulation de l'épissage. En effet, on retrouve fréquemment des mutations dans les sites de régulation de l'épissage de la protéine tau, qui est impliquée dans les maladies de Parkinson et d'Alzheimer. Ces mutations prédisposent à une agrégation de cette protéine, causant ainsi la maladie (Mills et Janitz, 2012; Yap et Makeyev, 2013; Fu et al, 2013). Un autre exemple important de maladie neurodégénérative grave provoquée par un changement dans l'épissage alternatif est la sclérose latérale amyotrophique (SLA). Le mécanisme est encore mal connu, mais il semble qu'une protéine impliquée dans l'épissage, TDP-43, ne soit plus localisée au noyau dans les neurones des patients atteints de SLA. Il est donc fort probable que TDP-43 perd ses fonctions au niveau de l'épissage dans les cellules malades (Neumann et al, 2006). Également, il semble que certains facteurs d'épissage tels que hnRNP H soient séquestrés dans des loci chez les patients atteints de SLA. Des études approfondies doivent toutefois être menées afin de comprendre les mécanismes exacts (Lee et al, 2013).

Une modification de l'épissage alternatif peut causer des problèmes au niveau du coeur en affectant son développement et en causant des problèmes de coagulation sanguine,

des déséquilibres au niveau de l'homéostasie du cholestérol, ainsi que des scléroses systémiques. Par exemple, des mutations au niveau du facteur d'épissage SF2 causera des cardiomyopathies chez certains patients (Cooper, 2005; Lara-Pezzi et al, 2012; Bogdanov, 2006; Medina et Krauss, 2013).

#### *1.3.3.1) L'importance de l'épissage alternatif dans le cancer*

Il est désormais bien connu que les cellules cancéreuses le deviennent en acquérant diverses caractéristiques qui changeront le phénotype de la cellule saine. On retrouve parmi ces caractéristiques la capacité de proliférer de manière soutenue, indépendamment des signaux externes, la résistance à l'apoptose, et bien d'autres qui favorisent la survie et la croissance des cellules. Récemment, il a été identifié qu'une autre caractéristique était importante pour l'acquisition de ce phénotype cancéreux: les modifications dans le mécanisme d'épissage alternatif (Ladomery, 2013). En effet, il a été démontré que l'épissage alternatif était souvent dérégulé dans plusieurs cancers, notamment les cancers de l'ovaire, de la prostate et du sein (Lapuk et al, 2014; Venables et al, 2008). Les auteurs de ces études ont analysé par RT-PCR (réaction en chaîne par polymérase suivant une rétro-transcription) à haut débit les patrons d'épissage alternatif de 26 tumeurs du sein. Ils ont découverts que 41 événements d'épissage alternatif étaient associés au cancer du sein sur les 600 gènes qui avaient été sondés (Venables et al, 2008). Le même groupe a également identifié 48 marqueurs spécifiques au cancer de l'ovaire à partir de 21 tissus cancéreux (Klinck et al, 2008). Sur les 41 événements reliés au cancer du sein, 13 se retrouvent également dans les tissus de cancer de l'ovaire, suggérant que certains événements seraient communs entre différents types de cancer. Pour ce qui est du cancer de la prostate, les auteurs ont conclu que plus de 200 gènes présentaient des modifications dans les patrons d'épissage alternatif entre les tissus sains et cancéreux, par analyse à l'aide de micro-puces de jonctions d'épissage (Lapuk et al, 2014).

Le facteur d'épissage SRSF1 (aussi nommé ASF/SF2) est un membre de la famille des régulateurs SR dont il a été mentionné précédemment. En 2007, un groupe a défini cette protéine comme étant un proto-oncogène tellement son rôle dans l'apparition de cancers est grand. SRSF1 a des conséquences variées au niveau du protéome cellulaire, puisqu'il va réguler l'épissage de nombreux transcrits. Entre autre, lorsque surexprimé, il va favoriser

l'expression de formes anti-apoptotiques de certains effecteurs cellulaires comme BIM et BINI. Aussi, SRSF1 favorise l'apparition de la forme pro-angiogénique du récepteur VEGF, lequel est impliqué dans la formation de nouveaux vaisseaux sanguins importants pour la tumeur (Karni et al, 2007; Anczuków et al, 2012). D'autres facteurs d'épissage, comme SRSF6, SRSF9 et hnRNP A2/B1 et H, auraient également des propriétés oncogéniques (Cohen-Eliav et al, 2013; Golan-Gerstl et al, 2011; Lefave et al, 2011). À l'inverse, certains facteurs agiraient principalement comme suppresseurs de tumeur (RBM5, RBM6 et RBM10) (Bechara et al, 2013). Étant donné leur large champ d'action, la surexpression de facteurs d'épissage oncogéniques ou l'inhibition de facteurs suppresseurs de tumeur peuvent avoir de grandes conséquences sur le phénotype de la cellule.

Dans plusieurs cancers, le récepteur CD44 est surexprimé. Cette surexpression est associée avec un stade avancé de la maladie, car CD44 est un acteur primordial dans l'invasion des cellules cancéreuses et aussi dans le maintien de la prolifération soutenue. Effectivement, le récepteur CD44 joue un rôle important dans l'invasion des cellules cancéreuses, car il induit le processus de transition épithélio-mésenchymateuse (EMT) lorsque le variant d'épissage se modifie. C'est ce processus qui confère aux cellules le pouvoir de se déplacer. Quant à la prolifération soutenue, l'épissage alternatif modifié de CD44 va mener à une forme du récepteur plus longue qui activera de manière constitutive la voie des MAPK par liaison avec les domaines cytoplasmiques des récepteurs à tyrosine kinases, menant à la prolifération des cellules (Gotley et al, 1996; Prochazka et al, 2014).

Aussi, plusieurs acteurs importants dans l'apparition du phénotype cancéreux sont régulés par épissage alternatif. En effet, les protéines APC, Ras, p53, Casp9 et Bcl-x ne sont que quelques exemples de protéines importantes dont un changement dans l'épissage alternatif peut mener à une protéine ayant un effet opposé à la forme d'épissage normalement prédominante dans une cellule saine (Neklason et al, 2004; Okumura et al, 2011). À titre d'exemple, dans une cellule saine, l'épissage alternatif favorise l'isoforme court de Bcl-x (Bcl-xs). Bcl-xs est connu pour ses propriétés pro-apoptotiques. Il va donc favoriser l'apoptose des cellules lorsque nécessaire. Par contre, une modification dans l'épissage alternatif de Bcl-x favorisera la forme longue, Bcl-xl, laquelle est anti-apoptotique et qui confère même un

pouvoir de chimio-résistance aux cellules (Adams et Cory, 2007; Akgul et al, 2004). De même, la caspase 9 produit deux isoformes avec des effets antagonistes, casp9a étant pro-apoptotique et casp9b anti-apoptotique. La modification de l'épissage de la caspase 9 est d'ailleurs fréquemment observée dans les cellules de cancers du poumon (Shultz et al, 2010). Ces quelques exemples démontrent comment un épissage alternatif modifié peut mener au cancer. Par contre, il existe plusieurs autres façons par lesquelles des changements au niveau de l'épissage alternatif peuvent transformer une cellule saine en cellule cancéreuse.

#### **1.4) HYPOTHÈSE ET OBJECTIF DE RECHERCHE**

**Est-ce que les virus de l'hépatite B et C peuvent modifier l'épissage alternatif de gènes cellulaires et de cette façon contribuer à causer le cancer du foie des suites de ces altérations?**

En vertu des évidences selon lesquelles les virus sont en mesure de transformer les cellules qu'ils infectent en cellules cancéreuses, et également que les modifications de l'épissage alternatif sont elles aussi une cause d'apparition de cancer, il est raisonnable de s'interroger sur l'implication des oncovirus dans les modifications des événements d'épissage alternatif.

##### ***Objectif***

**1-Déterminer si les virus de l'hépatite B et C sont capables de modifier l'épissage alternatif de certains gènes dans des tissus cancéreux.**

À l'aide des données de séquençage d'ARN à haut débit sur des tissus humains du foie, cancéreux ou non, fournies par le Cancer Genome Atlas, il est possible d'évaluer les changements au niveau de l'épissage alternatif selon la présence ou non de chacun des virus qui nous intéressent.

## ARTICLE

### 2.1) AVANT-PROPOS DE L'ARTICLE

Les carcinomes hépatocellulaires sont en majorité engendrés par une infection préalable dans la vie de l'individu par les virus de l'hépatite B ou C. L'implication de ces virus dans la carcinogénèse du foie n'est donc pas méconnue. Par contre, peu d'information est disponible quant à la manière que peuvent prendre ces virus pour modifier leur cellule hôte. Plusieurs effets ponctuels au niveau de la cellule ont été recensés, au sein de différents processus cellulaires ou encore de différentes voies de signalisation. Pour expliquer ces différents effets, il est raisonnable de penser que les virus peuvent agir sur des mécanismes de régulation beaucoup plus grands, qui auront un effet global sur l'ensemble des voies de la cellule hôte. Au cours des dernières années, il a été mis de l'avant que les modifications au niveau de l'épissage alternatif de la cellule pouvaient être responsables de la transformation de cette même cellule en cellule cancéreuse. Sachant que l'épissage alternatif est un mécanisme central de régulation de la diversité protéique d'une cellule, il est pertinent de s'intéresser à l'épissage alternatif dans le cadre d'une infection virale causant le cancer.

Les travaux de recherche qui constituent ce mémoire se sont penchés sur l'implication des virus de l'hépatite B et C dans le développement de carcinomes hépatocellulaires. En utilisant les données de séquençage d'ARN à haut débit du Cancer Genome Atlas sur des tissus hépatiques (sain et cancéreux), il nous a été possible d'étudier le rôle de ces virus dans l'épissage alternatif. Chaque tissu séquencé a pu être classifié selon qu'il était cancéreux ou non, et la présence de chacun de virus a pu être prise en compte afin de distinguer les cancers causés par le VHB, le VHC mais aussi les quelques tumeurs qui étaient positives pour les deux virus à la fois. Enfin, il nous a été possible d'observer ce qui se passe lorsque les tumeurs sont négatives pour la présence des virus, et donc analyser l'effet du cancer lui-même sur l'épissage alternatif. Tout ceci nous a permis de distinguer quelles modifications de l'épissage alternatif sont causées par le phénotype cancéreux en tant que tel, et quelles modifications sont engendrées par chacun des virus.

## 2.2) RÉSUMÉ DE L'ARTICLE ET IMPLICATION

L'épissage alternatif est un mécanisme central de régulation génétique qui augmente autant la variabilité et la diversité du protéome cellulaire en changeant la composition des protéines résultants en faisant un choix différentiel des exons à être inclus dans l'ARNm mature. L'épissage alternatif est un processus finement régulé et des dérégulations au niveau des patrons d'épissage ont été associées avec plusieurs pathologies humaines, incluant le cancer. Dans l'étude présentée ici, l'altération globale au niveau de l'épissage de l'ARN de gènes cellulaires a été investiguée à grande échelle sur 377 échantillons de tissus hépatiques en utilisant des données de séquençage d'ARN à haut débit. Ces données de séquençage sont tirées de patients âgés d'en moyenne 60 ans atteints ou non de cancer du foie à différents stades de la maladie et sont accessibles à partir du portail du Cancer Genome Atlas. L'alignement au génome humain des amplicons obtenus suite au séquençage a par la suite été fait afin d'identifier chacun des transcrits séquencés. Cette étude fournit le premier portrait des changements globaux au niveau des signatures d'épissage de l'ARN qui survient dans le carcinome hépatocellulaire. Notre étude identifie les modifications dans les patrons d'épissage alternatif de transcrits encodés par plus de 2500 gènes incluant des gènes suppresseurs de tumeur, des facteurs de transcription et des kinases. Les résultats démontrent des changements généralisés dans les carcinomes hépatocellulaires qui ont un impact au niveau du métabolisme cellulaire et de la prolifération de manière à contribuer à la carcinogénèse. Ces découvertes permettent également de comprendre les différences moléculaires entre des carcinomes hépatocellulaires associés au virus de l'hépatite B et au virus de l'hépatite C, et permettent l'identification de signatures uniques pour lesquelles l'épissage alternatif est dérégulé dans différents types de carcinomes hépatocellulaires. Cette identification de changements importants dans le profil d'épissage alternatif cellulaire dans les carcinomes hépatocellulaires représente le premier pas vers le développement d'agents anti-cancéreux basés sur les dérégulations d'épissage alternatif.

Mon implication au niveau de la réalisation de cet article est majoritaire. En effet, je suis celle qui ait analysé les résultats de séquençage autant au niveau de l'épissage que de l'expression des gènes. En effectuant l'analyse, j'ai également généré l'ensemble des figures, à l'exception de l'une d'elle qui a été généré par Andréa Allaire, une des co-auteurs de ce

manuscrit. Avec l'aide de Victoria E. S. Armero, une autre co-auteure, j'ai imaginé et élucidé comment bien présenter les figures qui composent cet article. Par contre, la rédaction du manuscrit a été réalisée par mon directeur de recherche, le Pr Martin Bisaillon. Les membres de la plateforme génomique de l'Université de Sherbrooke, Mathieu Durand, Philippe Thibault et Elvy Lapointe sont ceux qui ont effectué l'alignement des séquences sur le génome humain, la génération des heatmap, et la validation des candidats par PCR. Simon Boudreault, Maude Tremblay-Létourneau, Michelle S. Scott et Jean-Pierre Perreault ont contribué à l'élaboration du plan expérimental, et ont participé à la révision de l'article final.



## 2.3) ARTICLE

# **Global profiling of alternative RNA splicing events misregulated in hepatocellular carcinoma**

**Auteurs de l'article: Marie-Pier Tremblay<sup>1</sup>, Victoria E. S. Armero<sup>1</sup>, Andréa Allaire<sup>1</sup>, Simon Boudreault<sup>1</sup>, Mathieu Durand<sup>2</sup>, Elvy Lapointe<sup>2</sup>, Philippe Thibault<sup>2</sup>, Maude Tremblay-Létourneau<sup>1</sup>, Jean-Pierre Perreault<sup>1</sup>, Michelle Scott<sup>1</sup>, Martin Bisailon<sup>1,\*</sup>**

**Statut de l'article:** Soumis à Hepatology le 18 novembre 2015

<sup>1</sup> Département de biochimie, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Quebec, Canada J1E 4K8.

<sup>2</sup> Plateforme Rnominique, Université de Sherbrooke, Sherbrooke, Quebec, Canada J1E 4K8.

\*Contact information:

Martin Bisailon

Département de biochimie

Pavillon de recherche appliquée sur le cancer

Faculté de médecine et des sciences de la santé

Université de Sherbrooke

3201 Jean-Mignault

Sherbrooke, Qc J1E 4K8 Canada

[Martin.Bisailon@USherbrooke.ca](mailto:Martin.Bisailon@USherbrooke.ca)

Keywords: Liver cancer, Hepatitis B virus, Hepatitis C virus, RNA, Gene dysregulation.

## **ABSTRACT**

Alternative splicing (AS) is a central mechanism of genetic regulation which increases both the variability and diversity of the cellular proteome by changing the composition of resulting proteins through differential choice of exons to be included in mature mRNAs. AS is a tightly process and dysregulations in splicing patterns have been associated with many human diseases including cancer. In the present study, alterations to the global RNA splicing landscape of cellular genes were investigated in a large-scale screen from 377 liver tissue samples using high-throughput RNA sequencing data. This study provides the first comprehensive portrait of global changes in the RNA splicing signatures that occur in hepatocellular carcinoma (HCC). Our study identifies modifications in the AS patterns of transcripts encoded by more than 2500 genes such as tumor suppressor genes, transcription factors, and kinases. The results reveal widespread AS changes in HCC that impact cell metabolism and proliferation in a manner that likely contributes to tumorigenesis. These findings also provide insights into the molecular differences between Hepatitis B virus- and Hepatitis C virus-associated HCCs, and allowed the identification of unique gene signatures for which AS is misregulated in different types of HCC. This identification of extensive changes in the cellular AS landscape in HCC likely represents a first step toward the development of anticancerous agents based on AS dysregulation.

## **INTRODUCTION**

Hepatocellular carcinoma (HCC) is the fifth most prevalent tumor type and the third leading cause of cancer-related deaths worldwide.<sup>1</sup> Up to 80% of HCC cases occur in developing countries, although there has been an increasing trend over the past two decades in the incidence of HCC in developed western countries.<sup>2</sup> Chronic infection with hepatitis B virus (HBV) and hepatitis C virus (HCV) are the leading causes of HCC worldwide.<sup>3</sup> These two hepatotropic viruses represent the most important risk factors for the development of HCC, with an estimated 80% of all HCC cases globally.<sup>4</sup> Other less common risk factors include various nonviral factors, such as alcohol consumption and other chronic liver diseases.<sup>5</sup>

Hepatocarcinogenesis is a complex multistep process which involves genetic and epigenetic alterations, the activation of cellular oncogenes and/or the inactivation of tumor suppressor genes, and dysregulation of various signal transduction pathways.<sup>6</sup> HBV and HCV induce carcinogenesis via different molecular mechanisms. HBV infection causes HCC via both indirect and direct pathways.<sup>7</sup> HBV infection promotes chronic injury to the hepatocytes, with continuous necro-inflammation and regeneration activity.<sup>8</sup> As a consequence, an increase in hepatocyte turnover is observed which leads to the accumulation of potential critical mutations with subsequent malignant transformation and clonal expansion, leading to HCC.<sup>9</sup> Moreover, HBV infection can directly lead to HCC by integrating the viral genomic DNA into the host genome.<sup>10</sup> HBV integration can lead to numerous mutagenic consequences, including large inverted duplications, deletions, amplifications and translocation, resulting in chromosomal instability.<sup>11</sup> In contrast, HCV does not integrate into the hepatocyte genome. The main hypothesis for HCV carcinogenesis is that it occurs via indirect pathways through chronic inflammation and hepatocellular injury. This is supported by the fact that cirrhosis is frequently a pre-requisite for HCV-induced HCC.<sup>7</sup> However, HCC has been reported to develop in a very small proportion of non-cirrhotic patients, suggesting a more direct effect from the HCV itself.<sup>12-13</sup> Accordingly, various HCV proteins have been reported to have a role in HCC development in experimental models.<sup>14</sup>

During the last decade high-throughput molecular profiling studies have investigated changes in gene and protein expression that are found in HCC. These studies have shown that HBV- and HCV-associated HCC have very distinct differential patterns of gene expression. For instance, HBV-associated HCC is typically enriched with up-regulated genes related to signal transduction, transcription and metastasis, and in pathways related to immune response, cell cycle, and RNA metabolism.<sup>15</sup> In contrast, HCV-associated HCC is usually associated with an up-regulation of genes related to immune response and detoxification function of liver, and in both cancer-related and lipid metabolism pathways.<sup>15</sup> Interestingly, TGF- $\beta$  pathway, MAPK and p53 pathways are enriched in both HBV- and HCV-associated HCC.<sup>16</sup>

Alterations in RNA splicing of a small number of cellular genes have also been observed in

HCC, and although still limited, the available data suggest that splicing defects may play a role in hepatocarcinogenesis.<sup>17</sup> Higher eukaryotes use alternative splicing (AS) to change the composition of transcripts encoded by a gene through selection of different exons to be included in the mature mRNA, thus producing a variability at the proteomic level.<sup>18-20</sup> AS therefore diversifies the cellular proteome because it results in multiple transcripts and, as a consequence, multiple proteins are created from a single gene. These different proteins produced from the same gene can frequently promote different or even opposite biological effects. As a result, the relative abundance of each isoform may have consequences on cellular functions. An example of this is clearly demonstrated by the BCL-x gene. AS of the Bcl-x transcript results in two different mature mRNAs: Bcl-x(L), which has anti-apoptotic effects, and Bcl-x(S) which promotes apoptosis.<sup>21</sup> Therefore, it is not surprising that AS is tightly regulated and that variations in splicing patterns have been associated with many human diseases including cancer.<sup>22-25</sup> During the last few years, cancer-specific splice variants and cancer-associated changes in the relative levels of spliced isoforms of genes with an established role in carcinogenesis have been observed.<sup>26</sup> AS alterations can confer selective advantages to tumors, such as angiogenesis, proliferation, cell invasion and avoidance of apoptosis.<sup>27-31</sup> Moreover, in some cases the mechanisms leading to the appearance of these tumor-associated alterations in mRNA splicing are starting to emerge.<sup>32-33</sup> Recent evidences indicate that some of these can be used as prognostic or diagnostic biomarkers, and the identification of molecules capable of correcting and/or inhibiting pathological splicing events is also an important issue for future therapeutic approaches.<sup>34-39</sup>

Various studies have shown alterations in the AS patterns of a limited number of cellular genes in HCC.<sup>17</sup> One interesting case of dysregulated AS occurring in HCC is the overexpression of a splice variant of DNA methyltransferase 3b (DNMT3b), namely DNMT3b4.<sup>40</sup> An elevation of the ratio of DNMT3b4 to DNMT3b3 mRNA is correlated with DNA hypomethylation on pericentromeric satellite regions which may induce chromosomal instability, and is considered an early event during hepatocarcinogenesis.<sup>40</sup> Other examples of aberrantly-spliced genes detected in HCC include the serine/threonine kinase aurora kinase B (AURKB), the E3 ubiquitin ligase, the p53-antagonistic protein MDM2, the cell surface adhesion cadherin 17 protein, and the Hugel-1, Klf6, and p73 tumor suppressors.<sup>17,41</sup>

These examples illustrate the oncogenic potential of aberrantly spliced isoforms in HCC tissues.

In the present study, alterations to the global cellular AS landscape of more than 377 liver tissue samples from The Cancer Genome Atlas (TCGA) were investigated using high-throughput RNA sequencing data. This study provides the first comprehensive portrait of global changes in the RNA splicing signatures that occur in HCC. We identify modifications in the AS patterns of transcripts encoded by more than 2500 genes such as tumor suppressor genes, transcription factors, splicing factors, and kinases. These findings also provide crucial insights into the molecular differences between HBV- and HCV-associated HCCs, and highlight the presence of another important level of misregulation during gene expression in HCC. The results also reveal widespread AS changes in HCC that impact cell metabolism and proliferation in a manner that likely contributes to tumorigenesis. Finally, the data allowed the identification of unique gene signatures for which AS is misregulated in different types of HCC.

## **EXPERIMENTAL PROCEDURES**

### **RNA-seq Data Analysis**

Sequence reads were aligned on a transcriptome reference sequence database (UCSCGene Hg19) using Bowtie v2 aligner (default parameters) and associated gene isoforms quantified in transcript-per-million (TPM) using RSEM for each sequenced sample.<sup>42-43</sup> Alternative splicing events were automatically detected and quantified using the percent-spliced-in (PSI,  $\Psi$ ) metric based on isoform expression TPM for the long (L) and the short (S) forms using equation below:

$$\Psi = \frac{L}{L + S}$$

### **Gene Expression Analysis**

The gene list was initially filtered to keep only data present in at least two replicates for both virus-associated HCC and normal hepatic tissue. To ensure higher reproducibility, only genes with expression levels higher than one transcript per million (TPM) in either dataset were

conserved. Fold changes in base 2 logarithm were then calculated between both virus-associated HCC and normal hepatic tissue average TPM. Q-values were calculated in order to take into account multiple statistical hypothesis testing and results under 0.05 were considered significant.

### **Alternative Splicing Analysis**

The alternative splicing event (ASE) list was filtered to keep only data present in at least two replicates for both virus-associated HCC and normal hepatic tissue. Events with a P-value less than 0.05 were conserved. To ensure higher stringency, the ASEs were further filtered with a cutoff Q-value of less than 0.05. From these events, only those with a difference higher than 10% in PSI were considered biologically relevant.

### **Statistical Analysis**

Welch's t-test (Student's t-test with unequal sample sizes and unequal variances) was calculated through the GSL library (<http://www.gnu.org/software/gsl/>) integrated to Perl system analysis for gene expression and alternative splicing data. Also, false discovery rates were calculated with the Q-value package in R (<https://cran.r-project.org/src/contrib/Archive/qvalue/>) based on Storey and Tibshirani.<sup>44</sup> For all other analyses, Graph Pad Prism version 6.05 was used to run statistical analysis.

### **Gene Ontology Analysis**

Gene ontology analysis with PANTHER was performed using the database for annotation, visualization, and integrated discovery (DAVID).<sup>45</sup>

### **String Networks**

Using the STRING database (version 10)<sup>46</sup>, genes were submitted for generation of protein-protein interaction network from the *Homo sapiens* interactome. High resolution evidence views were created and saved.

### **Functional ASE Prediction**

Using the FAST-BD or EASANA suite, the splicing patterns of genes of interest were visualized. DNA sequences of representative transcripts presenting short and long isoforms were downloaded and translated into proteins using ExPASy translation tool.<sup>47</sup> Predicted proteins were then compared using Multalin (truncation and frameshift event)<sup>48</sup>, PFAM (loss or appearance of functional domain)<sup>49</sup>, and NLS Mapper (loss or gain of nuclear localisation signal).<sup>50</sup>

### **PCR Validation**

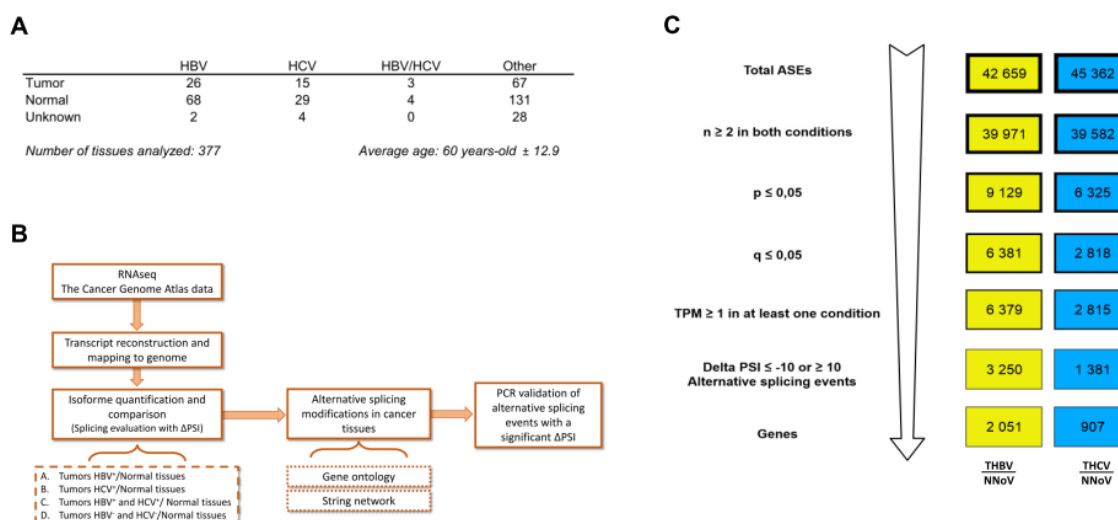
Liver Cancer cDNA Arrays from Origene TissueScan plates (Rockville, MD) were assessed for the expression of the various transcripts using the manufacturer's protocol. The plates contained cDNAs from 8 normal and 39 liver cancer tissues. All forward and reverse primers were individually resuspended to 20–100  $\mu$ M stock solution in Tris-EDTA buffer (IDT) and diluted as a primer pair to 1  $\mu$ M in RNase DNase-free water (IDT). PCR reactions were performed in 10  $\mu$ l in 96 well plates on a CFX-96 thermocycler (BioRad). The following cycling conditions were used: 3 min at 95°C; 50 cycles: 15 sec at 95°C, 30 sec at 60°C, 30 sec at 72°C. For every PCR run, control reactions performed in the absence of template were performed for each primer pair and these were consistently negative. The amplified products were analyzed by automated chip-based microcapillary electrophoresis on Caliper LC-90 instruments (Caliper LifeSciences). Amplicon sizing and relative quantitation was performed by the manufacturer's software.

## **RESULTS**

### **Modification of the cellular AS landscape in HCC**

Alterations to the global RNA splicing landscape of more than 377 liver tissue samples (Figure 1A) from The Cancer Genome Atlas (TCGA) were investigated using high-throughput RNA sequencing data. The overview of all the analyses performed in this study is outlined in Figure 1B. RNA-seq data were collected and mapped to the reference genome, followed by transcript assembly and analysis of RNA isoform abundance. We initially focused our study on HBV-associated HCC (THBV, Tumors with HBV) and HCV-associated HCC (THCV, Tumors with HCV), and we compared the global splicing patterns

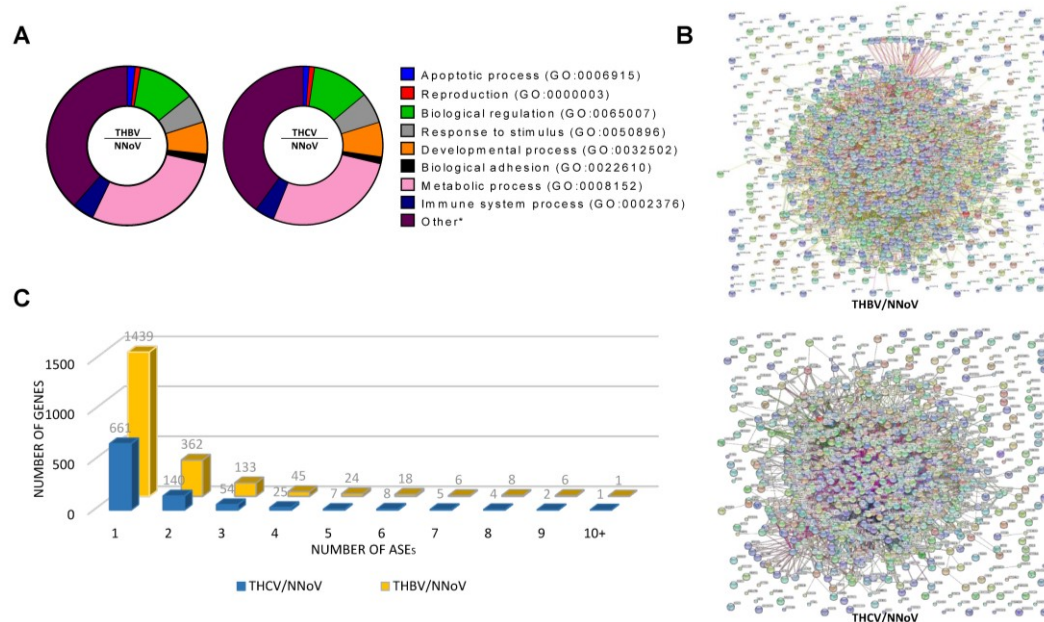
with normal tissues (NNoV, Normal, no virus). In order to identify the cellular AS patterns that are altered in HCC, we evaluated the changes in the relative splice abundances by quantifying all the primary transcripts that produce two or more isoforms. Alternative splicing events (ASEs) were detected and quantified using the percent-spliced-in (PSI) metric based on isoform expression TPM for the long and the short isoforms (see Experimental Procedures). The lists of ASEs (42,000 - 45,000 ASEs) were filtered to keep only data with at least two replicates for both types of HCC and normal hepatic tissue. The ASEs that differed between HCC tissues and normal hepatic tissue with a P-value of less than 0.05 were conserved. To ensure higher stringency, the ASEs were further filtered with a cutoff Q-value (false-discovery rate) of less than 0.05. From these events, only those with a difference higher than 10% in  $|\Delta\text{PSI}|$  were considered biologically relevant. Using such an approach, we identified 3250 primary transcripts belonging to 2051 genes for which the AS pattern was significantly modified in patients with HBV-associated HCC ( $Q \leq 0.05$ ,  $|\Delta\text{PSI}| \geq 10$ ) (Figure 1C). Moreover, the AS patterns of 1381 primary transcripts belonging to 908 genes were significantly modified in patients with HCV-associated HCC. The complete list of the differential ASEs associated with HBV- and HCV-associated HCC is provided in Supplemental Materials. The list includes six of the previously identified aberrantly spliced isoforms found in HCC in previous studies (*DNMT3b*, *AURKB*, *UBE3B*, *MDM2*, *KLF6*, and *TP73*). RNA-Seq data were not available for the two other previously identified examples of aberrantly-spliced genes detected in HCC (*CDH17* and *HUGLI*).





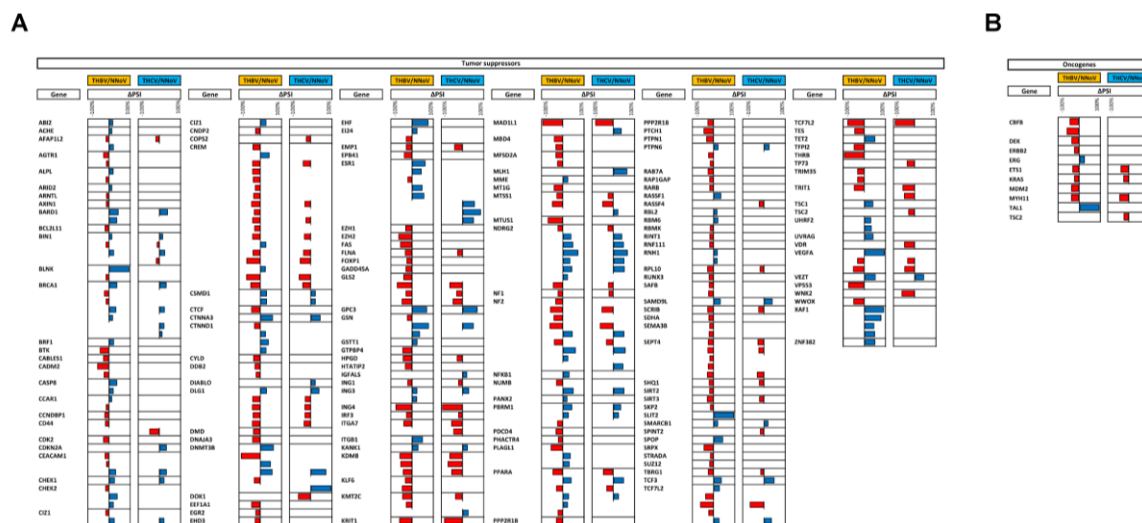


**Figure 2. Global profiling of alternative splicing event modifications in HBV- and HCV-associated HCC.** (A) Alternative splicing event modifications in HBV-associated HCC (THBV/NNoV) and HCV-associated HCC (THCV/NNoV) are presented with their associated changes in gene expression. ASEs were detected and quantified using the percent-spliced-in (PSI) values based on isoform expression for the long and the short isoforms (see Experimental Procedures). The graphs represent the relation between the difference of splicing for each hepatitis-associated HCC compared with healthy non-viral tissues, and the difference in gene expression for each of these alternative splicing event modifications. Only ASEs with a significant modification (PSI value under -10% or over 10%, with a Q value under 0.05) are represented in graphs. (B) Heatmap representations of isoform ratios (PSI values) for all tissues analyzed in the current study. The first heat map shows PSI values for HBV-associated HCC. HBV-associated HCC tissues (THBV) are shown in red, and the comparative healthy tissues are shown in green (NNoV). The second heat map shows PSI values for HCV-associated HCC. HCV-associated HCC tissues (THCV) are shown in blue, and the comparative healthy tissues are shown in green (NNoV).



**Figure 3. Characterization of ASEs that are modified in HBV- and HCV-associated HCC.** (A) Gene ontology analysis for alternative splicing modifications in HBV- and HCV-associated HCC. The asterisk (\*) indicates the other categories: Cellular process (GO:0009987), cellular component organization or biogenesis (GO:0071840), localization (GO:0051179), locomotion (GO:0040011) and multicellular organismal process (GO:0032501). (B) Interaction network for cellular genes for which AS is modified in both HBV- and HCC-associated HCC. The network was determined by uploading the genes into STRING. (C) Distribution of multiple alternative splicing events (ASEs) in each gene. Some of the genes which have changes in their splicing process present these modifications at the level of more than one splicing event per gene. HBV-associated HCC (THBV/NNoV) is represented in yellow while HCV-associated HCC (THCV/NNoV) is shown in blue color.

number of the modified ASEs are encoded by genes with important roles in metabolism. For instance, alterations in the splicing patterns of BCAT2, which encodes a branched chain aminotransferase found in mitochondria, DAP3, which helps in protein synthesis within the mitochondrion, and ECHDC2, which is involved in fatty acid biosynthesis, were found to be present both in HBV- and HCV-associated HCC. Many of these proteins form a major interacting cluster and interact either directly or indirectly with each other (Figure 3B). Through manual curation of functional annotations, we also identified numerous modifications in the AS patterns of genes potentially involved in oncogenic transformation. We identified many transcription factors, receptors, tumor suppressors, and kinases for which AS is significantly modified in HBV- and HCV-associated HCC (Table 1). For example, HBV-associated HCC displayed alterations in the AS patterns of 138 tumor suppressors, 156 transcription factors, 73 kinases, and 132 receptors. Some of the dysregulated ASEs that were identified include KRAS, a member of the small GTPase superfamily implicated in various malignancies (including lung adenocarcinoma, mucinous adenoma, ductal carcinoma of the pancreas and colorectal carcinoma), MDM2, a nuclear-localized E3 ubiquitin ligase which can promote tumor formation by targeting tumor suppressor proteins (such as p53) for proteasomal degradation, and BRCA1 which encodes a nuclear phosphoprotein that plays a role in maintaining genomic stability, also acts as a tumor suppressor and is frequently mutated in breast and ovarian cancers. Interestingly, some of these modifications are found both in HBV- and HCV-associated HCC while others are unique to the respective virus-associated HCC (Figure 4A-B and S2A-B). In the case of common misregulated ASEs, the direction of the splicing shift between normal and tumor tissues was frequently the same between HBV- and HCV-associated HCC (Figure 4A-B).



**Figure 4. AS modifications in transcripts encoded by tumor suppressors and oncogenes in HBV- and HCV-associated HCC. (A)** Alterations in the AS patterns of transcripts encoded by tumor suppressors in HBV- and HCV-associated HCC. Some transcripts have multiple ASEs that are modified. Red bars indicate negative Delta PSI values, and blue bars represent positive Delta PSI values. **(B)** Alterations in the AS patterns of transcripts encoded by oncogenes in HBV- and HCV-associated HCC.

**Table 1 –Protein families encoded by transcripts that are differentially spliced in HBV- and HCV-associated HCC**

Protein classification	THBV/NNoV		THCV/NNoV	
	Frequency	Percentage	Frequency	Percentage
Extracellular matrix protein (PC00102)	29	1.5%	12	1.4%
Protease (PC00190)	50	2.5%	18	2.0%
Cytoskeletal protein (PC00085)	100	5.1%	53	6.0%
Transporter (PC00227)	105	5.3%	56	6.3%
Transmembrane receptor regulatory/adaptor protein (PC00226)	4	0.2%	3	0.3%
Transferase (PC00220)	159	8.0%	64	7.2%
Oxidoreductase (PC00176)	68	3.4%	27	3.1%
Lyase (PC00144)	27	1.4%	11	1.2%
Cell adhesion molecule (PC00069)	52	2.6%	20	2.3%
Ligase (PC00142)	54	2.7%	32	3.6%
Nucleic acid binding (PC00171)	225	11.4%	114	12.9%
Signaling molecule (PC00207)	101	5.1%	39	4.4%
Enzyme modulator (PC00095)	147	7.4%	74	8.4%
Calcium-binding protein (PC00060)	40	2.0%	19	2.2%
Defense/immunity protein (PC00090)	63	3.2%	24	2.7%
Phosphatase (PC00181)	45	2.3%	17	1.9%
Hydrolase (PC00121)	187	9.5%	77	8.7%
Transfer/carrier protein (PC00219)	39	2.0%	22	2.5%
Membrane traffic protein (PC00150)	54	2.7%	22	2.5%
Transcription factor (PC00218)	156	7.9%	67	7.6%
Chaperone (PC00072)	15	0.8%	2	0.2%
Cell junction protein (PC00070)	10	0.5%	9	1.0%
Surfactant (PC00212)	6	0.3%	4	0.5%
Structural protein (PC00211)	9	0.5%	5	0.6%
Kinase (PC00137)	73	3.7%	32	3.6%
Storage protein (PC00210)	5	0.3%	1	0.1%
Receptor (PC00197)	132	6.7%	53	6.0%
Isomerase (PC00135)	20	1.0%	5	0.6%
Splicing factor	23	1.2%	13	1.6%
Tumor suppressor	138	6.7%	63	3.0%
Oncogene	9	0.5%	4	0.5%

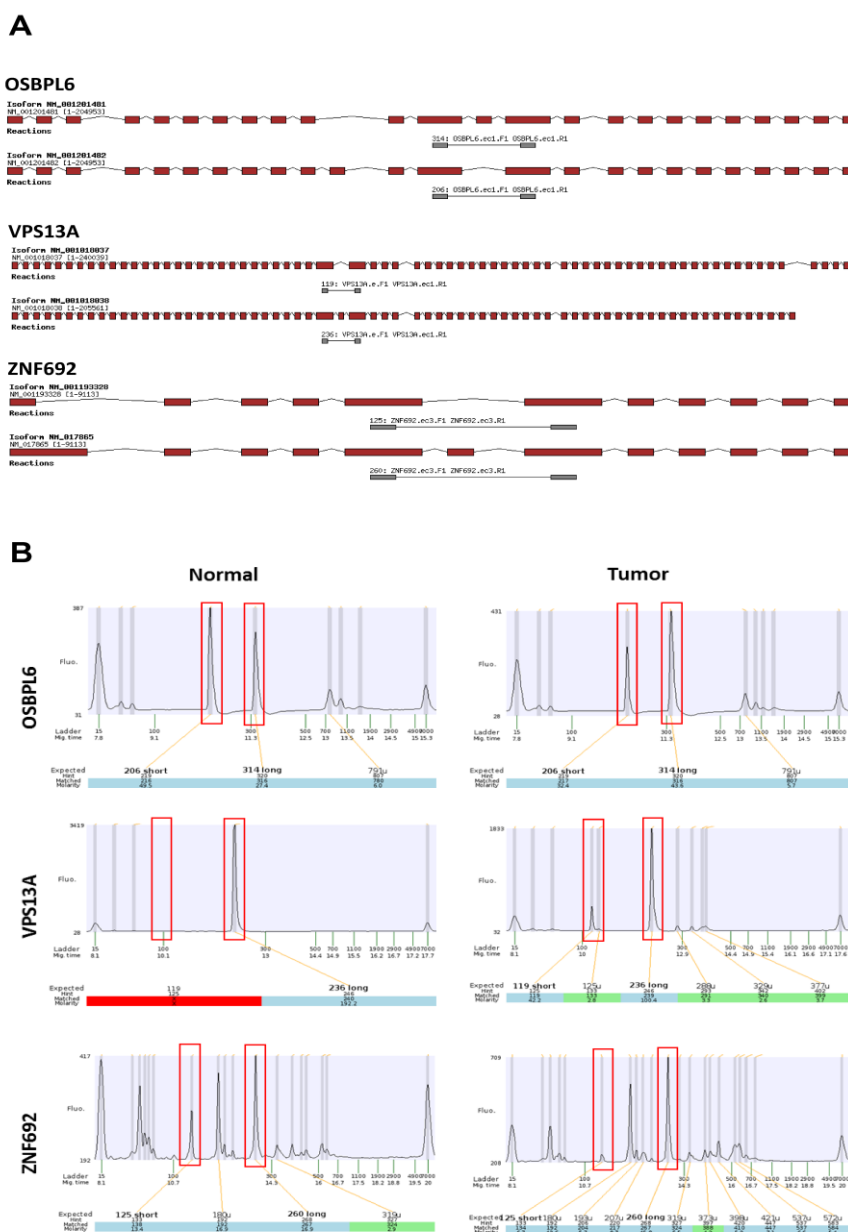
### **Characterization of the ASEs that are modified in HCC**

We next compared the profiles of the ASEs that are modified in HBV- and HCV-associated HCC. The vast majority of the ASEs that are modified in HCC were found at a level of one splicing event per gene (Figure 3C and S1B). Interestingly, analysis of the correlation between AS and gene expression indicated stable expression of most of the transcripts that are differentially spliced (Figure 2A). Notable exceptions included AURKB, one of three highly homologous serine/threonine kinases that play a critical role in regulating many of the processes that are pivotal to mitosis, CDC45 which is important for early steps of DNA replication, CDK1, which plays a key role in the control of the eukaryotic cell cycle by modulating the centrosome cycle as well as mitotic onset, and ALDH3A1, which is involved in the metabolism of corticosteroids, biogenic amines, neurotransmitters, and lipid peroxidation. Overall, no clear association between the cancer-specific splicing pattern and expression level was observed, indicating that there is no strict link between the regulation of transcription and of splicing in these cancer-specific genes.

Among the transcripts for which AS was significantly affected in HBV- and HCV-associated HCC, many ASEs we documented affect known protein domains. Table S1 displays the consequences of the differentially spliced transcripts on protein function for 60 transcripts that are differentially spliced in HCC. Among the differential ASEs, at least two ASEs resulted in the loss of predicted nuclear localization signals (NLS). These are RINT1, which plays a role in cell cycle checkpoint control and is essential for telomere length control, and EHD3, a paralog of EHD1, a protein which is thought to play a role in the endocytosis of IGF1 receptors. Twenty other ASEs were predicted to disrupt known functionally critical protein domains. Most of these genes encode for proteins involved in metabolic processes. For instance, some of the dysregulated ASEs lead to protein domain loss in HADHA which encodes the alpha subunit of the mitochondrial trifunctional protein and catalyzes the last three steps of mitochondrial beta-oxidation of long chain fatty acids, and ALDH2, the second enzyme of the major oxidative pathway of alcohol metabolism.

### **Experimental validation on HCC tissues**

We next sought to experimentally validate the results that were obtained from the RNA-Seq data. We used PCR analysis in order to experimentally validate the differential ASEs that were observed through RNA-seq studies. Using liver cDNA arrays from both healthy and cancerous tissues, we designed specific primers to allow detection of the various ASEs by PCR. Examples of differentially spliced transcripts are presented in [Figure 5A](#) illustrating the modifications in isoform usage in transcripts encoded by three different genes (*OSBPL6*, *VPS13A*, *ZNF692*). Our results demonstrated that changes in AS levels revealed through transcriptome sequencing could also be detected by PCR analysis ([Figure 5B](#)).



**Figure 5. Validation of ASEs dysregulated in HCC. (A)** Overview of the two isoforms encoded by *OSBPL6*, *VPS13A* and *ZNF692* genes. Exons are depicted in red and the intervening introns are shown as thin black lines (not to scale). The primers used to detect the isoforms by RT-PCR assays are shown in gray and the sizes of the expected amplicons are also indicated. **(B)** cDNA obtained from tissues were analyzed by PCR using specific primers to detect both isoforms of the transcripts encoded by the *OSBPL6*, *VPS13A* and *ZNF692* genes. The amplified products were analyzed by automated chip-based microcapillary electrophoresis. Capillary electropherograms of the PCR reactions are shown. The positions and the amplitude of the detected amplicons are highlighted by red boxes. The positions of the internal markers are also indicated. The data show the increase in the relative abundance of the long isoform (314nt) and the diminution in the abundance of the short

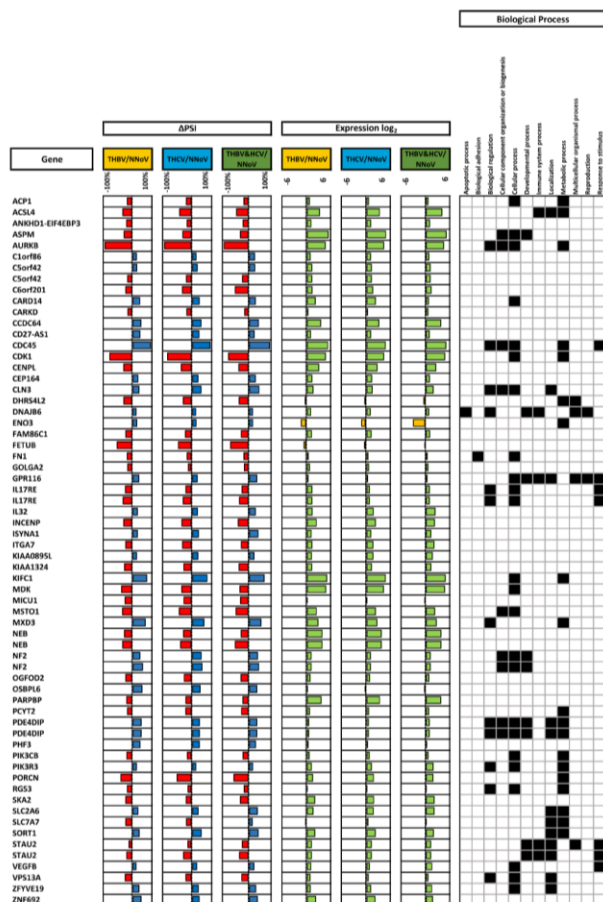


isoform (206nt) for the *OSBPL6* gene. For the *VPS13A* gene, the data shows the gain in abundance of the short isoform (119nt). The figure also presents the decrease in the abundance of the short isoform (125nt) for *ZNF692* gene.

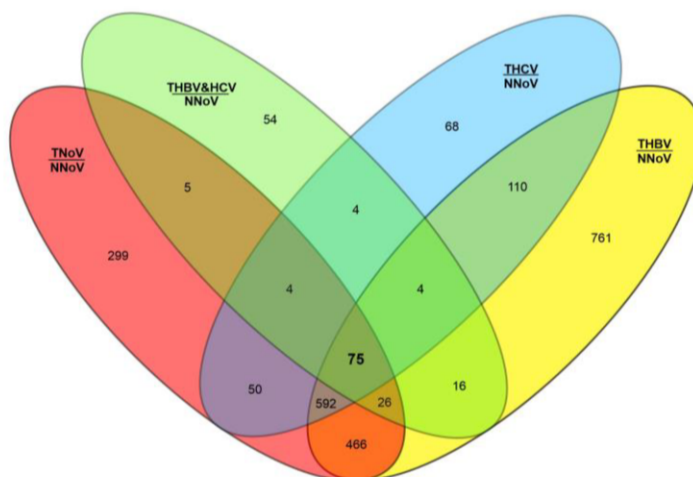
### **Comparison between different types of HCC**

The large size of the samples analyzed by high-throughput RNA sequencing allowed us to obtain data for the alterations to the global RNA splicing landscape of two other types of HCC, namely HCC associated with the co-presence of both HBV and HCV (THBV&HCV) and virus-free HCC (TNoV, Tumor no virus). Because of their shared modes of transmission, co-infection of HBV and HCV is not uncommon, particularly in countries with a high prevalence of HBV or HCV.<sup>51</sup> HBV and HCV co-infection results in more severe liver disease and in an increased risk of HCC than mono-infection.<sup>51</sup> Although the total number of HCC tissues associated with the presence of both HBV/HCV was relatively small in our total samples collection (n=3), our statistical analysis allowed us to identify 220 transcripts for which AS is modified in comparison to normal tissues (Figure S1C, S3A, Table S2-S6). Remarkably, 79 genes for which AS was modified were found to be common between the three types of virus-associated HCC (Figure 6A). This raised the possibility that these common ASEs might be related to a general characteristic of cancerous tissues. We therefore analyzed the cellular AS patterns that are altered in virus-free HCC of 131 patients (TNoV; tumor, no virus), and we evaluated the changes in relative splice abundances. Using such an approach, we identified 1517 genes for which the AS pattern was significantly modified in virus-free HCC ( $Q \leq 0.05$ ,  $|\Delta \text{PSI}| \geq 10$ ) (Figure S1C, S3B, Table S2-S6). Further analysis revealed that 75 of the 79 previously identified genes for which AS was modified in the three types of virus-associated HCC were also identified in virus-free HCC, thereby indicating that the modifications in AS found in those genes are associated with carcinogenic tissues. It should be noted that the AS of the four remaining genes were also found to be modified in virus-free HCC although they were not retained following the powerful statistical analysis i.e. the Q-values were slightly higher than 0.05 and/or the  $|\Delta \text{PSI}|$  were slightly lower than 10. Overall, our analysis allowed us to identify 761 unique genes for which AS is misregulated in HBV-associated HCC, while 68 are unique to HCV-associated HCC, 54 to HBV&HCV-associated HCC, and 299 to virus-free HCC (Figure 6B, Table S7-S11).

A



B

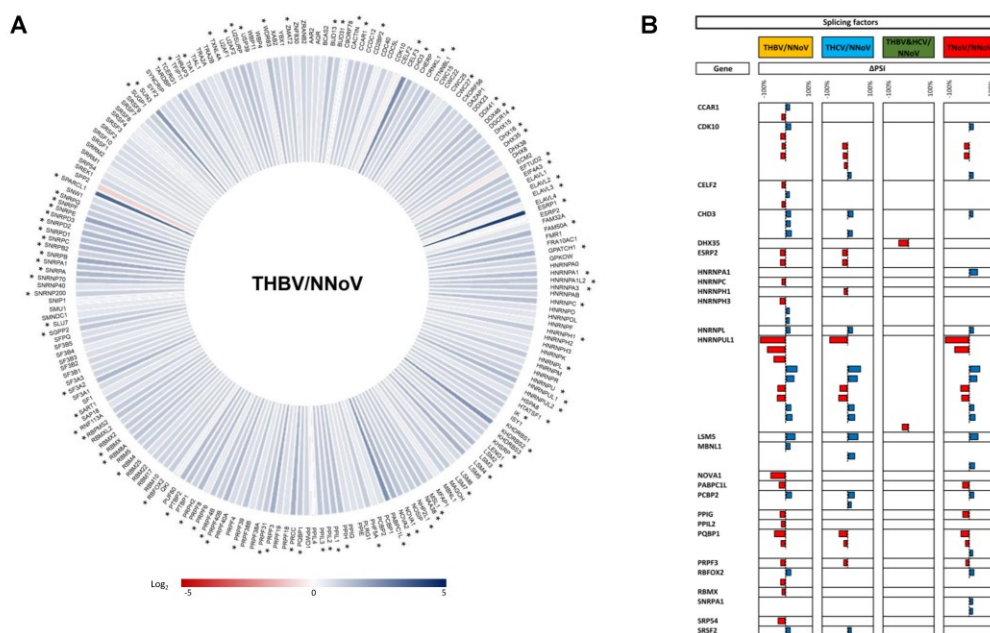


**Figure 6. Comparison between HBV-, HCV-, HBV&HCV-, and non-viral-associated HCC. (A)** The list displays the common differentially spliced transcripts for the various types of HCC with the corresponding Delta PSI values, the associated gene expression (in Log<sub>2</sub>), and the associated biological processes. **(B)** Comparison of the genes with dysregulated ASEs between HBV-, HCV-, HBV&HCV-, and virus-free HCC. 75 genes for which AS was

modified were found to be common between HBV-, HCV-, HBV&HCV-, and virus-free associated HCC.

### **RNA Splicing Factors**

The precise mechanism which leads to a modification of the AS landscape in HCC is unknown at the moment. Changes in splice site choice generally arise from modifications in the assembly of the spliceosome or by altering the binding of splicing factors to the RNA transcripts.<sup>52</sup> Although splicing is regulated by an abundant and yet incompletely characterized set of splicing factors, dysregulated expression of individual splicing factors has been shown to frequently result in aberrant splicing.<sup>53</sup> Taking these findings in consideration, we thus monitored both the expression levels and the changes in splicing patterns of RNA encoding for splicing factors and spliceosomal proteins. As shown in [Figure 7A](#), the expression of a number of splicing factors and spliceosomal proteins is indeed affected in HCC. The expression of 94 proteins involved in splicing was modulated by more than 2-fold in HBV-associated HCC. Interestingly, the expression of ESRP1, a splicing factor known to regulate diverse types of alternative splicing events<sup>54</sup>, was increased by more than 21-fold in HBV-associated HCC. Very similar expression profiles were also found in HCV-, HBV&HCV-, and virus-free associated HCC ([Figure S4](#)). Modifications to the AS patterns of transcripts encoding proteins involved in splicing (splicing factors and proteins of the spliceosome) was also investigated. Our study identified many splicing factors that were differentially spliced upon in HCC ([Figure 7B](#)). In the case of HBV-associated HCC, the AS patterns of 26 splicing factors were significantly modified. One example is RBFOX2, which encodes an RNA binding protein that is thought to be a key regulator of alternative exon splicing. Other examples include CELF2, which regulates pre-mRNA alternative splicing and may also be involved in mRNA editing, and translation, and MBNL1, which mediates pre-mRNA alternative splicing regulation by acting either as activator or repressor of splicing on specific pre-mRNA targets. Modifications to both the expression level and AS of splicing factors likely contribute to the observed modifications in the cellular AS landscape of HBV-, HCV, HBV&HCV- and virus-free associated HCC.



**Figure 7. RNA splicing factors in HCC. (A)** Iris Graph representing the expression profile of splicing factors for HBV-associated HCC. Differences in gene expression levels are shown on a logarithmic color scale ( $\text{Log}_2$ ), from red (negative changes in expression) to blue (increase in gene expression). The expression of 94 proteins involved in splicing was modulated by more than 2-fold (indicated by an asterisk). **(B)** Misregulation of splicing factors alternative splicing in HBV- (THBV/NNov, yellow), HCV- (THCV/NNov, blue), HBV&HCV- (THBV&HCV/NNov, green) and virus-free HCC (TNov/NNov, red). The delta PSI values are represented in red (negative delta PSI values) and in blue (positive delta PSI values).

## DISCUSSION

In the current study, we present the first large-scale screen of HCC-associated ASEs. The study demonstrated that the AS patterns of numerous cellular transcripts are modified in HCC. These data allowed us to identify unique signatures of genes for which AS is misregulated in the different types of HCC. Analysis of the AS landscape of the various types of HCC revealed numerous HCC-specific markers, which significantly increases the number of biomarkers that can currently be identified by standard expression profiling alone.

Currently,  $\alpha$ -fetoprotein (AFP) together with iconography and pathology detection are commonly used in clinical early diagnosis for liver cancer.<sup>55</sup> However, the widely used marker AFP does not always yield satisfactory results in the early diagnosis of HCC, particularly in the case of AFP-negative HCC, thereby limiting the universality of its application.<sup>55</sup> The use of other HCC markers (such as GPC3 and TGF- $\beta$ 1) is currently being explored<sup>56-57</sup>, which is expected to improve early diagnostic rate. Clearly, future developments in molecular genetics and proteomic analysis should lead to the identification of other useful HCC-specific markers. The current identification of unique signatures for genes in which AS is misregulated in the different types of HCC clearly constitutes a step toward that goal.

Tumor cells, in order to meet the increased requirements of proliferation, frequently display fundamental changes in pathways of energy metabolism and nutrient uptake.<sup>58</sup> Oncogenic transformation often coincides with both an increase in bioenergetic potential and the ability to take up nutrients in a cell-autonomous fashion.<sup>59</sup> The fact that metabolic enzymes such as succinate dehydrogenase (SDH) and fumarase (FH) function as tumor suppressors in human cells suggests that metabolic dysregulation can be an initiating event in cancer.<sup>59</sup> The recent discovery of activating mutations in isocitrate dehydrogenase (IDH1) in glioblastoma demonstrates that an activating mutation in a metabolic enzyme are selected during carcinogenesis.<sup>60</sup> The emergence of metabolic enzymes acting as important regulators of cancer cell growth highly suggests that metabolic control is a key element of tumor progression. In the current study, many transcripts that were differentially spliced in HCC samples are known to play key roles in cell metabolism. For example, the expression of the full-length ALDH2 transcript was found to be significantly reduced and replaced by a shorter form that is predicted to encode a truncated Aldh2 protein with a loss of a portion of the aldehyde dehydrogenase domain. Aldehyde dehydrogenase is the second enzyme of the major oxidative pathway of alcohol metabolism. Interestingly, the increased exposure to acetaldehyde in individuals with the catalytically inactive form has previously been shown to confer greater susceptibility to many types of cancer, including esophageal cancer.<sup>61</sup> Another example of a gene involved in metabolism for which AS is modified in HCC is *HADHA*, which encodes for the alpha subunit of the mitochondrial trifunctional protein

catalyzing the 3-hydroxyacyl-CoA dehydrogenase and enoyl-CoA hydratase activities. These steps are involved in mitochondrial beta-oxidation of long chain fatty acids. Our results showed that the expression of the full-length HADHA transcript was found to be significantly reduced and replaced by a shorter form that is predicted to encode a truncated Hadha protein with a loss of a portion of the enoyl-CoA domain. Interestingly, a link between *HADHA* and cancer has previously been observed in breast cancer where the *HADHA* gene was significantly under-expressed in cancerous tissues, especially in tumors with estrogen receptor-negative status.<sup>62</sup>

The expression of various splicing factors has been shown to be dysregulated in various human diseases. For instance, it was demonstrated that the expression of the Serine/Arginine-Rich Splicing Factor 1 (SRSF1) is increased in several tumor types<sup>63</sup>, and fibroblasts overexpressing SRSF1 caused tumor formation when injected into mice.<sup>64</sup> Similarly, an increase in the expression of the splicing factor hnRNPH has been shown to drive splicing switches of oncogenic target genes in gliomas.<sup>65</sup> Recently, RBM4, an RNA-binding factor involved in AS, was also demonstrated to control cancer-related splicing events, thereby affecting various cellular processes such as apoptosis, proliferation and migration.<sup>66</sup> In the current study, many changes in the expression levels and/or splicing patterns of splicing factors were detected in HCC samples. Interestingly, the transcript encoded by *ESRP1* was found to be significantly over-expressed in all types of HCC investigated in the current study (32-fold increase in HBV-associated HCC). How the altered expression dynamics of *ESRP1* (and other splicing regulators) contribute to AS homeostasis during virus infection remains to be investigated. Interestingly, a previous study identified ASEs regulated by *Esrp1* using RNA silencing technology in a human epithelial cell line.<sup>54</sup> Using such a strategy, the authors identified 148 alternative splicing events in a total of 134 different genes that were regulated by *Esrp1*.<sup>54</sup> In the present study, the splicing of 32 of these *Esrp1*-regulated transcripts are also found to be differentially spliced in HBV-associated HCC suggesting, at least in part, that some of the observed splicing alterations observed in HCC could be related to *Esrp1* over-expression. In the case of HCV-associated HCC and virus-free associated HCC, the splicing of 17 and 20 of these *Esrp1*-regulated transcripts were also found to be differentially spliced, respectively.

Large-scale studies are starting to reveal the extent of modifications that occur in AS in different types of cancer.<sup>22-24</sup> Interestingly, strategies to modulate AS by splice-switching oligonucleotides in order to correct aberrant ASEs, and/or to induce expression of therapeutic splice variants are currently being developed.<sup>34,38-39</sup> For instance, the AS of Bcl-x(L) in cancer cells can be redirected towards the pro-apoptotic variant Bcl-x(S) in order to reduce the tumor load in xenografts of metastatic melanoma.<sup>67</sup> It is therefore tempting to speculate that such a strategy could be applied to HCC. The current identification of extensive changes in the cellular AS landscape in HCC likely represents a first step toward the development of anticancerous agents based on the AS modifications identified in HCC. Molecular tools such as splice-switching oligonucleotides that can specifically alter the proportion of splice variants are also essential to assess the function of these splice variants.

### **COMPETING INTERESTS**

The authors declare that they have no competing interests.

### **AUTHORS' CONTRIBUTIONS**

M.P.T. contributed to experimental design and analysis of the data, performed validation experiments and helped in writing the manuscript; V.E.S.A. and A.A. helped in data analysis and in the generation of figures; E.L and M.D. performed PCR validations; P.T. performed informatics analyses; M.T.L., M.S. and J.P.P contributed to experimental design and analysis; M.B. contributed to experimental design, data analysis and wrote the manuscript with contributions from all the authors.

### **ACKNOWLEDGMENTS**

This work was supported by grants from the Centre de recherche du Centre hospitalier universitaire de Sherbrooke (M.B. and J.P.P) and the Faculté de médecine et des sciences de la santé de l'Université de Sherbrooke (M.B. and J.P.P). J.P.P holds the Université de Sherbrooke Research Chair in Structure and Genomic RNA. M.B. is a Chercheur-boursier Sénior from the Fonds de recherche du Québec-Santé (FRQ-S), and M.P.T. holds a graduate training award from the Faculté de médecine et des sciences de la santé de l'Université de Sherbrooke.

## SUPPLEMENTARY TABLES AND FIGURES

Supplementary table 1 - Bioinformatical prediction of functional changes caused by some of ASEs identified. (Negative delta-PSI means that the short isoform is favored and positive delta-PSI means that the long isoform is favored)

Gene	Delta PSI				Event type	Full name	Predicted residue variation caused by AS modification (AA)	Predicted functional consequence
	THBV/NNoV	THCV/NNoV	THBV&HCV/NNoV	TNoV/NNoV				
<b>A2M</b>	-13,1	N/S	N/S	N/S	exon-skipping	Alpha-2-macroglobulin	-150	Loss of signal peptide, loss of part of MG2 domain
<b>AAK1</b>	-10,3	N/S	N/S	N/S	alternate-3p	Adaptor-Associated Kinase 1	-1	Unknown/No predicted functional change
<b>ABCC10</b>	-15,1	N/S	N/S	-17,0	alternate-3p	ATP-Binding skipping Sub-Family C Member 10	28	Gain of signal peptide
<b>ACTA2</b>	-22,2	N/S	N/S	N/S	exon-skipping	Actin, Alpha 2, Smooth Muscle, Aorta	-45	Loss of part of actin domain
<b>ACVRL1</b>	-15,5	N/S	N/S	N/S	multiple-exon-skipping	Activin A receptor type II-like 1	-174	Loss of a transmembrane domain, loss of a TGFb domain, loss of a tyrosine kinase domain
<b>ADCY5</b>	16,2	N/S	N/S	N/S	exon-skipping	Adenylate cyclase 5	8	Unknown/No predicted functional change
<b>ALDH1A3</b>	-16,4	N/S	N/S	N/S	multiple-exon-skipping	Aldehyde dehydrogenase 1 family, member A3	-107	Loss of a part of aldehyde dehydrogenase domain
<b>ALDH2</b>	-12,1	N/S	N/S	N/S	exon-skipping	Aldehyde dehydrogenase 2 family (mitochondrial)	-47	Loss of a part of aldehyde dehydrogenase domain
<b>AMDHD1</b>	-18,4	N/S	N/S	N/S	exon-skipping	Amidohydrolase Domain Containing 1	-109	Loss a part of aminohydrolase family domain
<b>ANGPT1</b>	21,1	22,7	N/S	N/S	alternate-5p	Angiopoietin 1	1	Unknown/No predicted functional change
<b>ANKIB1</b>	-11,6	N/S	N/S	N/S	exon-skipping	Ankyrin repeat and IBR domain containing 1	-731	Loss of all domains (ankyrin repeat, ankyrin repeat 5, IBR domain (a half ring finger domain))
<b>ASPG</b>	-21,8	-19,7	N/S	-21,5	exon-skipping	Asparaginase	-18	Unknown/No predicted functional change
<b>AXIN1</b>	-10,9		N/S	N/S	exon-skipping	Axin 1	-36	Unknown/No predicted functional change
<b>BUB1B</b>	N/S	-24,6	N/S	-20,5	alternate-3p	BUB1 mitotic checkpoint serine/threonine kinase B	513	Gain a protein kinase domain and complete a Mad3/bub1 domain
<b>C16orf59</b>	29,0	31,9	N/S	33,6	alternate-3p	Chromosome 16 Open Reading Frame 59	43	Complete the DUF (domain of unknown function)



<b>C5orf34</b>	30,5	25,6	N/S	27,6	exon-skipping	Chromosome 5 Open Reading Frame 34	441	Complete the DUF4524 (domain of unknown function)
<b>CA13</b>	15,9	16,1	N/S	12,1	exon-skipping	Carbonic Anhydrase XIII	262 AA protein to no protein	Loss of the protein with a carbonic anhydrase domain due to NMD
<b>CAND1</b>	-15,7	N/S	N/S		alternate-5p+alternate-3p	Cullin-Associated And Neddylation-Dissociated 1	-490	Loss of transmembrane domains
<b>CENPL</b>	-25,4	-29,8	-34,3	-29,4	exon-skipping	Centromere protein L	-46	Unknown/No predicted functional change
<b>EHD3</b>	-12,7	N/S	N/S	N/S	multiple-exon-skipping	EH-Domain Containing 3	-360	Loss of 2 NLS, loss a part of dynamin domain
<b>EHF</b>	37,8	N/S	N/S	35,2	exon-skipping	Ets Homologous Factor	23	Unknown/No predicted functional change
<b>GTPBP4</b>	-22,6	N/S	N/S	N/S	exon-skipping	GTP Binding Protein 4	-116	Unknown/No predicted functional change
<b>HADHA</b>	-18,1	N/S	N/S	N/S	exon-skipping	Hydroxyacyl-CoA Dehydrogenase/3-Ketoacyl-CoA Thiolase/Enoyl-CoA	-87	Loss of a part of the enoyl-CoA hydratase/isomerase domain
<b>HAUS6</b>	-17,0	N/S	N/S	N/S	alternate-3p	HAUS Augmin-Like Complex, Subunit 6	-35	Unknown/No predicted functional change
<b>HBEGF</b>	-13,0	N/S	N/S	N/S	exon-skipping+alternate-5p+alternate-3p	Heparin-Binding EGF-Like Growth Factor	208 AA protein to no protein	Loss of the protein with an EGF-like domain due to the exclusion of the only exon coding for protein
<b>HGF</b>	-18,5	-17,6	N/S	-14,9	alternate-3p	Hepatocyte Growth Factor	-5	Unknown/No predicted functional change
<b>IGFALS</b>	N/S	10,1	N/S	N/S	alternate-3p	Insulin-Like Growth Factor Binding Protein, Acid Labile Subunit	38	Gain of transmembrane domains
<b>ITGB1</b>	24,7	N/S	N/S	N/S	exon-skipping	Integrin, Beta 1	3	Unknown/No predicted functional change
<b>KIFC1</b>	43,5	43,1	53,2	41,6	alternate-3p	Kinesin Family Member C1	8	Unknown/No predicted functional change
<b>KMO</b>	10,2	N/S	11,9	N/S	alternate-3p	Kynurenine 3-Monooxygenase	27	Unknown/No predicted functional change
<b>LRRC16B</b>	31,4	35,5	N/S	35,7	exon-skipping	Leucine Rich Repeat Containing 16B	947	Change of cellular localisation and loss of a part of CARMIL CTERM domain

<b>MARVELD2</b>	N/S	-15,1	N/S	N/S	exon-skipping	MARVEL (Membrane-Associating) Domain Containing 2	-12	Unknown/No predicted functional change
<b>MSS51</b>	-10,3	N/S	N/S	N/S	alternate-3p	MSS51 Mitochondrial Translational Activator	-221	Loss of a signal peptide, loss of a MYND finger domain
<b>MT1G</b>	-21,7	N/S	N/S	-13,8	alternate-3p	Metallothionein 1G	-1	Unknown/No predicted functional change
<b>MXD3</b>	39,5	34,2	42,7	35,3	intron-retention	MAX Dimerization Protein 3	195	Unknown/No predicted functional change
<b>NFKB1</b>	15,0	N/S	N/S	N/S	alternate-3p	Nuclear Factor Of Kappa Light Polypeptide Gene Enhancer In B-Cells 1	1	Unknown/No predicted functional change
<b>NOMO2</b>	N/S	N/S	21,1	N/S	exon-skipping	NODAL Modulator 2	167	Gain of a signal peptide
<b>PANX2</b>	10,1	N/S	N/S	N/S	exon-skipping	Pannexin 2	-134	Gain of innexin domain and of 2 transmembrane domains
<b>PHF14</b>	-12,1	-14,0	N/S	N/S	exon-skipping	PHD Finger Protein 14	-285	Unknown/No predicted functional change
<b>PRPF4B</b>	-34,5	-31,5	N/S	-27,0	alternate-5p+intron-retention	Pre-mRNA Processing Factor 4B	No protein to a 1007 AA protein	Gain of a tyrosine kinase protein (long isoform: NMD)
<b>PTPN1</b>	-22,0	N/S	N/S	N/S	exon-skipping	Protein Tyrosine Phosphatase, Non-Receptor Type 1	-73	Loss of a part of Y-phosphatase domain
<b>RAB7A</b>	-17,5	N/S	N/S	N/S	alternate-5p	Ras-Associated Protein RAB7	-47	Loss of half the Ras domain
<b>SFMBT1</b>	-10,1	N/S	N/S	N/S	exon-skipping	Scm-Like With Four Mbt Domains 1	-96	Loss of SAM1 domain
<b>SHQ1</b>	-13,5	-14,2	N/S	-13,1	alternate-3p	SHQ1, H/ACA Ribonucleoprotein Assembly Factor	89	Unknown/No predicted functional change
<b>SLIT2</b>	62,9	N/S	N/S	43,6	exon-skipping	Slit Homolog 2	8	Unknown/No predicted functional change
<b>SPAG4</b>	-27,2	N/S	-33,6	N/S	alternate-5p	Sperm Associated Antigen 4	-77	Unknown/No predicted functional change
<b>SPINT2</b>	N/S	-11,8	N/S	N/S	exon-skipping	Serine Peptidase Inhibitor, Kunitz Type 2	-57	Loss of Kunitz domain (inhibition of proteases)

<b>SUZ12</b>	-12,7	N/S	N/S	N/S	exon-skipping	Suppressor Of Zeste 12 Protein Homolog	-23	Unknown/No predicted functional change
<b>TPSB2</b>	N/S	N/S	19,3	N/S	alternate-3p	Tryptase Beta 2	152	Gain of signal peptide, gain of complete trypsin domain
<b>VPS13A</b>	-21,2	-16,2	-24,6	-18,0	exon-skipping	Vacuolar Protein Sorting 13 Homolog A	-39	Unknown/No predicted functional change
<b>WASF2</b>	-28,6	-14,5	N/S	-13,6	exon-skipping	Wiskott-Aldrich Syndrome Protein Family Member 2	-217	Loss of the WH2 motif  Gain of a protein with 2 AAA domains, a GRF zinc finger motif and a DUF2439 domain (short isoform: NMD)
<b>ZGRF1</b>	15,5	17,4	N/S	14,3	exon-skipping	Zinc Finger, GRF-Type Containing 1	No protein to a 2104 AA protein	Cytoplasmic protein (short isoform) goes to non- cytoplasmic (long isoform)
<b>ZNF35</b>	12,3	N/S	N/S	N/S	exon-skipping	Zinc Finger Protein 35	160	Unknown/No predicted functional change
<b>ZNF445</b>	-19,8	N/S	N/S	-20,2	alternate-5p	Zinc Finger Protein 445	-12	Unknown/No predicted functional change
<b>ZNF451</b>	-15,6	N/S	N/S	N/S	exon-skipping	Zinc Finger Protein 451, Coactivator For Steroid Receptors	-48	Unknown/No predicted functional change
<b>ZNF532</b>	-14,5	-12,7	N/S	N/S	alternate-5p	Zinc Finger Protein 532	-910	Unknown/No predicted functional change
<b>ZNF552</b>	-10,6	N/S	-16,4	-12,3	alternate-5p	Zinc Finger Protein 552	4	Unknown/No predicted functional change

---

**Supplementary table 2 – Protein families encoded by transcripts that are differentially spliced in various types of HCC**

Protein classification	THBV/NNoV		THCV/NNoV		THBV&HCV/NNoV		TNoV/NNoV	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Extracellular matrix protein (PC00102)	29	1.5%	12	1.4%	1	0.6%	18	1.2%
Protease (PC00190)	50	2.5%	18	2.0%	5	3.0%	38	2.5%
Cytoskeletal protein (PC00085)	100	5.1%	53	6.0%	10	6.0%	88	5.7%
Transporter (PC00227)	105	5.3%	56	6.3%	7	4.2%	91	5.9%
Transmembrane receptor regulatory/adaptor protein (PC00226)	4	0.2%	3	0.3%	1	0.6%	4	0.3%
Transferase (PC00220)	159	8.0%	64	7.2%	11	6.6%	121	7.8%
Oxidoreductase (PC00176)	68	3.4%	27	3.1%	5	3.0%	60	3.9%
Lyase (PC00144)	27	1.4%	11	1.2%	2	1.2%	16	1.0%
Cell adhesion molecule (PC00069)	52	2.6%	20	2.3%	3	1.8%	44	2.9%
Ligase (PC00142)	54	2.7%	32	3.6%	7	4.2%	49	3.2%
Nucleic acid binding (PC00171)	225	11.4%	114	12.9%	18	10.8%	173	11.2%
Signaling molecule (PC00207)	101	5.1%	39	4.4%	11	6.6%	61	4.0%
Enzyme modulator (PC00095)	147	7.4%	74	8.4%	16	9.6%	103	6.7%
Calcium-binding protein (PC00060)	40	2.0%	19	2.2%	5	3.0%	31	2.0%
Defense/immunity protein (PC00090)	63	3.2%	24	2.7%	4	2.4%	49	3.2%
Hydrolase (PC00121)	187	9.5%	77	8.7%	15	9.0%	144	9.3%
Transfer/carrier protein (PC00219)	39	2.0%	22	2.5%	2	1.2%	34	2.2%
Membrane traffic protein (PC00150)	54	2.7%	22	2.5%	7	4.2%	36	2.3%
Phosphatase (PC00181)	45	2.3%	17	1.9%	3	1.8%	38	2.5%
Transcription factor (PC00218)	156	7.9%	67	7.6%	15	9.0%	128	8.3%
Chaperone (PC00072)	15	0.8%	2	0.2%	0	0.0%	4	0.3%
Cell junction protein (PC00070)	10	0.5%	9	1.0%	2	1.2%	10	0.6%
Surfactant (PC00212)	6	0.3%	4	0.5%	0	0.0%	4	0.3%
Structural protein (PC00211)	9	0.5%	5	0.6%	1	0.6%	9	0.6%
Receptor (PC00197)	132	6.7%	53	6.0%	8	4.8%	112	7.3%
Kinase (PC00137)	73	3.7%	32	3.6%	6	3.6%	61	4.0%
Storage protein (PC00210)	5	0.3%	1	0.1%	0	0.0%	4	0.3%
Isomerase (PC00135)	20	1.0%	5	0.6%	1	0.6%	11	0.7%
Splicing factor	23	1.16%	13	1.6%	2	1.2%	15	1.0%
Tumor suppressor	138	6.7%	63	3.0%	10	6.0%	109	7.1%
Oncogene	9	0.5%	4	0.5%	1	0.6%	5	0.3%

Supplementary table 3 – List of tumor suppressors for which AS is dysregulated in various types of HCC

THBV/NNoV			THCV/NNoV		THBV&HCV/NNoV	TNoV/NNoV		
ABI2	EZH1	RAB7A	AFAP1L2	SAMD9L	CASP8	ABI2	HPGD	SRPX
ACHE	EZH2	RAP1GAP	BARD1	SCRIB	CTNND1	AFAP1L2	IKZF1	STRADA
AFAP1L2	FAS	RARB	BIN1	SEPT4	FHIT	AGTR1	IKZF2	TBRG1
AGTR1	FLNA	RASSF1	BRCA1	SHQ1	HACE1	ALDH1A2	ING1	TCF3
ALPL	FOXP1	RBL2	CD44	SIRT3	ITGA7	ALPL	ING4	TCF7L2
ARID2	GADD45A	RBM6	CDKN2A	SMARCB1	KRIT1	ARNTL	IRF3	TFPI2
ARNTL	GLS2	RBMX	CEACAM1	SPINT2	LIMA1	BARD1	ITGA7	THRB
AXIN1	GPC3	RINT1	CHEK1	TBRG1	NDRG2	BCL2L11	KDM8	TRIT1
BARD1	GSN	RNF111	CIZ1	TCF3	NF2	BIN1	KMT2C	TSC1
BCL2L11	GSTT1	RNH1	COPS2	TCF7L2	TET2	BLNK	KRIT1	VDR
BIN1	GTPBP4	RPL10	CREM	TP73		BRCA1	LRP1B	VEGFA
BLNK	HPGD	RUNX3	CSMD1	TRIT1		BRF1	MAD1L1	VEZT
BRCA1	HTATIP2	SAFB	CTNNA3	TSC2		CADM2	MFSD2A	VPS53
BRF1	ING1	SAMD9L	DIABLO	VDR		CASP8	MLH1	WVVOX
BTK	ING3	SCRIB	DLG1	VEGFA		CCDC136	MT1G	XAF1
CABLES1	ING4	SDHA	DNMT3B	VEZT		CCNDBP1	MTSS1	ZMYND11
CADM2	IRF3	SEMA3B	DOK1	WNK2		CD4	MTUS1	ZNF382
CASP8	ITGA7	SEPT4	EMP1			CD44	NDRG2	
CCAR1	ITGB1	SHQ1	ESR1			CDK2	NF2	
CCNDBP1	KANK1	SIRT2	FLNA			CEACAM1	NPAS2	
CD44	KDM8	SIRT3	GLS2			CHEK1	NRCAM	
CDK2	KLF6	SKP2	GPC3			CIZ1	NUMB	
CEACAM1	KMT2C	SLIT2	GSN			CREM	PARK2	
CHEK1	KRIT1	SMARCB1	HPGD			CSMD1	PBRM1	
CHEK2	MAD1L1	SPOP	IGFALS			CTCFL	PLAGL1	
CIZ1	MBD4	SRPX	ING1			CTNNA3	PLCB3	
CNDP2	MFSD2A	STRADA	ING3			CTNND1	PML	
CREM	MME	SUZ12	ING4			CUX1	PPARA	
CSMD1	MT1G	TBRG1	IRF3			CYLD	PRR5	
CTCF	MTSS1	TCF3	ITGA7			DICER1	PTCH1	
CTNNA3	MTUS1	TCF7L2	KANK1			DLG1	PTPN6	
CTNND1	NDRG2	TES	KDM8			DNMT3B	RAP1GAP	
CYLD	NF1	TET2	KMT2C			ECT2	RASSF1	
DDB2	NF2	TFPI2	KRIT1			EDA2R	RASSF4	
DLG1	NFKB1	THRB	MAD1L1			EGR2	RBM6	
DMD	NUMB	TRIM35	MLH1			EHF	RNH1	
DNAJA3	PANX2	TRIT1	MTSS1			EI24	RPL10	
DNMT3B	PBRM1	TSC1	NDRG2			ESR1	SCRIB	
EEF1A1	PDCD4	UHRF2	NF1			EZH1	SEC14L2	
EGR2	PHACTR4	UVRAG	NF2			EZH2	SEMA3B	
EHD3	PLAGL1	VEGFA	NUMB			FOXP1	SEPT4	
EHF	PPARA	VEZT	PBRM1			GLS2	SHQ1	
EI24	PPP2R1B	VPS53	PPARA			GPC3	SIRT3	
EMP1	PTCH1	WVVOX	PTPN6			GSN	SLIT2	
EPB41	PTPN1	XAF1	RASSF4			GSTT1	SMARCA4	
ESR1	PTPN6	ZNF382	RPL10			HOPX	SMARCB1	

Supplementary table 4 - List of oncogenes for which AS is dysregulated in various types of HCC

---

THBV/NNoV	THCV/NNoV	THBV&HCV/NNoV	TNoV/NNoV
CBFB	ETS1	CBFB	AXL
DEK	KRAS		CBFB
ERBB2	MYH11		MDM2
ERG	TSC2		MYH11
ETS1			PML
KRAS			
MDM2			
MYH11			
TAL1			

---

Supplementary table 5 - List of kinases for which AS is dysregulated in various types of HCC

THBV/NNoV		THCV/NNoV	THBV&HCV/NNoV	TNoV/NNoV	
ALPK1	MAST4	AURKA	AURKB	AURKA	MPP1
AURKA	MELK	AURKB	CDK1	AURKB	MPP3
AURKB	MORN4	CAMK2D	PI4KAP1	BRSK2	MPP4
AURKC	MPP3	CDC42BPA	PIK3CB	CAMK2D	MVK
BRSK2	MVK	CDK1	RPS6KA4	CAMK2G	NEK2
BTK	NRBP2	CDK10	SLAMF7	CD244	NRBP2
CAMK2D	PANK1	CHKB		CDC42BPA	PAPSS2
CD244	PAPSS2	DGKZ		CDK1	PASK
CD84	PASK	DGUOK		CDK10	PI4KAP1
CDC42BPA	PFKM	DMPK		CDK2	PIK3CG
CDK1	PI4KAP1	FES		CDK4	PRPF4B
CDK10	PIK3CB	FYN		CDKL3	PRPS2
CDK14	PIP4K2C	GK5		CERKL	PTK7
CDK2	PIP5K1A	IP6K2		CHKB	RIPK2
CDK20	PRKDC	KHK		CLK3	SGK3
CDK4	PRPF4B	MAGIX		CLK4	SLAMF7
CDKL3	PRPS1	MAP3K6		CSNK1G1	SPHK2
CHKB	PRPS2	MORN1		DGKG	SRPK1
CIT	RIPK1	MORN4		DGKH	STK16
CMPK1	RIPK2	PAPSS2		DGKZ	TK2
DGKG	SGK1	PI4KAP1		DGUOK	TNK2
DGKZ	SGK3	PIK3CB		DSTYK	ULK3
DGUOK	SLAMF6	PRPF4B		FASTK	VRK2
EIF2AK2	SLAMF7	PTK6		FES	WNK4
FES	SPHK2	RIPK1		FGGY	
FYN	SRPK1	RPS6KA4		FYN	
GK5	STK16	SPHK2		GK5	
GNE	TANC1	STK16		IKBKB	
HCK	TESK2	ULK3		IP6K2	
IKBKB	TK2	WNK1		KHK	
IKBKE	TNK2	WNK2		KSR1	
IP6K2	ULK3	WNK4		MAP3K6	
KHK	VRK2			MAPK13	
KSR1	WNK1			MARK1	
MAGIX	WNK3			MAST4	
MAPK13	WNK4			MELK	
MARK3				MORN4	

Supplementary table 6 - List of transcription factors for which AS is dysregulated in various types of HCC

THBV/NNoV				THCV/NNoV		THBV&HCV/NNoV	TNoV/NNoV		
ARHGAP5	KBTBD4	RNH1	ZNF35	BCL11A	TCEA3	ARHGAP5	ARNTL	NCOR1	ZBTB24
ARNTL	KDM4A	RRN3	ZNF382	CARD8	TCF20	HDGFRP2	ATXN3	NFAT5	ZBTB38
ARNTL2	KDM5D	RUNX3	ZNF408	CCNL2	TCF3	KAT6A	BCL11A	NFATC4	ZBTB41
BCL11A	KRBOX4	SFMBT1	ZNF419	CNOT1	TFDP2	LIMA1	BRF1	NFE2	ZFP64
BRF1	LMO2	SMARCE1	ZNF445	CNOT8	TP73	MXD3	C3orf67	NFE2L2	ZMYM6
BTBD1	LRRC16B	SOX13	ZNF451	CREM	UBAC2	NLRC5	CARD8	NFIA	ZMYND11
C3orf67	LRRFIP2	SPOP	ZNF502	CRIP2	UTY	SMARCC2	CBFA2T2	NFYA	ZNF107
CAND1	MKL1	SREBF1	ZNF512	E2F7	VDR	SUPT20H	CCNL2	NFYC	ZNF138
CARD8	MLXIPL	STAT6	ZNF532	ETS1	XPNPEP3	TLE2	CDIP1	NLRC5	ZNF169
CBFA2T2	MORF4L2	SUPT20H	ZNF550	FHL2	YY1AP1	ZBTB24	CNOT8	NLRP6	ZNF195
CCNL2	MPND	TAF11	ZNF552	FOSB	ZBTB38	ZMYM1	CREB3L4	NPAS2	ZNF200
CDIP1	MSS51	TAF6	ZNF567	FOSL1	ZEB1	ZMYND8	CREM	NR113	ZNF233
CIITA	MXD3	TAL1	ZNF618	FOXM1	ZHX3	ZNF211	CRIP2	NR2C1	ZNF248
CLOCK	NCOA3	TBX3	ZNF707	GATA3	ZNF138	ZNF552	CTCF	NR3C1	ZNF266
CNOT1	NCOR1	TCEA3	ZNF793	HDGFRP2	ZNF169	ZNF667	CUX1	PDLIM2	ZNF274
CNOT8	NFAT5	TCF20	ZNF85	HMGB2	ZNF195		E2F6	PDLIM5	ZNF285
CREM	NFATC3	TCF3	ZNF852	HNF1B	ZNF248		E2F7	PHC2	ZNF331
CRIP2	NFIA	TEAD2	ZSCAN2	HOXB6	ZNF532		EHF	PLAG1	ZNF346
CTCF	NFKB1	TFDP2		ING1	ZNF550		ELK1	PLAGL1	ZNF382
E2F6	NEFYA	TFEB		ING4	ZNF559		FHL2	POGZ	ZNF408
E2F7	NLRC5	TFEC		IRF3	ZNF707		FOSB	PPARA	ZNF431
EHF	NLRP6	THRB		JMJD1C			FOSL1	PQBP1	ZNF445
ELF1	NR1D1	TLE2		KDM5D			FOXM1	PRDM16	ZNF446
ELK1	NR113	TOX4		LRRC16B			FOXP1	RARG	ZNF502
ERG	NR2C1	TRIM35		LRRFIP2			FOXP2	REST	ZNF512
ETS1	NR3C1	UBAC2		MARVELD2			GATA3	RNH1	ZNF527
FHL2	NR3C2	USF2		MLXIPL			HDGFRP2	RRN3	ZNF550
FOSB	PDDC1	UTY		MORF4L2			HMGB2	SMARCE1	ZNF552
FOSL1	PDLIM2	WIZ		MXD3			HOXB6	STAT6	ZNF582
FOXK2	PDLIM5	XPNPEP3		MYNN			IKZF1	SUPT20H	ZNF618
FOXM1	PEPD	YY1AP1		NCOA1			IKZF2	TAF6	ZNF707
FOXP1	PHC2	ZBTB24		NCOA3			ING1	TCEA3	ZNF738
GABPA	PHF14	ZBTB38		NCOR1			ING4	TCF12	ZNF79
GATA3	PLAG1	ZBTB41		NFYA			IRF3	TCF20	ZNF793
GCFC2	PLAGL1	ZEB1		NR113			IVNS1ABP	TCF3	ZNF821
GNE	POGZ	ZEB2		NR2C1			JMJD1C	TEAD2	ZSCAN2
HDGFRP2	PPARA	ZFP64		NR3C1			KDM5D	TEAD4	
HMGB2	PPARD	ZHX3		PHC2			LMO2	TFDP2	
HNF1B	PQBP1	ZNF107		PHF14			LRRC16B	THRB	
HOXB6	PRDM16	ZNF131		PLAG1			LRRFIP2	TLE2	
ING1	PRRC2B	ZNF195		POGZ			MKL1	UBAC2	
ING4	RARA	ZNF200		PPARA			MLXIPL	UTY	
IRF3	RARB	ZNF211		PQBP1			MORF4L2	VDR	
IRF6	RARG	ZNF233		PRRC2B			MPND	XPNPEP3	
IVNS1ABP	RBL2	ZNF248		PRRC2C			MXD3	YY1AP1	
JMJD1C	RELA	ZNF331		RRN3			NCOA3	ZBTB17	



**Supplementary table 7 - List of genes for which AS is  
dysregulated in all types of HCC**

ASEs misregulated in the four types of HCC				
ACP1	CENPL	INCENP	OGFOD2	SLC2A6
ACSL4	CEP164	ISYNA1	OSBPL6	SLC7A7
ASPM	CLN3	ITGA7	PARPBP	SORT1
ATXN2L	DHRS4L2	KIAA0895L	PCYT2	ST3GAL3
AURKB	ENO3	KIAA1324	PDE4DIP	STAU2
C1orf86	FAM86C1	KIFC1	PHF3	STK25
C5orf42	FDFT1	KRIT1	PI4KAP1	TAMM41
C6orf201	FETUB	MDK	PIK3R3	TJP2
CARD14	FN1	MICU1	PMM2	VEGFB
CARKD	GOLGA2	MSTO1	POLL	VPS13A
CCDC64	GPR116	MXD3	PORCN	WASH2P
CD27-AS1	HDGFRP2	NDRG2	RCC1	WASH7P
CDC45	HNRNPUL1	NEB	RGS3	ZFYVE19
CDK1	IL17RE	NF2	SEC31B	ZNF655
CDS2	IL32	NIPA2	SKA2	ZNF692

Supplementary table 8 - List of genes uniquely dysregulated in HBV-associated HCC

ASEs misregulated in only THBV/NoV										
A2M	BAZ2A	CR2	F7	GUSBP4	LOC100506054	NKTR	POFUT2	RSU1	STEAP1B	TSPO
AAK1	BBS4	CRYZ	F9	HADHA	LOC155060	NMRAL1	POGK	RTN2	STOM	TTC7A
ABCB7	BCL2L14	CSDE1	FADS6	HAL	LOC619207	NOL11	POLB	RTN3	STRAP	TUFT1
ABCD3	BCS1L	CSF3R	FAM153A	HAUS6	LOC654433	NOVA1	POLDIP3	RUNX3	STT3A	TXN2
ABHD10	BDP1	CTCF	FAM179B	HBEGF	LRIF1	NPEPPS	POLR2M	SAFB	STX16	UACA
ABI3BP	BIVM	CTH	FAM198A	HCK	LRRRC61	NPHP3	POLR3D	SAP130	STX5	UBAP1
ACAD8	BRAP	CTNNA1	FAM199X	HEATR5A	LSS	NR1D1	PPARD	SBF1	SULF1	UBAP2L
ACADSB	BRWD3	CWF19L2	FAM206A	HEATR5B	LTBP3	NR3C2	PPIL2	SBNO1	SUMF2	UBE2K
ACCS	BSDC1	CXorf38	FAM210A	HECTD1	LUZP6	NRBF2	PPM1E	SCAND2P	SUN1	UBL7
ACHE	BTBD1	CYBRD1	FAM228A	HECTD3	LYPLA1	NRG3	PPP1R10	SCARB1	SUN2	UBQLN4
ACLY	BTK	CYFIP2	FAM3C	HERC4	LYSMD3	NSG1	PPP1R8	SDC4	SURF4	UCHL5
ACMSD	C10orf118	CYP2B6	FAM57A	HHAT	MAGT1	NUCB2	PPP2R1A	SDHA	SUV420H1	UHRF2
ACSS2	C14orf159	CYP2C8	FAM63B	HINT3	MAN2B2	NUDT9	PPP2R1B	SDHAP2	SUZ12	USF2
ACTA2	C15orf39	CYP39A1	FAM65C	HMGCS2	MANEAL	NUP50	PPP2R5D	SEC23A	SWAP70	USP15
ACTL6A	C16orf62	CYP4F2	FAS	HNRNPC	MAP4	NUP54	PPP4R1	SEC23IP	SYNE1	USP36
ACVRL1	C21orf58	DAO	FASTKD2	HNRNPH3	MAPKAP1	OGFOD1	PRG4	SEC24D	SYNJ2	UVRAG
ADAM10	C2CD5	DAXX	FBLN7	HP1BP3	MARCH6	OGT	PRKCSH	SEC61A1	SYP	VAC14
ADAR	C6orf106	DAZAP2	FBXL5	HSD11B1L	MARK3	ORC4	PRKDC	SECISBP2	TAF11	VAMP7
ADCY3	CABLES1	DCAF7	FBXO3	HSD17B4	MATN3	OXR1	PRMT2	SEMA5B	TAL1	VASH2
ADCY5	CABYR	DCDC5	FGFR4	HSD3B7	MAVS	P2RX4	PRMT5	SENP6	TANC1	VMP1
ADH6	CACNA1H	DDDB2	FKBP10	HSDL2	MBD1	P4HA2	PRNP	SEPT10	TARS	VNN2
ADPGK	CALCOCO2	DEK	FKBP5	HSF4	MBD4	PADI1	PRPS1	SERPINF2	TBC1D4	VPS16
AGPAT1	CALD1	DET1	FKRP	HSPH1	MBNL3	PAK4	PRSS8	SETD4	TBRG4	VPS26A
AGXT2L1	CAND1	DGAT2	FLJ00385	HTATIP2	MCCC1	PANK1	PSMC1	SFMBT1	TBX3	VPS41
ALB	CANX	DGCR2	FLYWCH1	HYLS1	MCCC2	PANX2	PSMD3	SFXN4	TCP11L1	VSIG4
ALDH18A1	CAP1	DHCR24	FLYWCH2	IAH1	MCOLN3	PARP9	PSME4	SGCD	TCTN3	VWF
ALDH1A3	CAPN1	DHX36	FBNP1L	IARS	MEAF6	PCBP1-AS1	PTCD3	SGCE	TDRD9	WAC
ALDH2	CBWD3	DHX40	FOLH1	IBTK	MEF2D	PCGF3	PTDSS1	SGK1	TECPR2	WDR44
ALPK1	CCAR1	DHX58	FOPNL	ICAM1	MEMO1	PCGF6	PTPLAD1	SGSM3	TES	WDR6
AMD1	CCDC111	DIAPH1	FO XK2	IDH3A	METTL23	PCK1	PTPN1	SH3PXD2A	TESK2	WIPF2
AMDHD1	CCDC135	DIO1	FTSJ2	IDI1	METTL5	PCK2	PTPN21	SHANK2	TFEB	WIZ
AMPD3	CCDC25	DMD	FUT6	IFNLR1	MFN1	PCNX	PVRL4	SIGLEC10	TFEC	WNK3
ANKHD1	CCDC90B	DMGDH	FZD6	IGSF9	MFSD12	PCNXL4	PWWP2A	SIMC1	TGM2	WSB2
ANKIB1	CCM2	DNAJA1	G3BP1	IKBKE	MGAT4B	PCYOX1	PYGL	SIRT2	THEMIS	XIRP1
ANKRD19P	CCNG1	DNAJA3	G6PC	IL10RA	MGME1	PDCD1	RAB18	SKIV2L2	THSD7B	XPC
ANKS6	CCSER2	DNAJC19	G6PC3	IL18RAP	MINK1	PDCD4	RAB2B	SKP2	THTPA	XRCC6
ANXA8	CD200	DNM2	GABPA	IL6ST	MIOS	PDDC1	RAB5A	SLAMF6	TIMM10B	XRN2
AOX1	CD300A	DNPEP	GABPB1	INTS9	MLIP	PDE3B	RAB5B	SLC13A5	TLR2	YWHAE
AP1AR	CD34	DPP3	GADD45A	IREB2	MLLT4	PDE8B	RAB6A	SLC15A4	TLR4	ZBTB21
AP5M1	CD68	DRAM2	GCFC2	IRF6	MLPH	PDGFC	RAB7A	SLC17A1	TM9SF2	ZC3H14
APH1A	CD84	DSCR3	GDA	ITCH	MME	PDHB	RABEP1	SLC23A2	TMBIM1	ZC3HC1
API5	CDC25B	DUS3L	GDE1	ITFG3	MMP19	PDHX	RABGAP1	SLC25A13	TMEM116	ZCCHC6
APMAP	CDK14	DUSP16	GDI2	ITGB1	MMRN2	PDXDC1	RAI2	SLC25A43	TMEM136	ZCWPW1

APP	CDK20	DYM	GEM	IWS1	MNS1	PDXDC2P	RAP1B	SLC27A2	TMEM161B	ZDHHC19
AQP3	CDS1	DYNC1L1	GEMIN7	JKAMP	MOB2	PDZK1	RARA	SLC2A14	TMEM167B	ZDHHC23
AQP6	CECR2	DYNC2L1	GEMIN8	KBTBD4	MON1B	PEPD	RARB	SLC2A2	TMEM170A	ZDHHC5
ARCN1	CELF2	EBF4	GFM2	KCNT2	MRC2	PFKFB4	RARS	SLC30A5	TMEM216	ZEB2
ARFGAP3	CERS2	ECD	GFOD2	KDM4A	MROH7	PFKM	RASSF6	SLC35B2	TMEM254	ZFYVE9
ARHGEF10L	CFHR4	EEF1A1	GLE1	KIAA0125	MRPL3	PGGT1B	RBBP5	SLC38A2	TMEM254-AS1	ZMAT5
ARID2	CHD6	EFTUD2	GLYATL1	KIAA0232	MSS51	PGRMC1	RBL2	SLC38A7	TMEM30A	ZNF131
ARL5A	CHEK2	EHD3	GMPS	KIAA1598	MST1P2	PHACTR4	RBM23	SLC44A3	TMEM44	ZNF302
ARL6IP1	CHRD	EIF2AK2	GNAS	KIDINS220	MTA1	PI16	RBM24	SLC7A9	TMEM63B	ZNF326
ARL8B	CIDEC	ELAC2	GNB1	KLF6	MTF2	PIBF1	RBM42	SLC9A8	TMEM66	ZNF35
ARNTL2	CIITA	ELF1	GNB5	KLF7	MTFR1	PIGA	RBM47	SLMAP	TMEM8A	ZNF410
ASB13	CIRH1A	ELN	GNE	KLHL12	MTM1	PIGG	RBMX	SMARCD1	TMPRSS4	ZNF415
ASH2L	CIT	EML4	GNL3	KLHL2	MTMR14	PIGT	RCBTB2	SMEK1	TMTC1	ZNF419
ASPH	CLASRP	ENOPH1	GNS	KLRD1	MYD88	PIK3R5	RDH16	SMIM8	TNKS1BP1	ZNF451
ASPN	CLOCK	ENOX2	GOLGA5	KPNB1	MYLIP	PIP4K2C	RELA	SNAP23	TOLLIP	ZNF567
ASUN	CMPK1	ENPP2	GOLPH3L	KRBOX4	MYO9B	PIP5K1A	RETSAT	SNTA1	TOP2B	ZNF839
ATG16L1	CNDP2	EOGT	GOT2	LAMP1	NAA60	PLA2G1B	RGPD1	SOX13	TOR3A	ZNF85
ATP2B2	CNPY3	EPB41	GPNMB	LASP1	NAE1	PLEKHH3	RHPN2	SPAG9	TOX4	ZNF852
ATP2B4	CNTRL	ERBB2	GPR137	LGMN	NARG2	PLEKHO2	RINT1	SPECC1L	TPRG1L	
ATP2C1	COG3	ERBB3	GPR161	LILRB1	NARS	PLS3	RMDN2	SPOP	TRAP1	
ATP5SL	COL18A1	ERC1	GPR180	LILRB2	NBEAL1	PMPCA	RNF111	SPSB3	TRAPPC11	
ATP6V1C1	COL1A2	ERG	GRAMD1B	LILRB3	NCAPG2	PMS2CL	RNF4	SRD5A1	TRAPPC6A	
ATP7B	COL6A6	ERMP1	GRB2	LILRB5	NDC1	PMS2P5	RNGTT	SREBF1	TRAPPC6B	
ATP9B	COQ5	ERVK13-1	GSS	LIPC	NDUFS5	PNPO	RPH3AL	SRP54	TRIM35	
AURKC	CORO1C	ESYT2	GSTO2	LIPG	NET1	PODN	RRAS2	SRP72	TRIM41	
AXIN1	CPD	F11R	GTPBP4	LMAN2L	NFATC3	PODXL	RRM1	SSFA2	TRIM5	
BACE1	CPSF7	F3	GUCY1B3	LMBR1	NFKB1	POFUT1	RSAD1	ST6GAL1	TRPV3	

**Supplementary table 9 - List of genes uniquely  
dysregulated in HCV-associated HCC**

ASEs misregulated in only THCV/NNv			
AASS	FCGR3A	MORN1	SLC10A7
ANXA2	FGFR10P	MRPL55	SMARCAD1
ATP6V0B	FLJ39739	MTHFSD	SNX5
CACNB4	FUT10	MYNN	SPINT2
CCDC7	GLT1D1	N4BP2L1	SUSD1
CD22	HNRNPH1	NAALAD2	TFPI
CDKN2A	IGFALS	NCOA1	TGIF1
CHRNA4	IL9R	NEIL2	TP73
COPS2	INIP	PHGDH	TROVE2
DIABLO	JHDM1D	PHYHD1	TSC2
DMPK	KAZN	PPP3CB	WNK2
DNAJC18	KLHDC1	PRRC2C	WNT2
DOK1	LINC00875	PTK6	ZNF106
EFCAB1	LOC388152	R3HDM2	ZNF30
EMR1	MAEL	RACGAP1	ZNF559
FAM83A-AS1	MARS	RYK	ZNF568
FBXO28	MARVELD2	SAMD4A	ZNF638

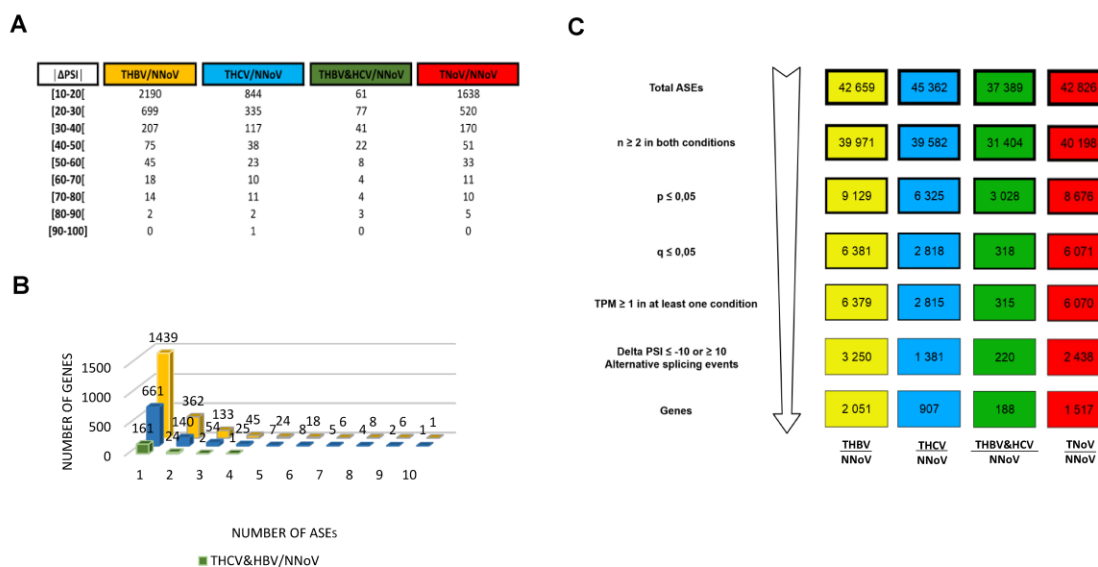
**Supplementary table 10 - List of genes uniquely  
dysregulated in HBV&HCV-associated HCC**

ASEs misregulated in only THEV&HCV/NNv		
ALAD	HACE1	RGN
ALOX12P2	KAT6A	RPL15
ARHGAP44	KIFC3	SMARCC2
ATP6V1H	KMT2E	SNRPN
C7orf63	LANCL1	SYBU
CD28	LIMA1	TCIRG1
CD8B	LIN54	TPSB2
CNTNAP3	MCTP1	TRABD2A
CYP21A2	MIB2	VANGL1
DHX35	NBPF10	WDPCP
FAM135A	NBPF8	WEE2-AS1
FAM13A	NCEH1	XRCC1
FHIT	NOMO2	ZCCHC7
FYB	NR2F1-AS1	ZFAT
GDPD5	ORAI2	ZMYM1
GIGYF2	PACSIN3	ZMYND8
GNL1	PMEPA1	ZNF182
GRB10	RBMS1	ZNF667

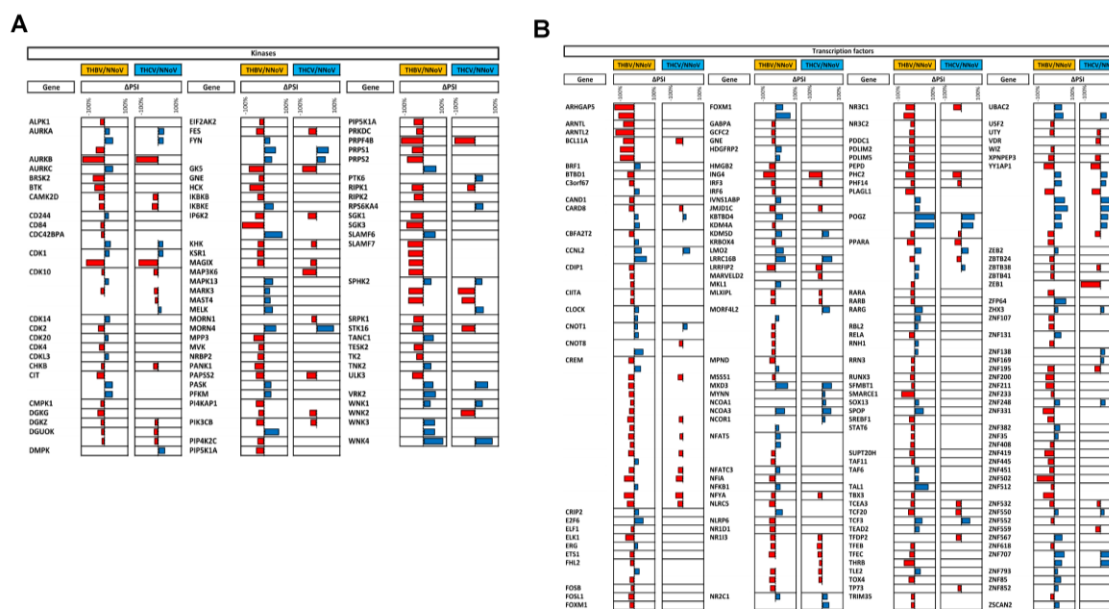
Supplementary table 11 - List of genes uniquely dysregulated in virus-free HCC

ASEs misregulated in only TNoV/NNoV				
AADAT	CSNK1G1	JOSD2	PIK3CG	STAC3
ACSF2	CTCF	KANSL1	PLCB3	STAG3L3
ACSL6	CUL9	KCNK17	PLEKHS1	STARD3
ADORA2A	CUTC	KCNMB3	PML	SUV39H2
AKAP17A	CUX1	KIAA0391	POLR1B	SVEP1
ALDH1A2	CWF19L1	KIAA0895	PPCS	TBC1D17
AMZ2	CYB561D1	KIF20B	PPP2R2B	TBCK
APOBEC3B	DCAF11	KLHL3	PRDX1	TCF12
ARHGAP30	DCUN1D2	KLRF1	PROM1	TDP1
ARMC7	DDX11L1	LGALS9	PRR5	TEAD4
ATP13A4	DGKH	LIMCH1	PSMD13	TESPA1
ATRN1	DICER1	LINC00910	PSPH	TIMD4
ATXN1	DIO2	LOC100287534	PSTPIP1	TM4SF19
ATXN3	DIO3AS	LOC400927	PTK7	TM7SF2
ATXN7L3	DIP2A	LOXL2	PTPN20A	TMEM67
AXL	DMAP1	LOXL3	PTPN20B	TNFRSF25
BBS1	DSTYK	LRP1B	PTPRS	TOMM40L
BBX	ECT2	LRRC23	PYCR1	TPK1
BCAS1	EDA2R	LRRC7	RAB37	TREML1
BCL7B	EPHA5	LRRK1	RABL5	TRIM39
BDKRB2	ERCC6L2	LRRN2	RAD51AP1	TRPC4
BPY2	FAM213B	LUZP2	RAPGEF1	TLL3
BTBD7	FAM21A	MARK1	RAPGEF3	TLL7
BTNL8	FAM227A	MAZ	RBM14-RBM4	TTYH1
BTRC	FAM92A1	MCM2	REPS1	UBE2J2
C12orf65	FASTK	MDH1B	RERG	UBE2W
C1orf109	FBXL2	MICAL2	REST	UNK
C1QTNF6	FGGY	MPC2	RGL4	UPF3B
C21orf91	FOXP2	MPP1	RGS11	USP25
CADPS	FRMD1	MPP4	RIMBP2	UXS1
CALML4	FUZ	MRPL52	RINL	VIPR2
CAMK2G	GALNT14	MS4A14	RPGRIP1	VWA1
CAPN12	GALNT6	MSANTD2	RPL23AP7	VWDE
CCDC136	GAPVD1	MTCH1	RRAGC	WASF3
CCDC149	GATS	NDUFAF7	RTEL1-TNFRSF6B	WASH1
CCDC91	GDAP1	NECAB2	RUNDC3B	WDR20
CD320	GEMIN5	NEK2	SCAMP4	WDR81
CD4	GEN1	NFASC	SCARA5	WSCD1
CDAN1	GLIS3	NFATC4	SEC14L2	ZBTB17
CDC25A	GPM6A	NFE2	SENP7	ZDHHC15
CEP97	GRAP2	NFE2L2	SH2D1A	ZMYM6
CERKL	GUCY1B2	NFYC	SIGMAR1	ZMYND11
CFLAR	HDAC7	NHSL2	SLC12A4	ZNF266
CHN1	HDAC8	NLGN4X	SLC14A1	ZNF274
CHRNA7	HELB	NPAS2	SLC25A12	ZNF285
CIB2	HERC2P3	NRCAM	SLC25A19	ZNF334
CLCN7	HFE	NXF3	SLC26A3	ZNF346
CLEC4A	HLA-G	ORAOV1	SLC30A10	ZNF431

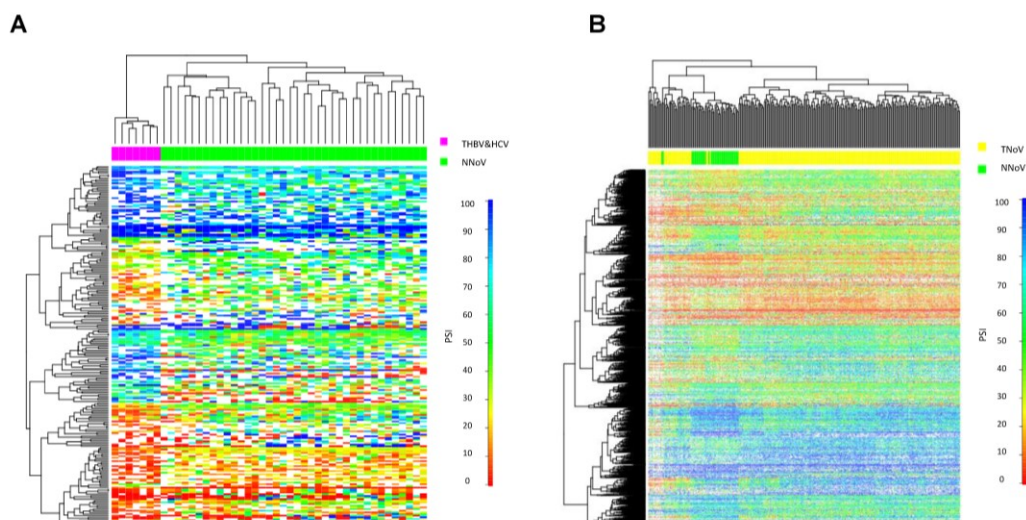
CLIP1	HMCN1	P2RX2	SLC35F6	ZNF446
CLK3	HNRNPA1	P4HA1	SLC35G1	ZNF454
CLK4	HOPX	PAK3	SLFN12	ZNF527
CNIH2	HPSE	PAOX	SMARCA4	ZNF534
CNTN5	ICA1	PAPD5	SNAP91	ZNF582
COL24A1	IGFLR1	PARK2	SNCAIP	ZNF738
CPA5	IKZF1	PCDH18	SNRPA1	ZNF79
CRCP	IKZF2	PDCD1LG2	SPATA9	ZNF791
CREB3L4	IL34	PDE10A	SPDL1	ZNF821
CRYZL1	INF2	PGBD2	SPEF2	ZNF823
CSF2RA	ISCU	PGPEP1	SSPO	ZSCAN25
CSF2RB	IYD	PIFO	ST7L	



**Figure S1. Characterization of splicing dysregulation in HCC. (A)** Distribution of Delta PSI values in various types of HCC. **(B)** Distribution of multiple alternative splicing events (ASEs) in each gene for various types of HCC. **(C)** The splicing events lists for various types of HCC were filtered to keep only data with at least two replicates for both virus-induced HCC and virus-free healthy tissues (NNoV). Fold changes were then calculated between HCC and healthy tissues. Q-values (false-discovery rate) were calculated in order to take into account multiple statistical hypothesis testing, and results under 0.05 were considered significant. Events with a P-value of less than 0.05 were conserved. Events with expression levels higher than one TPM in at least one dataset were conserved.

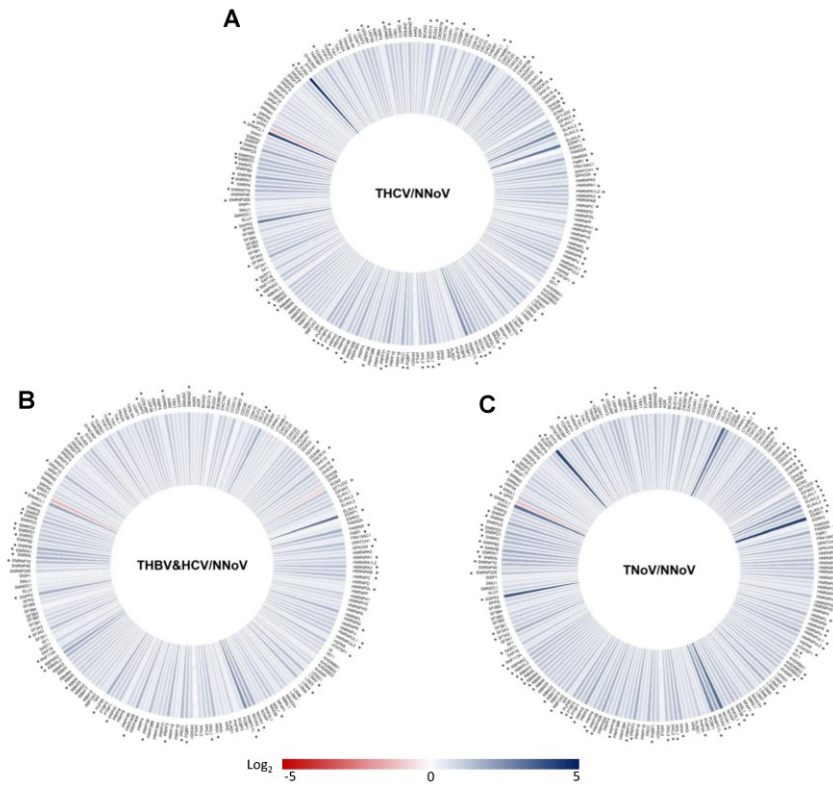


**Figure S2. AS modifications in transcripts encoded by kinases and transcription factors in HBV- and HCV-associated HCC. (A)** Alterations in the AS patterns of transcripts encoded by kinases in HBV- and HCV-associated HCC. Some transcripts have multiple ASEs that are modified. Red bars indicate negative Delta PSI values, and blue bars represent positive Delta PSI values. **(B)** Alterations in the AS patterns of transcripts encoded by transcription factors in HBV- and HCV-associated HCC.



**Figure S3. Global profiling of ASE modifications in both HBV&HCV-associated HCC and virus-free-associated HCC. (A)** Heatmap representation of ASEs (PSI values) for cellular transcripts in HBV&HCV-associated HCC. Data obtained from THBV&HCV (HBV&HCV-associated HCC) tissues are shown in purple and data obtained from healthy tissues (NNoV) are shown in green. **(B)** Heatmap representation of ASEs (PSI values) for cellular transcripts in virus-free-associated HCC. Data obtained from TNoV (virus-free-associated HCC) tissues are shown in yellow and data obtained from healthy tissues (NNoV) are shown in green.

cellular transcripts in virus-free HCC. Data obtained from virus-free-associated HCC tissues are shown in yellow and data obtained from healthy tissues (NNoV) are shown in green.



**Figure S4. RNA splicing factors in HCC.** Iris Graph representing the expression profile of splicing factors for (A) HCV-, (B) HBV&HCV-, and (C) virus-free HCC. Differences in gene expression levels are shown on a logarithmic color scale (Log<sub>2</sub>), from red (negative changes in expression) to blue (increase in gene expression). The expression of proteins involved in splicing modulated by more than 2-fold is indicated by an asterisk.



## REFERENCES

1. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res* 2014 ;74:2913-21.
2. McGlynn KA, Londonb WT. The Global Epidemiology of Hepatocellular Carcinoma, Present and Future. *Clin Liver Dis* 2011;15:223–244.
3. El-Serag HB. Epidemiology of Viral Hepatitis and Hepatocellular Carcinoma. *Gastroenterology* 2012;142:1264–1273.
4. Tsai WL, Chung RT. Viral hepatocarcinogenesis. *Oncogene* 2010;29 :2309–2324;
5. Hamed MA, Ali SA. Non-viral factors contributing to hepatocellular carcinoma. *World J Hepatol* 2013;5:311–322.
6. Jeong SW, Jang JY, Chung RT. Hepatitis C virus and hepatocarcinogenesis. *Clin Mol Hepatol* 2012;18:347–356.
7. Fung J, Lai CL, Yuen MF. Hepatitis B and C virus-related carcinogenesis. *Clin Microbiol Infect.* 2009;15:964-70.
8. Chisari FV, Isogawa M, Wieland SF. Pathogenesis of hepatitis B virus infection. *Pathol Biol (Paris)* 2010;58:258-66.
9. Rogler CE, Chisari FV. Cellular and molecular mechanisms of hepatocarcinogenesis. *Sem Liver Dis* 1992;12:265–278.

10. Murakami Y, Saigo K, Takashima H, Minami M, Okanoué T, Bréchet C, Paterlini-Bréchet P. Large scaled analysis of hepatitis B virus (HBV) DNA integration in HBV related hepatocellular carcinomas. *Gut* 2005;54:1162–1168.
11. Jiang Z, Jhunjhunwala S, Liu J, Haverty PM, Kennemer MI, Guan Y, Lee W, Carnevali P, Stinson J, Johnson S, Diao J, Yeung S, Jubb A, Ye W, Wu TD, Kapadia SB, de Sauvage FJ, Gentleman RC, Stern HM, Seshagiri S, Pant KP, Modrusan Z, Ballinger DG, Zhang Z. The effects of hepatitis B virus integration into the genomes of hepatocellular carcinoma patients. *Genome Res* 2012;22:593-601.
12. Bralet MP, Régimbeau JM, Pineau P, Dubois S, Loas G, Degos F, Valla D, Belghiti J, Degott C, Terris B. Hepatocellular carcinoma occurring in nonfibrotic liver: epidemiologic and histopathologic analysis of 80 French cases. *Hepatology* 2000;32:200–204.
13. Grando-Lemaire V, Guettier C, Chevret S, Beaugrand M, Trinchet JC. Hepatocellular carcinoma without cirrhosis in the West: epidemiological factors and histopathology of the non-tumorous liver. *J Hepatol* 1999;31:508–513.
14. Heindryckx F, Colle I, Vlierberghe HV. Experimental mouse models for hepatocellular carcinoma research. *Int J Exp Pathol* 2009; 90:367–386.
15. Iizuka N, Oka M, Yamada-Okabe H, Mori N, Tamesa T, Okada T, Takemoto N, Tangoku A, Hamada K, Nakayama H, Miyamoto T, Uchimura S, Hamamoto Y. Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data on the basis of a supervised learning method. *Cancer Res* 2002;62:3939-44.
16. Lee SY, Song KH, Koo I, Lee KH, Suh KS, Kim BY. Comparison of pathways associated with hepatitis B- and C-infected hepatocellular carcinoma using pathway-based class discrimination method. *Genomics* 2012;99:347-54.

17. Berasain C, Goñi S, Castillo J, Latasa MU, Prieto J, Ávila MA. Impairment of pre-mRNA splicing in liver disease: Mechanisms and consequences. *World J Gastroenterol* 2010;16:3091–3102.
18. Black DL. Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 2003;72:291-336.
19. Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. Alternative isoform regulation in human tissue transcriptomes. *Nature* 2008;456:470-476.
20. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 2008;40:1413-1415.
21. Boise LH, González-García M, Postema CE, Ding L, Lindsten T, Turka LA, Mao X, Nuñez G, Thompson CB. Bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. *Cell* 1993;74:597-608.
22. Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. *Genes Dev.* 2003; 17: 419-437.
23. Biamonti G, Catillo M, Pignataro D, Montecucco A, Ghigna C. The alternative splicing side of cancer. *Semin Cell Dev Biol.* 2014; 32: 30-36.
24. Germann S, Gratadou L, Dutertre M, Auboeuf D. Splicing programs and cancer. *J Nucleic Acids* 2012;269570.
25. Prudencio M, Belzil VV, Batra R, Ross CA, Gendron TF, Pregent LJ, Murray ME, Overstreet KK, Piazza-Johnston AE, Desaro P, et al. Distinct brain transcriptome profiles in C9orf72-associated and sporadic ALS. *Nat Neurosci* 2015;18:1175-1182.

26. Bonomi S, Gallo S, Catillo M, Pignataro D, Biamonti G, Ghigna C. Oncogenic alternative splicing switches: role in cancer progression and prospects for therapy. *Int J Cell Biol* 2013;2013:962038.
27. Amin EM, Oltean S, Hua J, Gammons MVR, Hamdollah-Zadeh M, Welsh GI, Cheung MK, Ni L, Kase S, Rennel ES, et al. WT1 mutants reveal SRPK1 to be a downstream angiogenesis target by altering VEGF splicing. *Cancer Cell*.2011;20:768–780.
28. Bechara EG, Sebestyén E, Bernardis I, Eyras E, Valcárcel J. RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Mol Cell* 2013;52:720–733.
29. Venables JP, Brosseau JP, Gadea G, Klinck R, Prinos P, Beaulieu JF, Lapointe E, Durand M, Thibault P, Tremblay K, et al. RBFOX2 is an important regulator of mesenchymal tissue-specific splicing in both normal and cancer tissues. *Mol Cell Biol* 2013;33:396–405
30. Izquierdo JM, Majós N, Bonnal S, Martínez C, Castelo R, Guigó R, Bilbao D, Valcárcel J. Regulation of Fas alternative splicing by antagonistic effects of TIA-1 and PTB on exon definition. *Mol Cell* 2005;19:475–484.
31. Sebestyén E, Zawisza M, Eyras E. Detection of recurrent alternative splicing switches in tumor samples reveals novel signatures of cancer. *Nucleic Acids Res* 2015;43:1345-56.
32. Pettigrew CA, Brown MA. Pre-mRNA splicing aberrations and cancer. *Front Biosci* 2008;13:1090–1105.
33. Körner M, Miller LJ. Alternative splicing of pre-mRNA in cancer: focus on G protein-coupled peptide hormone receptors. *Am J Pathol* 2009;175:461–472.
34. Oltean S, Bates DO. Hallmarks of alternative splicing in cancer. *Oncogene* 2014;33:5311-5318.

35. Venables JP, Klinck R, Koh C, Gervais-Bird J, Bramard A, Inkel L, Durand M, Couture S, Froehlich U, Lapointe E, et al. Cancer-associated regulation of alternative splicing. *Nat Struct Mol Biol* 2009;16:670-676.
36. Venables JP, Klinck R, Bramard A, Inkel L, Dufresne-Martin G, Koh C, Gervais-Bird J, Lapointe E, Froehlich U, Durand M, et al. Identification of alternative splicing markers for breast cancer. *Cancer Res* 2008;68:9525-9231.
37. Klinck R, Bramard A, Inkel L, Dufresne-Martin G, Gervais-Bird J, Madden R, Paquet ER, Koh C, Venables JP, Prinos P, et al. Multiple alternative splicing markers for ovarian cancer. *Cancer Res* 2008;68:657-663.
38. Brosseau JP, Lucier JF, Lamarche AA, Shkreta L, Gendron D, Lapointe E, Thibault P, Paquet E, Perreault JP, Abou Elela S, Chabot B. Redirecting splicing with bifunctional oligonucleotides. *Nucleic Acids Res* 2014;42:e40.
39. Alló M, Buggiano V, Fededa JP, Petrillo E, Schor I, de la Mata M, Agirre E, Plass M, Eyras E, Elela, SA, et al. Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat Struct Mol Biol* 2009;16:717-724.
40. Saito Y, Kanai Y, Sakamoto M, Saito H, Ishii H, Hirohashi S. Overexpression of a splice variant of DNA methyltransferase 3b, DNMT3b4, associated with DNA hypomethylation on pericentromeric satellite regions during human hepatocarcinogenesis. *Proc Natl Acad Sci USA* 2002;99:10060–10065.
41. Yasen M, Mizushima H, Mogushi K, Obulhasim G, Miyaguchi K, Inoue K, Nakahara I, Ohta T, Aihara A, Tanaka S, et al. Expression of Aurora B and alternative variant forms in hepatocellular carcinoma and adjacent tissue. *Cancer Sci* 2009;100:472–480.
42. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nature Met* 2012;9:357-35

43. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 2011;12:323.
44. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA* 2003;100:9440-9445.
45. Jiao X, Sherman BT, Huang DW, Stephens R, Baseler MW, Lane HC, Lempicki RA. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics* 2012;28:1805–1806.
46. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 2015;43: D447–D452.
47. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res* 2012;40:W597-W603.
48. Corpet F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 1988;16:10881-10890.
49. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M. The Pfam protein families database. *Nucleic Acids Res* 2014;Database Issue 42:D222-D230.
50. Kosugi S, Hasebe M, Tomita M, Yanagawa H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci USA* 2009;106:10171–10176.

51. Liu Z, Hou J. Hepatitis B Virus (HBV) and Hepatitis C Virus (HCV) Dual Infection. *Int J Med Sci* 2006;3(2):57–62.
52. Grosso AR, Martins S, Carmo-Fonseca M. The emerging role of splicing factors in cancer. *EMBO Rep* 2008;9:1087-1093.
53. Fackenthal JD, Godley LA. Aberrant RNA splicing and its functional consequences in cancer cells. *Dis Model Mech* 2008;1:37-42.
54. Warzecha CC, Shen S, Xing Y, Carstens RP. The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. *RNA Biol* 2009;6:546-562.
55. Zhao YJ, Ju Q, Li GC. Tumor markers for hepatocellular carcinoma. *Mol Clin Oncol* 2013;1:593–598.
56. Shirakawa H, Kuronuma T, Nishimura Y, Hasebe T, Nakano M, Gotohda N, Takahashi S, Nakagohri T, Konishi M, Kobayashi N, Kinoshita T, Nakatsura T. Glypican-3 is a useful diagnostic marker for a component of hepatocellular carcinoma in human liver cancer. *Int J Oncol* 2009;34:649-56.
57. Dong ZZ, Yao DF, Yao M, Qiu LW, Zong L, Wu W, Wu XH, Yao DB, Meng XY. Clinical impact of plasma TGF-beta1 and circulating TGF-beta1 mRNA in diagnosis of hepatocellular carcinoma. *Hepatobiliary Pancreat Dis Int* 2008;7:288-95.
58. Garber K. Energy deregulation: Licensing tumors to grow. *Science* 2006;312:1158–1159.
59. Jones RG, Thompson CB. Tumor suppressors and cell metabolism: a recipe for cancer growth. *Genes Dev* 2009;23:537-48.
60. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H,

Siu IM, Gallia GL, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science* 2008;321:1807–1812

61. Lee CH, Wu DC, Wu IC, Goan YG, Lee JM, Chou SH, Chan TF, Huang HL, Hung YH, Huang MC, Lai TC, Wang TN, Lan CC, Tsai S, Lin WY, Wu MT. Genetic modulation of ADH1B and ALDH2 polymorphisms with regard to alcohol and tobacco consumption for younger aged esophageal squamous cell carcinoma diagnosis. *Int J Cancer* 2009;125:1134-42.

62. Mamtani M, Kulkarni H. Association of HADHA expression with the risk of breast cancer: targeted subset analysis and meta-analysis of microarray data. *BMC Res Notes* 2012;5:25.

63. Anczuków O, Rosenberg AZ, Akerman, M, Das S, Zhan L, Karni R, Muthuswamy SK, Krainer AR. The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nat Struct Mol Biol* 2012;19:220-228.

64. Karni R, de Stanchina E, Lowe SW, Sinha R, Mu D, Krainer AR. The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat Struct Mol Biol* 2007;14:185-193.

65. LeFave CV, Squatrito M, Vorlova S, Rocco GL, Brennan CW, Holland EC, Pan YX, Cartegni L. Splicing factor hnRNPH drives an oncogenic splicing switch in gliomas. *EMBO J* 2011;30:4084-4097.

66. Wang Y, Chen D, Qian H, Tsai YS, Shao S, Liu Q, Dominguez D, Wang Z. The splicing factor RBM4 controls apoptosis, proliferation, and migration to suppress tumor progression. *Cancer Cell* 2014;26:374-389.

67. Bauman JA, Kole R. Modulation of RNA splicing as a potential treatment for cancer. *Bioeng Bugs* 2011;2:125-128.



## DISCUSSION

Les travaux présentés dans ce mémoire constituent les premières preuves concernant l'implication des virus dans les modifications du mécanisme d'épissage alternatif. Le rôle de l'épissage alternatif dans l'apparition de cancers était déjà bien documenté, tout comme l'était l'implication des virus dans les processus tumorigéniques. Depuis plusieurs années, la communauté scientifique était au fait de l'implication de l'épissage alternatif dans le cancer, puisque certains gènes ont été identifiés bien spécifiquement comme ayant un épissage alternatif modifié dans certains cancers, tels que les cancers du sein, de l'ovaire et de la prostate. Par exemple, dans le cancer de la prostate, des acteurs importants dans l'apparition du cancer, comme le récepteur de l'androgène et le récepteur de facteur de croissance FGFR2 subissent des changements majeurs au niveau de l'épissage alternatif (Lapuk et al, 2014). Également, comme il a été mentionné plus tôt, plus d'une quarantaine de gènes subissent des modifications dans l'épissage alternatif dans les cas de cancers de l'ovaire et du sein, certains d'entre eux étant des suppresseurs de tumeurs importants comme NOTCH3 et DSC3 (Venables et al, 2008). Il n'est pas étonnant que les changements au niveau de l'épissage alternatif puissent avoir des répercussions à plusieurs niveaux dans les processus de survie et d'apoptose cellulaire, étant donné la grande proportion de gènes subissant l'épissage alternatif. En raison de leur nature de parasites intracellulaires, les virus ont intérêt à ce que la cellule ait un métabolisme actif, soit capable de survivre à l'apoptose et aussi qu'elle soit en mesure d'échapper à la surveillance du système immunitaire afin de pouvoir se répliquer le plus efficacement possible. Pour y arriver, les virus ont développé différentes stratégies afin de moduler ces différents processus cellulaires. Par exemple, la protéine E6 du virus du papillome humain est capable d'inhiber l'apoptose par l'inhibition du suppresseur de tumeur p53 (Mantovani et Banks, 2001). Aussi, la protéase virale NS3/NS4A du virus de l'hépatite C aide à l'échappement du virus au système immunitaire en clivant des protéines impliquées au sein des voies de signalisation cellulaire responsables de l'activation d'une réponse immunitaire (clivage et inhibition des effecteurs TRIF et TRAM par exemple) (Jin et al, 2007; Imran et al, 2012). Ce faisant, le VHC pourra se répliquer dans la cellule en échappant à la surveillance de l'hôte. Par contre, il arrive que la modification de ces facteurs cellulaires mène à la transformation de la cellule en cellule cancéreuse.

L'article présenté dans ce mémoire démontre que l'épissage alternatif est modifié dans le cancer du foie, autant lorsque les tissus sont exempts de virus, mais de manière encore plus importante lorsque les carcinomes découlent d'une exposition aux virus de l'hépatite B ou C. En effet, comme nous l'avons vu plus tôt, ces deux virus sont capables d'engendrer des tumeurs au niveau du foie, suite à une infection chronique. Certains de ces cancers apparaîtraient de manière indirecte, c'est-à-dire que les virus, en infectant chroniquement le tissu, engendraient une inflammation prolongée, et un remplacement plus fréquent des cellules hépatiques, ce qui favoriserait la transformation du tissu (Morales-Sánchez et Fuentes-Pananá, 2014). Par contre, comme nous l'avons démontré, les virus peuvent transformer directement le tissu. Lors de mon projet de recherche, j'ai pu identifier des modifications au niveau de l'épissage alternatif de gènes importants dans l'apparition de cancer, comme des gènes suppresseurs de tumeur, des oncogènes et différentes autres cibles comme les kinases et les facteurs de transcription, lesquels ont aussi une grande importance dans les cancers. L'étude présentée dans l'article qui constitue ce mémoire est une étude de séquençage à haut débit sur des tumeurs provenant de différents patients, atteints à différents stades de la maladie, et porteurs ou non d'un des deux virus à hépatite. Les résultats engendrés sont d'une grande importance, car ils constituent la première preuve concernant l'implication des oncovirus dans les modifications des événements d'épissage alternatif dans un tissu cancéreux. Par contre, nous nous sommes également intéressés à en connaître davantage sur l'origine des modifications. La seconde partie de mon projet de maîtrise traite donc des modifications des événements d'épissage alternatif lorsque l'on exprime un oncogène viral dans des cellules eucaryotes, de manière très ponctuelle.

### **3.1 MODIFICATION DES ÉVÉNEMENTS D'ÉPISSAGE ALTERNATIF PAR UN ONCOGÈNE VIRAL**

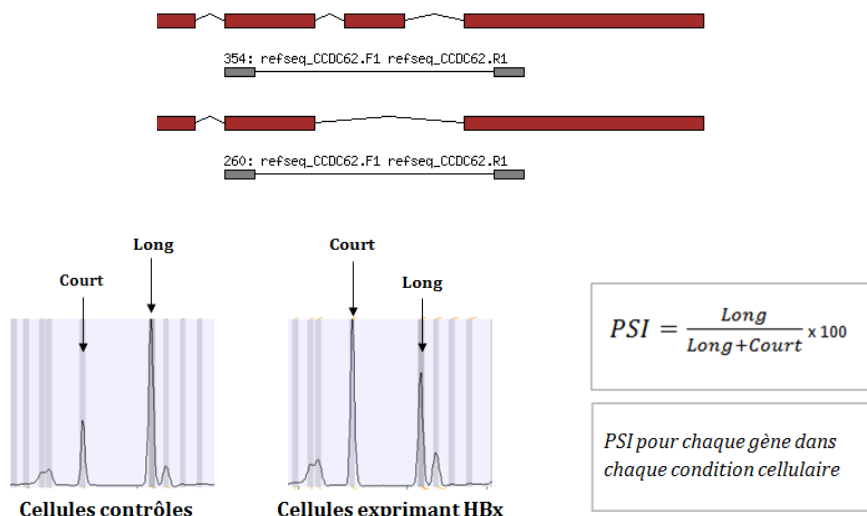
Dans le cadre de l'étude présentée ici, l'oncogène viral qui a été choisi est la protéine HBx du virus de l'hépatite B. En effet, certains mécanismes sont connus quant à la manière qu'a le virus de l'hépatite B, par l'entremise de sa protéine HBx, de transformer une cellule hépatique saine en cellule cancéreuse formant un carcinome hépatocellulaire. Plusieurs fonctions ont déjà été attribuées à la protéine HBx, entre autre l'inhibition de suppresseurs de

tumeur, ainsi que l'activation des voies de survie cellulaire. De plus, puisque HBx peut également moduler l'expression de certains gènes de par son rôle d'activateur de la transcription, son implication au niveau de la carcinogénèse est déjà bien connue. Par contre, aucun groupe d'étude jusqu'à présent n'a pu identifier précisément comment la protéine peut mener au cancer, ses effets étant étendus à plusieurs processus biologiques variés. Notre étude, en attribuant à HBx un pouvoir de modification de l'épissage alternatif, permet de décrire comment la protéine peut affecter autant de processus menant au cancer, puisque comme il a été mentionné, l'épissage alternatif est responsable de la variété protéique au sein d'une cellule. Ce faisant, HBx pourrait étendre ces effets à plusieurs processus, en altérant un plus grand processus qui en contrôle plusieurs.

### ***3.1.1 Identification de modifications d'épissage alternatif par la protéine HBx***

Outre les modifications dans l'épissage alternatif observées dans les tissus cancéreux infectés par les virus de l'hépatite B et C, des résultats préliminaires nous montrent qu'une protéine virale, la protéine HBx du virus de l'hépatite B, pourrait réguler l'épissage alternatif. En effet, lors de mes études graduées, j'ai également évalué le potentiel de la protéine HBx à moduler l'épissage alternatif de certains gènes impliqués dans le cancer, démontrant avec plus de confiance que c'est l'aspect viral en soit qui est responsable de la modification de l'épissage alternatif et non que ces effets sont attribuables simplement à la présence du phénotype cancéreux.

Pour cette étude, j'ai travaillé en collaboration avec la plateforme de génomique fonctionnelle de l'Université de Sherbrooke afin de quantifier les différences au niveau des patrons d'épissage alternatif de 1200 gènes impliqués dans le cancer, dans des cellules eucaryotes qui expriment la protéine HBx du virus de l'hépatite B.

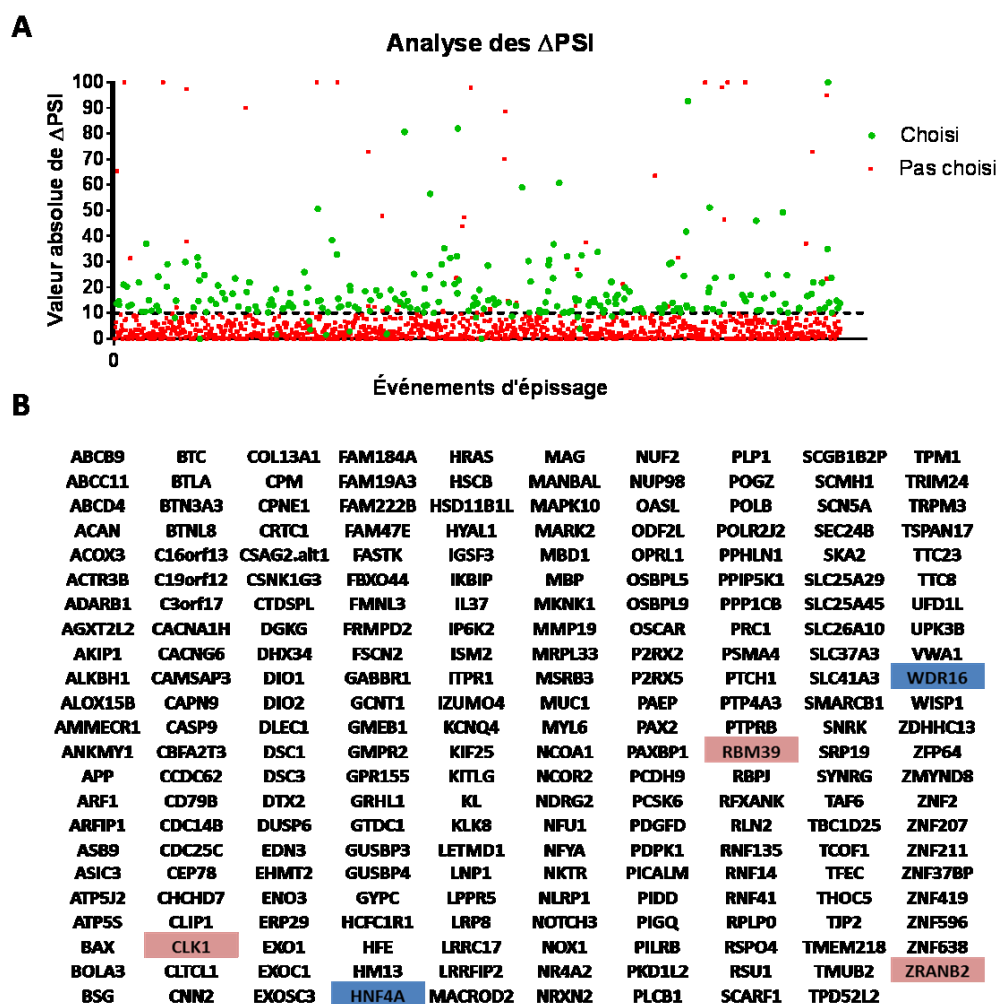


**Figure 14 – Méthodologie pour l’analyse des modifications d’épissage alternatif**

La protéine HBx a été exprimée dans des cellules HEK293T pendant 5 jours suite à une transfection transitoire du plasmide pLenti6V5A-HBx. Les cellules ont été récoltées et l’ARN a été extrait à l’aide du trizol. La plateforme LISA (Layered and Integrated system for Splicing isoform Annotation) du laboratoire de génomique fonctionnelle de l’Université de Sherbrooke amplifie spécifiquement (par RT-PCR) la région d’épissage, à l’aide d’amorces situées dans la jonction exonique. Suite à l’amplification, on observe sur les électrophorégrammes la présence des isoformes long et court, et on calcule le ratio entre ces isoformes à l’aide d’une valeur appelée PSI (percent splicing index) qu’on associe à chaque événement d’épissage. Lorsque l’on veut comparer deux conditions entre elles, dans le cas qui nous intéresse les cellules exprimant HBx et les cellules non-transfectées, il suffit de faire la différence entre les valeurs de PSI des deux conditions, ce qui nous fournit un delta PSI. Le delta PSI est l’indice qui nous permet de savoir si l’événement d’épissage alternatif a changé de manière significative entre les deux conditions. Lors de l’étude, nous considérons qu’une valeur de delta PSI qui est supérieure à dix représente un changement au niveau de l’épissage alternatif.

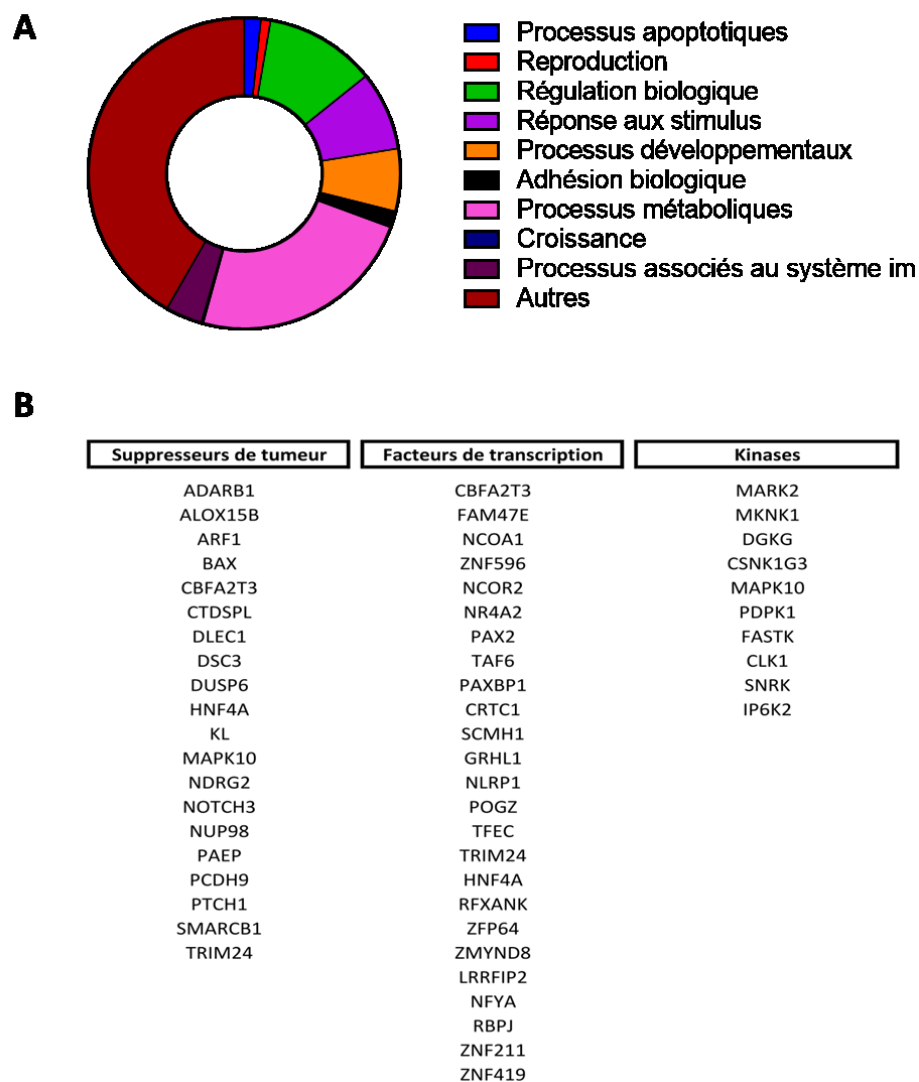
De cette façon, il nous a été possible d’identifier 229 gènes (240 événements) dont l’épissage alternatif des transcrits est modifié par la présence de la protéine HBx. De ceux-ci, certains sont connus comme étant impliqués dans le cancer du foie, comme les gènes WDR16 dont l’activité est augmentée dans les carcinomes hépatocellulaires et le facteur HNF4 $\alpha$ , l’expression de ce dernier étant réduite dans les cancers du foie (représentés en bleu sur la figure 15). Il est également intéressant de constater que la liste comprend quelques protéines impliquées dans l’épissage alternatif (en rouge sur la figure 15), et donc on peut penser que si l’épissage alternatif est modifiée pour ces facteurs, ceci aura un impact

important sur l'épissage de transcrits qui dépendent de ces facteurs. De cette liste de gènes, on retrouve également plusieurs suppresseurs de tumeurs (20) dont l'épissage alternatif est modifié en présence de la protéine HBx, ainsi que des facteurs de transcription (25) et des kinases (10), tout comme dans l'analyse présentée dans l'article (figure 16).



**Figure 15 – Candidats obtenus pour l'épissage alternatif modifié lorsque la protéine HBx est exprimée en cellules eucaryotes**

**A** Représentation des deltas PSI de chaque événement d'épissage alternatif analysé. Certains faux positifs apparaissent (les points rouges au-dessus de la ligne située à 10% de variation) ainsi que des faux négatifs (points verts situés sous la barre). Au total, 229 gènes ont au moins un événement d'épissage alternatif modifié en présence de la protéine HBx du VHB. **B** Liste de ces candidats. En bleu, il est représenté deux acteurs importants dans l'apparition des carcinomes hépatocellulaires et en rouge, trois protéines impliquées dans l'épissage dont l'épissage alternatif des transcrits a été modifié en présence de HBx.



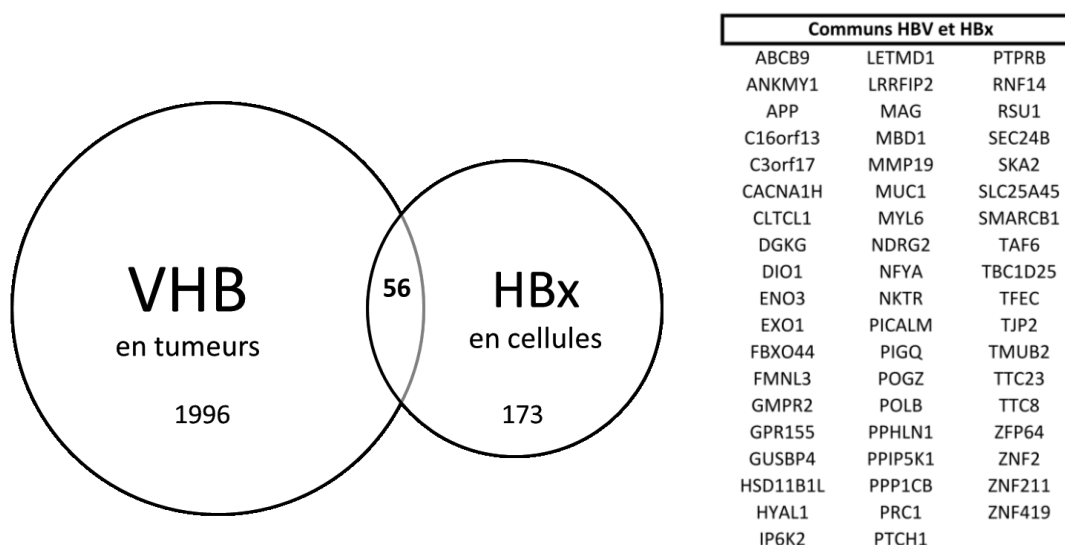
**Figure 16 – Distribution des candidats au sein de différents processus biologiques impliqués dans l'apparition de cancers**

**A** Ontologie des 229 gènes avec épissage alternatif modifié par HBx selon les processus biologiques. L'analyse ontologique est effectuée à l'aide du site web Panther, lequel se sert de la banque de données DAVID pour l'annotation des catégories d'ontologie. **B** Représentation de certaines classes de protéines importantes dans l'apparition de cancer, toujours selon la classification effectuée par Panther.

### 3.1.2 Comparaison avec les résultats obtenus pour les tissus positifs pour le VHB

Comme nous l'avons vu dans l'article, lorsque l'on compare les événements d'épissage alternatif entre des tumeurs positives pour le virus de l'hépatite B et des tissus sains, on

observe plusieurs changements. Il serait donc intéressant de voir si certains de ces changements sont également présents lorsque l'on exprime HBx dans les cellules eucaryotes. Bien entendu, le modèle cellulaire d'expression de la protéine HBx seule dans des cellules d'origine rénale n'est pas parfait, toutefois les changements observés dans ces cellules représentent globalement ce qui pourrait se produire lorsque HBx est présent dans les cellules hépatiques.



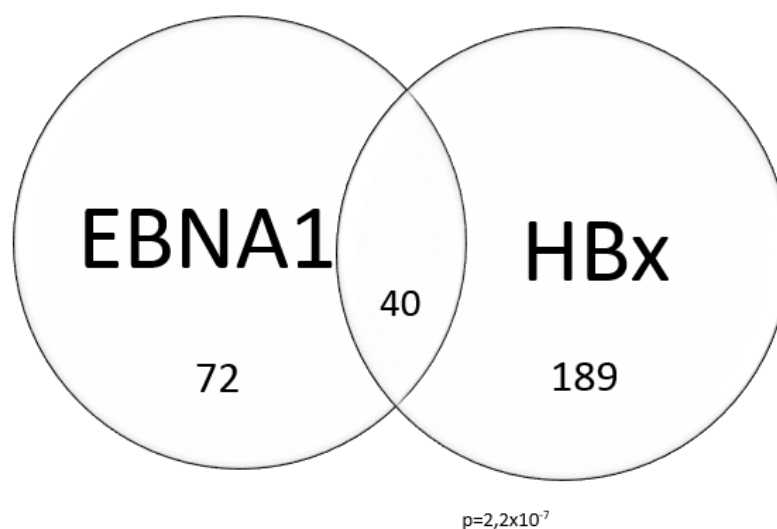
**Figure 17 – Comparaison avec les résultats obtenus dans les tissus cancéreux positifs pour le virus de l'hépatite B**

Des deux différentes analyses, il ressort que 56 modifications d'épissage alternatif seraient communes aux tumeurs positives pour le virus de l'hépatite B et aux cellules eucaryotes exprimant une protéine du VHB, l'oncoprotéine HBx. Toutefois, lorsque l'on effectue un Fisher's exact test (en prenant en compte seulement les gènes qui ont été analysés par PCR lors de l'analyse des modifications de HBx en cellules), on remarque que le chevauchement des données est non-significatif ( $p = 0,12$ ). Il est tout de même important de noter que ces résultats sont tirés d'une analyse faite en simplicité (les analyses de PCR sur les modifications d'épissage alternatif engendrées par la présence de la protéine HBx n'ayant été effectuées qu'une seule fois).

### 3.1.3 Comparaison avec une autre protéine virale: EBNA1

Comme il y a six autres virus qui causent le cancer chez l'humain, il serait pertinent de savoir si d'autres virus peuvent également modifier l'épissage alternatif. Une autre étudiante du laboratoire a identifié que la protéine EBNA1 du virus Epstein-Barr était également en

mesure de modifier l'épissage alternatif de certains pré-ARNm (112) de cette même liste de 1200 gènes. En comparant ces deux listes de gènes dont l'épissage alternatif des transcrits a été modifié par une protéine virale, il est possible de constater que 39 de ceux-ci sont communs à la présence des protéines EBNA1 et HBx en cellule. Ceci laisse présager que certaines modifications au niveau de l'épissage alternatif pourraient être des voies communes. Toutefois, il est intéressant de constater que certains changements dans les événements d'épissage alternatif restent spécifiques à l'expression de chacune des protéines virales. Ceci laisse croire que certains changements seraient spécifiques à une condition donnée, dans le cas présent à la présence d'une protéine virale spécifique.



**Figure 18 – Comparaison avec la protéine EBNA1 du virus epstein-barr**

Le diagramme de venn représente le fait qu'il y ait 39 candidats dont l'épissage alternatif est modifié autant dans les cellules eucaryotes exprimant EBNA1 que dans les cellules exprimant HBx. Les résultats sur les changements d'épissage engendrés par la présence de la protéine EBNA1 ont été tirés de la même façon que ceux obtenus pour la protéine virale HBx, soit à l'aide d'analyses par RT-PCR effectuées par la plateforme LISA. Toutefois, les cellules utilisées expriment de manière permanente la protéine EBNA1 à partir d'un vecteur différent (pMSCV), comparativement à une expression transitoire pour les cellules exprimant le vecteur pLenti6V5A-HBx.



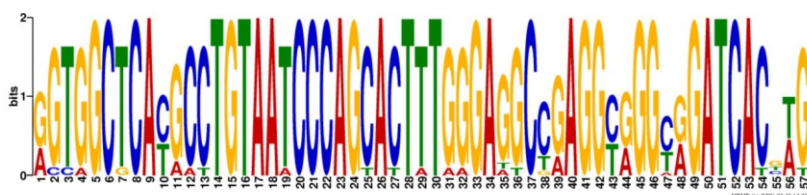
**Tableau 1 – Modifications communes entre les cellules exprimant EBNA1 et celles exprimant HBx et l'isoforme favorisée**

Gènes	EBNA1		HBx	
	$\Delta$ PSI (valeur absolue)	Isoforme favorisé	$\Delta$ PSI (valeur absolue)	Isoforme favorisé
ALOX15B	76,1	Long	37	Long
ANKMY1	13,7	Court	10,9	Court
C16ORF13	17,7	Court	24,8	Long
CAPN9	19	Long	15,5	Court
CASP9	6,8	Court	17,6	Long
CCDC62	42,3	Court	12,6	Long
CLK1	12,5	Long	13,3	Long
COL13A1	17	Long	19,4	Court
DLEC1	53,8	Court	3,5	Long
DTX2	11	Court	15,1	Long
FAM47E	8,6	Court	18,9	Court
FSCN2	79,8	Court	15	Court
GCNT1	25,4	Long	16,8	Long
ITPR1	15,3	Long	21,5	Long
IZUMO4	12,2	Long	35,3	Long
KIF25	42	Long	21,4	Long
KL	28,8	Court	82	Long
LETMD1	20,1	Long	11	Court
LNP1	9,8	Long	13	Court
LPPR5	23,2	Court	16,1	Court
MAPK10	6,4	Long	28,6	Court
NFU1	12	Court	14,2	Long
NRXN2	11,8	Court	60,8	Court
ODF2L	13,1	Court	10,3	Long
OPRL1	30,7	Court	10,8	Long
OSBPL9	18,3	Court	14,8	Court
OSCAR	17,5	Court	8,5	Court
P2RX5	25,4	Long	24,9	Long
PAEP	69,3	Court	32,5	Court
PAX2	15,6	Court	13	Court
PAXBP1	15,4	Court	14,2	Court
PCDH9	12	Long	15,4	Court
PKD1L2	19	Long	14,5	Court
PRC1	15,4	Court	18,3	Long
PTP4A3	19,4	Court	11,8	Long
RNF135	24,7	Court	29,7	Long
RNL2	16,5	Court	29,2	Court
IL37	38,6	Court	10	Long
TTC23	12,7	Court	24,7	Court
ZRANB2	11,5	Court	13,8	Long

Les 39 modifications communes aux cellules exprimant EBNA1 et les cellules exprimant HBx sont présentées ici, avec leur valeur absolue de delta PSI indiquant l'impact de la modification ainsi que l'information à savoir si c'est la forme longue ou courte du transcrit qui sera favorisée dans chacune des deux conditions.

### 3.1.5 Recherche de séquence consensus

Des résultats préliminaires pourraient nous permettre de comprendre comment la présence de la protéine HBx dans les cellules eucaryotes a un effet sur l'épissage alternatif de ces 229 gènes. À l'aide du logiciel MEME, nous avons observé qu'une séquence consensus de 57 nucléotides était présente au niveau de la jonction exonique étudiée dans l'événement d'épissage.



**Figure 19 – Présence d'une séquence consensus de 57 nucléotides dans 129 événements d'épissage alternatif modifiés en présence de la protéine HBx**

Sur les 240 événements étudiés, 129 possèdent cette séquence consensus au niveau de la jonction exon-exon. Le logiciel MEME (version 4.10.1) a été utilisé pour déterminer une séquence consensus au niveau de la région intron-exon-intron de chacun des candidats. Le logiciel a été programmé afin d'obtenir un motif d'une longueur maximale de 75 nucléotides ainsi qu'un maximum de 5 séquences possibles dans le but d'obtenir seulement des séquences pertinentes.

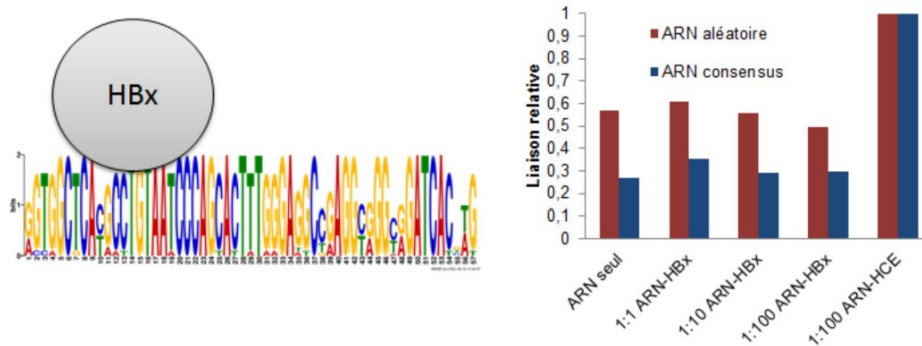
**Tableau 2 – Présence de la séquence consensus de 57 nucléotides dans 121 gènes dont l'épissage alternatif est modifié en présence de la protéine HBx**

Événements d'épissage avec la séquence consensus			
ABC89	CTDSPL	MSRB3	SCARF1
ABCD4	DIO1	NCOA1	SCMH1
ACAN	DTX2	NFU1	SCN5A
ACOX3	ERP29	NFYA	SEC24B
ACTR3B	FAM184A	NKTR	SKA2
AKIP1	FAM222B	NLRP1	SLC25A29
ALKBH1	FAM47E	NOTCH3	SLC25A45
AMMERC1	FMNL3	NRXN2	SLC37A3
APP	FRMPD2	NUP98	SNRK
ARF1	GCNT1	OASL	SRP19
ARFIP1	GMEB1	ODF2L	TAF6
ASB9	GPR155	OSBPL9	TBC1D25
ATP5J2	GTDC1	OSCAR	TCOF1
BOLA3	GUSBP4	PAXBP1	THOC5
BSG	HFE	PCDH9	TMUB2
BTC	HSCB	PDGFD	TPM1
BTLA	HSD11B1L	PDPK1	TRIM24
C19orf12	IKBIP	PICALM	TRPM3
C3orf17	ISM2	PKD1L2	TTC23
CACNG6	ITPR1	PLCB1	TTC8
CAPN9	KITLG	POGZ	WDR16
CASP9	KL	POLB	WISP1
CCDC62	LNP1	POLR2J2	ZDHHC13
CDC14B	LPPR5	PPHLN1	ZFP64
CDC25C	LRRFIP2	PPP1CB	ZMYND8
CLIP1	MACROD2	PTPRB	ZNF2
CLTCL1	MAG	RBM39	ZNF596
CNN2	MANBAL	RBPJ	ZNF638
COL13A1	MAPK10	RNF135	
CPNE1	MBP	RNF41	
CSNK1G3	MMP19	RSU1	

Les 121 gènes sont listés ci-haut. Il est à noter que la séquence se retrouve à plusieurs reprises dans différents événements d'épissage au travers de ces 121 gènes, augmentant le nombre d'événements d'épissage qui comporte cette séquence à 129.

La découverte de cette séquence consensus nous a amené à nous poser quelques questions, à savoir premièrement si cette séquence de 57 nucléotides avait été répertoriée dans la littérature. En faisant quelques recherches, nous avons découvert que celle-ci semble être une séquence appartenant à la famille des petits éléments nucléaires intercalés (SINE). Plus précisément, cette séquence consensus de 57 nucléotides ferait partie des éléments Alu, lesquels sont des séquences d'environ 300 nucléotides. Les éléments Alu sont très présents dans le génome humain; ils représentent 11% du génome. Malgré cette présence élevée, ces séquences semblent être enrichies au sein de nos candidats, puisqu'elle est retrouvée dans plus de la moitié de ceux-ci. Ainsi, il serait intéressant de savoir si cette séquence est impliquée dans les modifications d'épissage alternatif observées. Nous avons donc

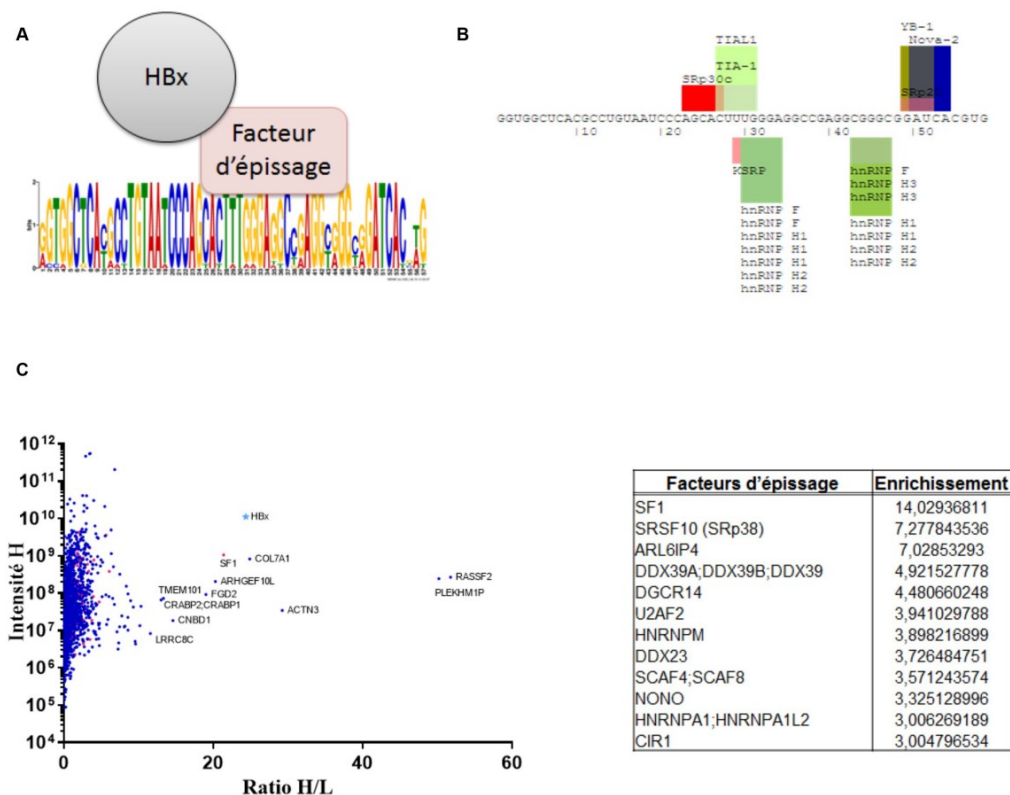
investigué l'implication potentielle de cette séquence dans l'altération des événements d'épissage alternatif induite en présence de la protéine virale HBx. Si cette séquence est impliquée, nous avons pensé qu'il y avait deux moyens possibles. D'abord, il serait raisonnable de penser que la protéine HBx peut lier directement cette séquence nucléotidique, et que cette liaison directe de HBx à l'ARN soit responsable de la modification dans le statut d'épissage alternatif : HBx pourrait inhiber ou activer l'épissage de manière directe, ou encore en empêchant le recrutement de certains facteurs d'épissage par compétition. Pour vérifier cette première hypothèse présentée à la figure 20, l'expérience à réaliser consistait en un essai de retardement sur gel. Les essais de retardement sur gel que j'ai effectué entre HBx et l'ARN ont démontré que la protéine virale ne lie pas l'ARN de séquence consensus préférentiellement à un ARN de 57 nucléotides de séquence aléatoire (voir figure 20). Ce résultat préliminaire nous démontre que l'effet de HBx sur l'épissage alternatif ne semble pas être lié à l'attachement direct de la protéine sur cette séquence d'ARN consensus.



**Figure 20 – Possibilité de liaison de HBx à la séquence consensus**

Un essai de retardement sur gel de HBx en présence d'un ARN radiomarqué correspondant à la séquence consensus a permis de démontrer que HBx ne lie pas cette séquence consensus préférentiellement à une séquence d'ARN de 57 nucléotides aléatoire.

Afin d'expliquer la grande présence de cette séquence sur nos candidats dont l'épissage alternatif a été modifié, la seconde option à laquelle nous avons pensé est que si, comme nous l'avons démontré, HBx ne lie pas cette séquence, la protéine virale pourrait toutefois lier des facteurs d'épissage, lesquels peuvent attacher cette séquence. De cette façon, les facteurs d'épissage, puisque liés par HBx, auraient plus ou moins d'affinité pour l'ARN, et donc auraient des capacités à épisser l'ARN différentes de la normale. Pour que cette deuxième option puisse être valable, deux conditions doivent absolument être respectées. La première condition est que cette séquence consensus doit pouvoir être liée par des facteurs d'épissage. Afin de valider ceci, j'ai utilisé le logiciel en ligne SpliceAid2. Celui-ci recense plusieurs facteurs d'épissage et leur site de liaison à l'ARN. Ainsi, il nous a permis de voir que certains facteurs d'épissage sont bel et bien en mesure de reconnaître cette séquence consensus, en ayant des séquences de reconnaissance de quatre à cinq nucléotides de long sur cette grande séquence. La première condition étant respectée, il nous a fallu tester la seconde condition, laquelle est que la protéine virale HBx doit être capable de lier des facteurs d'épissage et donc des protéines cellulaires. Pour savoir si cette deuxième condition est respectée, j'ai procédé à un SILAC (stable isotope labelling by aminoacids in cell culture). L'immunoprécipitation de HBx suivi de l'analyse par spectrométrie de masse nous a permis d'identifier plusieurs protéines qui semblent interagir avec la protéine virale en cellule. Le panneau C de la figure 21 illustre les résultats obtenus pour cette expérience. Le graphique démontre que certaines protéines cellulaires sont plus enrichies que d'autres lors de l'analyse par spectrométrie de masse, et donc que ces protéines pourraient lier la protéine HBx. De plus, il est intéressant de constater que de celles-ci, plusieurs sont des protéines impliquées dans l'épissage alternatif. Ces facteurs d'épissage sont indiqués en rose sur le graphique, et le tableau à la droite de ce graphique présente ceux qui sont le plus présents.



**Figure 21 – Possibilité de liaison de la protéine HBx à des facteurs d'épissage qui lieraient la séquence consensus**

**A** Hypothèse selon laquelle HBx pourrait interagir avec des facteurs d'épissage cellulaires. Une fois liés par HBx, ces facteurs auraient plus ou moins d'affinité pour la séquence d'ARN consensus, qui pourrait être impliquée dans la reconnaissance des sites d'épissage. **B** Sites de liaison potentielle de certains facteurs d'épissage à la séquence consensus de 57 nucléotides évalués à l'aide du logiciel en ligne SpliceAid2 **C** Résultats de spectrométrie de masse sur les cellules exprimant HBx afin d'évaluer les partenaires d'interaction en cellule de la protéine virale. La spectrométrie de masse a été effectuée suite à un SILAC (stable isotope labelling by aminoacids in cell culture) sur trois conditions de cellules HEK293T : des cellules n'exprimant pas HBx et pour les deux autres conditions, des cellules qui expriment HBx depuis cinq jours. Avant de procéder à l'immunoprécipitation de HBx, un des échantillons cellulaires positifs pour HBx a été traité avec de la RNase et de la DNase afin d'éliminer les acides nucléiques, ce qui nous fournira comme information seulement les liaisons directes protéines-protéines. C'est la comparaison entre cette condition traitée à la RNase et à la DNase et des cellules non-transfectées qui est représentée sur ce graphique. La protéine HBx semble pouvoir lier certaines protéines cellulaires, dont quelques facteurs d'épissage représentés en rose dans le graphique et mis sous forme de tableau à droite.

À la lumière de ces résultats, on peut penser que la deuxième condition serait respectée. Par contre, les résultats de spectrométrie de masse devaient être validés. Nous avons choisi de valider la liaison potentielle de HBx avec certains facteurs d'épissage en reproduisant l'immunoprécipitation contre HBx, et en ciblant ces facteurs d'épissage par immunobuvardage de type western. Des anticorps spécifiques pour six facteurs d'épissage ont été utilisés sur les culots d'immunoprécipitation. Malheureusement, les résultats obtenus ne se sont pas avérés concluants. En effet, les facteurs d'épissage qui auraient dû être visibles sur l'immunobuvardage ne se sont pas avérés positifs, et ceux qui ne devaient pas apparaître se sont révélés positifs. Après plusieurs essais, je n'ai pas été en mesure de reproduire les résultats obtenus suite au SILAC. Il m'est donc finalement impossible de conclure sur la condition selon laquelle HBx devrait être capable de lier des facteurs d'épissage cellulaires. Donc, pour l'instant, nous devons mettre de côté l'hypothèse selon laquelle la séquence consensus servirait à lier des facteurs d'épissage qui eux lient HBx.

### **3.2 PERSPECTIVES**

Mes travaux de recherche m'ont permis de démontrer que les virus de l'hépatite C et B sont en mesure de modifier l'épissage alternatif au sein des cellules qu'ils infectent. En effet, autant les travaux présentés dans l'article qui constitue ce mémoire que les résultats présentés sur la protéine virale HBx du VHB nous ont permis d'observer des changements au niveau de l'épissage alternatif de certains gènes importants dans le développement du cancer.

Afin de pousser l'étude un peu plus loin, il faudra valider si les changements observés au niveau de l'ARNm se transposent aux protéines qui en sont traduites. En effet, ce sont les modifications au niveau des protéines qui auront réellement un impact sur les fonctions biologiques des différents effecteurs cellulaires. Pour y arriver, il faudra tout d'abord identifier des candidats pour lesquels les modifications au niveau de l'épissage retirent ou ajoutent un domaine fonctionnel important. Ce faisant, le phénotype de la cellule pourra être transformé, puisque les fonctions biologiques ne seront plus fonctionnelles comme à l'habitude.

Également, des études plus approfondies devront être menées afin de connaître avec précision comment les virus sont en mesure de modifier l'épissage alternatif. Les travaux présentés dans la discussion de ce mémoire nous ont permis de découvrir que des protéines virales isolées exprimées en cellule sont capables de modifier les événements d'épissage alternatif. Toutefois, aucun mécanisme précis n'a pu être identifié, malgré la découverte de la présence d'une séquence consensus.

Ultimement, les travaux de recherche présentés dans ce mémoire pourraient nous permettre de développer de nouveaux biomarqueurs pour le dépistage du cancer du foie induit par les virus de l'hépatite C et B. Aussi, en comprenant mieux les différences associées à chacun des virus, il serait peut-être possible éventuellement de prétendre à des traitements antinéoplasiques basés sur les découvertes engendrées dans cette étude.



## REMERCIEMENTS

Au cours de mon passage dans le laboratoire, j'ai côtoyé plusieurs personnes exceptionnelles, qui m'ont toutes aidée à leur façon. D'abord, je dois remercier mon directeur de recherche, Martin Bisailon pour m'avoir accordé suffisamment de confiance pour exploiter ce nouveau sujet de recherche qui lui tenait à cœur. Aussi, sa bonne humeur et ses bons conseils m'ont grandement aidée à évoluer en tant que personne et en tant que scientifique. Ce que j'ai appris au laboratoire me servira certainement dans ma future carrière en recherche. Également, je dois remercier mon mentor, Michelle Scott, qui a été une oreille attentive à mes besoins. Les deux assistants de recherche, Frédéric et Maude, ont été d'une grande aide tout au long du projet, plus particulièrement Frédéric qui m'a aidée à démarrer les travaux à mon arrivée au laboratoire. Je dois également remercier les autres étudiants à la maîtrise, plus spécifiquement Vicky avec qui j'ai eu à travailler en étroite collaboration, et qui est devenue une bonne amie. Merci à tous les stagiaires qui ont fait diverses allées et venues au laboratoire, en particulier Andréa qui m'a beaucoup aidée à faire avancer mon projet, et Cyntia pour les quelques expériences qu'elle a faites pour moi. Globalement, l'ambiance générale du laboratoire créé par tous les membres m'a permis d'être positive et motivée tout au long de mon parcours. Un gros merci également aux trois membres de la plateforme génomique de l'Université de Sherbrooke, Elvy, Mathieu et plus spécialement Philippe qui ont fourni un travail remarquable qui m'ont permis de générer l'ensemble des résultats présents dans l'article qui constitue ce mémoire.

Je remercie la Faculté de médecine et des sciences de la santé pour la généreuse bourse d'étude qui m'a été offerte.

Finalement, je remercie chaleureusement ma famille pour m'avoir épaulé tout au long de mon parcours scolaire. Le plus gros des mercis à mon amoureux, Jean-Philip, pour son écoute, sa patience, sa générosité et ses précieux conseils, en particulier lorsque aux moments des présentations orales mais aussi lors de la rédaction de textes en anglais.

## LISTE DES RÉFÉRENCES

Bartenschlager, R, and H Schaller. 1992. "Hepadnaviral Assembly Is Initiated by Polymerase Binding to the Encapsidation Signal in the Viral RNA Genome." *The EMBO Journal* 11 (9) (September 2): 3413–20.

Bayer, M E, B S Blumberg, and B Werner. 1968. "Particles Associated with Australia Antigen in the Sera of Patients with Leukaemia, Down's Syndrome and Hepatitis." *Nature* 218 (5146) (June 6): 1057–9.

Benn, J, and R J Schneider. 1995. "Hepatitis B Virus HBx Protein Deregulates Cell Cycle Checkpoint Controls." *Proceedings of the National Academy of Sciences of the United States of America* 92 (24) (November 2): 11215–9.

Bock, C T, P Schranz, C H Schröder, and H Zentgraf. 1994. "Hepatitis B Virus Genome Is Organized into Nucleosomes in the Nucleus of the Infected Cell." *Virus Genes* 8 (3) (July 5): 215–29.

Bock, C T, S Schwinn, S Locarnini, J Fyfe, M P Manns, C Trautwein, and H Zentgraf. 2001. "Structural Organization of the Hepatitis B Virus Minichromosome." *Journal of Molecular Biology* 307 (1) (March 5): 183–96.

Bruss, Volker. 2007. "Hepatitis B Virus Morphogenesis." *World Journal of Gastroenterology : WJG* 13 (1) (January): 65–73.

Canadian liver foundation. [<http://www.liver.ca/fr/liver-disease/types/cirrhosis/>] "Cirrhosis". Consulted online in 2015.

Canadian liver foundation. [<http://www.liver.ca/liver-disease/types/liver-cancer.aspx>] "Liver Cancer" . Consulted online in 2015

Cancer Research UK. 2014. [<http://www.cancerresearchuk.org/about-cancer/cancers-in-general/treatment/>] "Cancer treatments" Consulted online in 2015.

Cha, Man-Young Y, Chang-Myeong M Kim, Young-Min M Park, and Wang-Shick S Ryu. 2004. "Hepatitis B Virus X Protein Is Essential for the Activation of Wnt/beta-Catenin Signaling in Hepatoma Cells." *Hepatology (Baltimore, Md.)* 39 (6) (June 2): 1683–93.

Chang, C, G Enders, R Sprengel, N Peters, H E Varmus, and D Ganem. 1987. "Expression of the Precore Region of an Avian Hepatitis B Virus Is Not Required for Viral Replication." *Journal of Virology* 61 (10) (October 4): 3322–5.

Chang, C, S Zhou, D Ganem, and D N Standing. 1994. "Phenotypic Mixing between Different Hepadnavirus Nucleocapsid Proteins Reveals C Protein Dimerization to Be Cis Preferential." *Journal of Virology* 68 (8) (August 1): 5225–31.

Chen, H S, S Kaneko, R Girones, R W Anderson, W E Hornbuckle, B C Tennant, P J Cote, J L Gerin, R H Purcell, and R H Miller. 1993. "The Woodchuck Hepatitis Virus X Gene Is Important for Establishment of Virus Infection in Woodchucks." *Journal of Virology* 67 (3) (March 1): 1218–26.

Chen, H S, M C Kew, W E Hornbuckle, B C Tennant, P J Cote, J L Gerin, R H Purcell, and R H Miller. 1992. "The Precore Gene of the Woodchuck Hepatitis Virus Genome Is Not Essential for Viral Replication in the Natural Host." *Journal of Virology* 66 (9) (September 2): 5682–4.

Chen, Jun, and Aleem Siddiqui. 2007. "Hepatitis B Virus X Protein Stimulates the Mitochondrial Translocation of Raf-1 via Oxidative Stress." *Journal of Virology* 81 (12) (June 5): 6757–60.

Chlumsky, J, and V Krtek. 1982. "HBsAg in Ascitic Fluid of Patients with Hepatic Cirrhosis." *Hepato-Gastroenterology* 29 (3) (July 5): 106–7.

Chu, Chia-Ming M, Shao-Jen J Hung, Justin Lin, Dar-In I Tai, and Yun-Fan F Liaw. 2004. "Natural History of Hepatitis B E Antigen to Antibody Seroconversion in Patients with Normal Serum Aminotransferase Levels." *The American Journal of Medicine* 116 (12) (June 2): 829–34.

Churin, Yuri, Martin Roderfeld, and Elke Roeb. 2015. "Hepatitis B Virus Large Surface Protein: Function and Fame." *Hepatobiliary Surgery and Nutrition* 4 (1) (February): 1–10.

Coppolino, Giuseppe, Mariadelina Simeoni, Chiara Summari, Maria C Postorino, Laura Rivoli, Alessio Strazzulla, Carlo Torti, and Giorgio Fuiano. 2015. "The Case of Chronic Hepatitis B Treatment with Tenofovir: An Update for Nephrologists." *Journal of Nephrology* 28 (4) (August 6): 393–402.

Dane, D S, C H Cameron, and M Briggs. 1970. "Virus-like Particles in Serum of Patients with Australia-Antigen-Associated Hepatitis." *Lancet (London, England)* 1 (7649) (April 6): 695–8.

EDMONDSON, H A, and P E STEINER. 1954. "Primary Carcinoma of the Liver: A Study of 100 Cases among 48,900 Necropsies." *Cancer* 7 (3) (June): 462–503.

Francki, RIB, Fauquet, CM, Knudson, DL and Brown F. 1991. "Classification and nomenclature of viruses" *Springer-Verlag Wien. (1) 90-405*

Ganem, Don, and Alfred M Prince. 2004. "Hepatitis B Virus Infection--Natural History and Clinical Consequences." *The New England Journal of Medicine* 350 (11) (March 4): 1118–29.

Glebe, Dieter, and Corinna M Bremer. 2013. "The Molecular Virology of Hepatitis B Virus." *Seminars in Liver Disease* 33 (2) (May 3): 103–12.

Glebe, Dieter, and Stephan Urban. 2007. "Viral and Cellular Determinants Involved in Hepadnaviral Entry." *World Journal of Gastroenterology : WJG* 13 (1) (January): 22–38.

Hanahan, D, and R A Weinberg. 2000. "The Hallmarks of Cancer." *Cell* 100 (1) (January 5): 57–70.

Hirsch, R C, D D Loeb, J R Pollack, and D Ganem. 1991. "Cis-Acting Sequences Required for Encapsidation of Duck Hepatitis B Virus Pregenomic RNA." *Journal of Virology* 65 (6) (June 6): 3309–16.

Hu, Liang, Lei Chen, GuangZhen Yang, Liang Li, HanYong Sun, Yanxin Chang, QianQian Tu, MengChao Wu, and HongYang Wang. 2011. "HBx Sensitizes Cells to Oxidative Stress-Induced Apoptosis by Accelerating the Loss of Mcl-1 Protein via Caspase-3 Cascade." *Molecular Cancer* 10 (January 6): 43.

Kann, M, A Bischof, and W H Gerlich. 1997. "In Vitro Model for the Nuclear Transport of the Hepadnavirus Genome." *Journal of Virology* 71 (2) (February 6): 1310–6.

Karayiannis, P, D M Novick, A S Lok, M J Fowler, J Monjardino, and H C Thomas. 1985. "Hepatitis B Virus DNA in Saliva, Urine, and Seminal Fluid of Carriers of Hepatitis B E Antigen." *British Medical Journal (Clinical Research Ed.)* 290 (6485) (July 6): 1853–5.

Khattar, Ekta, Atish Mukherji, and Vijay Kumar. 2012. "Akt Augments the Oncogenic Potential of the HBx Protein of Hepatitis B Virus by Phosphorylation." *The FEBS Journal* 279 (7) (April): 1220–30.

Kim, Seong-Jun J, Mohsin Khan, Jun Quan, Andreas Till, Suresh Subramani, and Aleem Siddiqui. 2013. "Hepatitis B Virus Disrupts Mitochondrial Dynamics: Induces Fission and Mitophagy to Attenuate Apoptosis." *PLoS Pathogens* 9 (12) (January 2): e1003722.

Kumar, Ajay, Manisha Dwivedi, S P Misra, Sushil Narang, Balesh K Tiwari, and Renu Pandey. 2011. "Clinical Profile, Genotype and Management Updates of Hepatitis B Virus." *Indian Journal of Virology : An Official Organ of Indian Virological Society* 22 (1) (July): 1–10.

Lambert, Carsten, Tatjana Döring, and Reinhild Prange. 2007. "Hepatitis B Virus Maturation Is Sensitive to Functional Inhibition of ESCRT-III, Vps4, and Gamma 2-Adaptin." *Journal of Virology* 81 (17) (September 6): 9050–60.

Lee, Wei-Ping P, Keng-Hsin H Lan, Chung-Pin P Li, Yee Chao, Han-Chieh C Lin, and Shou-Dong D Lee. 2012. "Pro-Apoptotic or Anti-Apoptotic Property of X Protein of Hepatitis B Virus Is Determined by Phosphorylation at Ser31 by Akt." *Archives of Biochemistry and Biophysics* 528 (2) (December 6): 156–62.

Locarnini, Stephen, and Fabien Zoulim. 2010. "Molecular Genetics of HBV Infection." *Antiviral Therapy* 15 Suppl 3 (January 5): 3–14.

MacCALLUM, F O. 1946. "Homologous Serum Hepatitis." *Proceedings of the Royal Society of Medicine* 39 (September 3): 655–7.

Mason, W S, C Aldrich, J Summers, and J M Taylor. 1982. "Asymmetric Replication of Duck Hepatitis B Virus DNA in Liver Cells: Free Minus-Strand DNA." *Proceedings of the National Academy of Sciences of the United States of America* 79 (13) (July 4): 3997–4001.

Mast, E E, M J Alter, and H S Margolis. 1999. "Strategies to Prevent and Control Hepatitis B and C Virus Infections: A Global Perspective." *Vaccine* 17 (13-14) (March 5): 1730–3.

Medical Imaging & Technology Alliance. [<http://www.medicalimaging.org/>] Consulted online in 2015.

Milich, D R, M K Chen, J L Hughes, and J E Jones. 1998. "The Secreted Hepatitis B Precore Antigen Can Modulate the Immune Response to the Nucleocapsid: A Mechanism for Persistence." *Journal of Immunology (Baltimore, Md. : 1950)* 160 (4) (February): 2013–21.

Milich, D R, J E Jones, J L Hughes, J Price, A K Raney, and A McLachlan. 1990. "Is a Function of the Secreted Hepatitis B E Antigen to Induce Immunologic Tolerance in Utero?" *Proceedings of the National Academy of Sciences of the United States of America* 87 (17) (September 6): 6599–603.

Nassal, M, and H Schaller. 1993. "Hepatitis B Virus Replication." *Trends in Microbiology* 1 (6) (September 3): 221–8.

Nassal, M. 1996. "Hepatitis B Virus Replication--an Update." *Journal of Viral Hepatitis* 3 (5) (September): 217–26.

Nassal, Michael. 2008. "Hepatitis B Viruses: Reverse Transcription a Different Way." *Virus Research* 134 (1-2) (June): 235–49.

National Cancer Institute. [<http://imaging.cancer.gov/>] "Cancer Imaging Program" Consulted online in 2015.

National Institutes of Health, U.S. National Library of Medicine. 2014. [<https://www.nlm.nih.gov/medlineplus/druginfo/meds/a682019.html>] "Methotrexate" Consulted online in 2015.

Newbold, J E, H Xin, M Tencza, G Sherman, J Dean, S Bowden, and S Locarnini. 1995. "The Covalently Closed Duplex Form of the Hepadnavirus Genome Exists in Situ as a Heterogeneous Population of Viral Minichromosomes." *Journal of Virology* 69 (6) (June 4): 3350–7.

Pollicino, Teresa, Irene Cacciola, Francesca Saffioti, and Giovanni Raimondo. 2014. "Hepatitis B Virus PreS/S Gene Variants: Pathobiology and Clinical Implications." *Journal of Hepatology* 61 (2) (August 5): 408–17.

Qadri, Ishtiaq, Kaneez Fatima, and Hany AbdeL-Hafiz. 2011. "Hepatitis B Virus X Protein Impedes the DNA Repair via Its Association with Transcription Factor, TFIID." *BMC Microbiology* 11 (January 6): 48.

Quasdorff, M, and U Protzer. 2010. "Control of Hepatitis B Virus at the Level of Transcription." *Journal of Viral Hepatitis* 17 (8) (August): 527–36.

Rabe, Birgit, Angelika Vlachou, Nelly Panté, Ari Helenius, and Michael Kann. 2003. "Nuclear Import of Hepatitis B Virus Capsids and Release of the Viral Genome." *Proceedings of the National Academy of Sciences of the United States of America* 100 (17) (August 2): 9849–54.

Seeger, C, and W S Mason. 2000. "Hepatitis B Virus Biology." *Microbiology and Molecular Biology Reviews : MMBR* 64 (1) (March 3): 51–68.

Seeger, Christoph, and William S Mason. 2015. "Molecular Biology of Hepatitis B Virus Infection." *Virology* 479-480 (May 5): 672–86.

Société Canadienne du Cancer. [<http://www.cancer.ca/en/cancer-information/diagnosis-and-treatment/diagnosis/?region=qc>] "Diagnosis" Consulted online in 2015

Statistique Canada. 2011. "Principales causes de décès au Canada". CANSIM Tableau 102-0522

Stewart BW, and Wild, CP. 2014. "World Cancer Report". International Agency for Research on Cancer.

Summers, J, and W S Mason. 1982. "Replication of the Genome of a Hepatitis B--like Virus by Reverse Transcription of an RNA Intermediate." *Cell* 29 (2) (June 2): 403–15.

Summers, J, P M Smith, and A L Horwich. 1990. "Hepadnavirus Envelope Proteins Regulate Covalently Closed Circular DNA Amplification." *Journal of Virology* 64 (6) (June 5): 2819–24.

Sun, T T, and Y Y Chu. 1984. "Carcinogenesis and Prevention Strategy of Liver Cancer in Areas of Prevalence." *Journal of Cellular Physiology. Supplement* 3 (January): 39–44.

Szmuness, W, C E Stevens, E J Harley, E A Zang, W R Oleszko, D C William, R Sadovskiy, J M Morrison, and A Kellner. 1980. "Hepatitis B Vaccine: Demonstration of Efficacy in a Controlled Clinical Trial in a High-Risk Population in the United States." *The New England Journal of Medicine* 303 (15) (November 2): 833–41.

Tiollais, P, C Pourcel, and A Dejean. 1985. "The Hepatitis B Virus." *Nature* 317 (6037) (January 2): 489–95.

Tong, Aiping, Lantu Gou, Quek C Lau, Bin Chen, Xia Zhao, Jiong Li, Hong Tang, et al. 2009. "Proteomic Profiling Identifies Aberrant Epigenetic Modifications Induced by Hepatitis B Virus X Protein." *Journal of Proteome Research* 8 (2) (February): 1037–46..

Tortora, GJ, Funke, BR, Case CL, Martin, L. 2003. "Introduction à la microbiologie" ERPI, 1<sup>ère</sup> édition.

Twu, J S, and R H Schloemer. 1987. "Transcriptional Trans-Activating Function of Hepatitis B Virus." *Journal of Virology* 61 (11) (November): 3448–53.

Urban, Stephan, and Philippe Gripon. 2002. "Inhibition of Duck Hepatitis B Virus Infection by a Myristoylated Pre-S Peptide of the Large Viral Surface Protein." *Journal of Virology* 76 (4) (February 5): 1986–90.

Villarejos, V M, K A Visoná, A Gutiérrez, and A Rodríguez. 1975. "Role of Saliva, Urine and Feces in the Transmission of Type B Hepatitis." *The New England Journal of Medicine* 291 (26) (January 6): 1375–8.

Vivanco, Igor, and Charles L Sawyers. 2002. "The Phosphatidylinositol 3-Kinase AKT Pathway in Human Cancer." *Nature Reviews. Cancer* 2 (7) (July 1): 489–501.

Wang, Zhanhui, Yasuhito Tanaka, Yuehua Huang, Fuat Kurbanov, Jinjun Chen, Guobing Zeng, Bin Zhou, Masashi Mizokami, and Jinlin Hou. 2007. "Clinical and Virological Characteristics of Hepatitis B Virus Subgenotypes Ba, C1, and C2 in China." *Journal of Clinical Microbiology* 45 (5) (May 2): 1491–6.

WHO. 2015. [<http://www.who.int/mediacentre/factsheets/fs297/en/>] "Cancer" Consulted online in 2015

WHO. 2001. "Introduction of hepatitis B vaccine into childhood immunization services.Geneva", 1211 Geneva 27, Switzerland. Consulted online in 2015.

WHO. 2015. [<http://www.who.int/csr/disease/hepatitis/whocdscsrlyo20022/en/index3.html>] "Hepatitis B - The disease" Consulted online in 2015.

WHO. 2015. [<http://www.who.int/mediacentre/factsheets/fs164/en/>] "Hepatitis C" Consulted online in 2015

WHO. 2015. [<http://www.who.int/cancer/detection/en/>] "Early detection of cancer" Consulted online in 2015

Xie, Kun-Lin L, Yan-Ge G Zhang, Jun Liu, Yong Zeng, and Hong Wu. 2014. "MicroRNAs Associated with HBV Infection and HBV-Related HCC." *Theranostics* 4 (12) (January 3): 1176–92.

Zahm, P, P H Hofschneider, and R Koshy. 1988. "The HBV X-ORF Encodes a Transactivator: A Potential Factor in Viral Hepatocarcinogenesis." *Oncogene* 3 (2) (August 1): 169–77.

Zheng, Yanyan, Wen-ling L Chen, Stan G Louie, T S Yen, and Jing-hsiung James H Ou. 2007. "Hepatitis B Virus Promotes Hepatocarcinogenesis in Transgenic Mice." *Hepatology (Baltimore, Md.)* 45 (1) (January 1): 16–21.

Zoulim, F, J Saputelli, and C Seeger. 1994. "Woodchuck Hepatitis Virus X Protein Is Required for Viral Infection in Vivo." *Journal of Virology* 68 (3) (March 2): 2026–30.