



**UNIVERSITA' DEGLI STUDI DI MESSINA**

**DIPARTIMENTO DI MEDICINA CLINICA E SPERIMENTALE**

**Dottorato di Ricerca in  
Biotechnologie Mediche e Chirurgiche  
XXXIV Ciclo  
A.A. 2020/2021  
S.S.D. MED/04**

**Coordinatore: Ch.mo Prof.re G. Squadrito**

**CARATTERIZZAZIONE MOLECOLARE DEI SITI DI  
INTEGRAZIONE DEL VIRUS DELL'EPATITE B NEL  
CARCINOMA EPATOCELLULARE**

**Tesi di dottorato di:  
Dott.ssa Valeria Chines**

**Tutor:  
Prof.ssa Teresa Pollicino**



# INDICE

<b>INTRODUZIONE</b> .....	5
VIRUS DELL'EPATITE B (HBV) .....	5
Replicazione di HBV .....	6
Proteine virali .....	8
CARCINOMA EPATOCELLULARE (HCC).....	9
Meccanismi oncogenetici.....	9
Integrazione virale nel genoma umano ed epatocarcinogenesi.....	10
TECNICHE DI RILEVAZIONE DELL'INTEGRAZIONE VIRALE NEL DNA DELLA CELLULA OSPITE.....	11
<b>SCOPO DELLA TESI</b> .....	14
<b>MATERIALI E METODI</b> .....	15
PAZIENTI E LINEE CELLULARI .....	15
PREPARAZIONE DELLA LIBRERIA DA SEQUENZIARE .....	15
VALIDAZIONE DEI SITI DI INTEGRAZIONE OTTENUTI IN NGS .....	18
TRAMITE SEQUENZIAMENTO DI SANGER	
ANALISI COMPUTAZIONALI.....	18
Determinazione dei siti di integrazione virale.....	18
Determinazione dei siti di hotspot e dei motivi sovrarappresentati.....	19
<b>RISULTATI</b> .....	20
SITI DI INTEGRAZIONE .....	22
Breakpoint nel genoma umano.....	22
Breakpoint nel genoma di HBV.....	23
IDENTIFICAZIONE DELLE INTEGRAZIONI NELLE ESPANSIONI CLONALI.....	24
<b>DISCUSSIONI</b> .....	28
<b>BIBLIOGRAFIA</b> .....	30



# INTRODUZIONE

## VIRUS DELL'EPATITE B (HBV)

Il virus dell'epatite B (HBV) appartiene alla famiglia degli *Hepadnaviridae*. È stato stimato che circa 240 milioni di persone nel mondo sono affette da epatite B. L'infezione da questo virus può causare un'epatite acuta che può risolversi o progredire in un'infezione cronica. L'infezione cronica da HBV rappresenta un importante fattore di rischio per lo sviluppo del carcinoma epatocellulare (HCC) [P. Karayiannis, 2017]. Si stima che nel mondo muoiano circa 780.000 persone all'anno a causa del cancro epatico [H. B. El-Serag, 2012]. Il virione dell'HBV ha un diametro di 42 nm ed è chiamato "particella di Dane". Tale particella presenta un involucro lipoproteico esterno, detto *envelope*, e un rivestimento interno detto *capside*, che contiene al suo interno il genoma virale. Il genoma dell'HBV è un doppio filamento di DNA circolare, incompleto (rcDNA), di circa 3200 paia di basi. Il filamento di DNA a polarità positiva ha una lunghezza variabile, mentre il filamento a polarità negativa è lungo 3200bp [Figura 1].

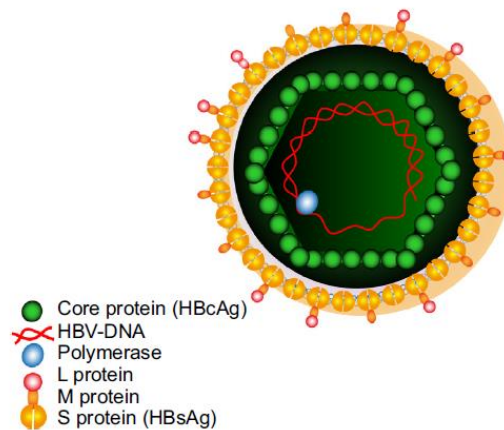


Figura 1. Rappresentazione schematica del virus dell'epatite B  
[Pollicino et al. 2014]

Il genoma di HBV risulta essere variabile portando alla selezione di quasi-specie virali che contengono diverse mutazioni; la variabilità del genoma è dovuta alla mancata capacità di *proof-reading* della polimerasi virale. Sono stati identificati nove differenti genotipi denominati con le lettere dell'alfabeto (da A ad I) e 35 sottogenotipi [A. Kramvis, 2014]. I diversi genotipi di HBV sono distribuiti geograficamente nel mondo: il genotipo A si trova principalmente in Europa, Nord

America, Africa e India; i genotipi B e C sono predominanti in Asia; il genotipo D nell'Europa orientale e nei paesi mediterranei; i genotipi E, F, G e H sono distribuiti nell'America Centrale, nell'America del Sud e in Europa centrale; mentre il genotipo I è maggiormente presente in Cina e Vietnam. [A. O. Santos *et al.*, 2010. S. Schaefer, 2007].

### **Replicazione di HBV**

Il sito di replicazione di HBV è rappresentato dalle cellule epatiche. Affinché la replicazione virale abbia inizio, si instaura un'interazione a bassa affinità tra il virus e i proteoglicani presenti sulla superficie cellulare. Avviene anche un'interazione ad alta affinità tra il dominio PreS1 di HBV e il recettore NTCP (sodium taurocholate cotransporting polypeptide) presente nella membrana basolaterale degli epatociti. Dopo l'interazione tra le proteine virali e i recettori cellulari segue, tramite endocitosi mediata, l'internalizzazione del virus [K. Watashi, 2014]. Segue il disassemblaggio del virione e il trasporto del genoma virale nel nucleo [H. Y. Kang, 2006]. HBV, pur essendo un virus a DNA, replica attraverso uno stampo a RNA, con un meccanismo simile a quello utilizzato dai retrovirus. Il primo evento della replicazione virale consiste nel convertire l'rcDNA in una forma circolare, completa e chiusa covalentemente, il cccDNA; questo avviene grazie alla sintesi della porzione mancante del filamento a polarità positiva e al legame dei due filamenti adesso completi. È stato suggerito che durante gli eventi di conversione dell'rcDNA in cccDNA intervengono enzimi cellulari [J. Kock, 1993]. Il cccDNA è una molecola molto stabile e si accumula nel nucleo della cellula ospite organizzandosi in una struttura simile ad un minicromosoma grazie alla presenza di proteine istoniche [P. Karayiannis, 2017], questo è responsabile della persistenza dell'infezione negli epatociti. Inoltre, il cccDNA funge da template per la trascrizione degli RNA messaggeri (mRNA) virali da parte della RNA polimerasi II presente nelle cellule ospiti. Gli mRNA includono 3 mRNA di diverse dimensioni che serviranno per la traduzione delle proteine virali, e un RNA pregenomico (pgRNA) che fungerà da template per la trascrizione inversa che porterà a generare nuove molecole di DNA genomico. Gli RNA prodotti presentano all'estremità 3' una coda poliadenilata, mentre all'estremità 5' è presente un cap. L'RNA pregenomico presenta una lunghezza di 3.5 kb, inoltre questo mRNA include anche informazioni relative alla porzione degli antigeni HBeAg e HBcAg, oltre che della polimerasi. I restanti RNA virali codificano per la proteina L dell'envelope (RNA di lunghezza pari a 2.4 kb), per l'antigene HBsAg e la proteina M (2.1 kb), e per la proteina HBx (0.8 kb). La maggior parte degli RNA migrano nel citoplasma della cellula dove sono utilizzati per la sintesi delle proteine virali. Nel citoplasma la polimerasi va a legare una struttura a stem-loop all'estremità 5' del pgRNA (*encapsidation signal* "epsilon"), in questo modo sia l'RNA virale che la polimerasi stessa vengono sequestrati all'interno del capsido virale sintetizzato dalla polimerasi

cellulare. [J. R. Pollac, 1993. G. H. Wang, 1994]. Il legame di epsilon all'RNA pregenomico è facilitato dal complesso molecolare chaperon hps90 della cellula ospite, infatti per inibire la replicazione di HBV vengono utilizzati anche farmaci che sono in grado di interferire con la funzionalità di hps90 [J. Hu, 1997]. La polimerasi, grazie alla sua funzione di trascrittasi inversa, utilizza il pgRNA come stampo per iniziare la sintesi del DNA virale a polarità negativa. Durante la sintesi della nuova catena di DNA, grazie all'attività dell'RNaseH, l'RNA pregenomico viene degradato. Successivamente avviene la sintesi del filamento a polarità positiva; in questo modo si ottiene il genoma virale a doppio filamento. Il capsid virale contenente al suo interno il genoma di nuova sintesi, viene rivestito con le proteine dell'envelope e i virioni maturi vengono rilasciati in circolo.

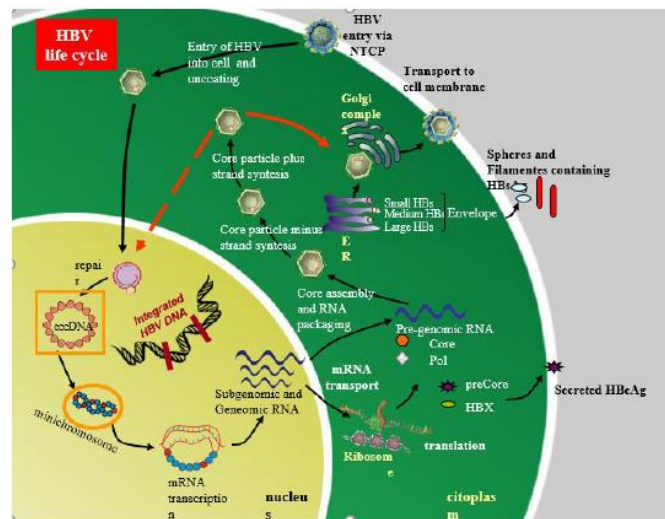


Figura 2. Rappresentazione schematica del ciclo vitale di HBV [Raimondo et al. 2007]

Diversamente da quanto si osserva nei retrovirus, l'integrazione del genoma virale all'interno dei cromosomi umani non svolge un ruolo nella replicazione di HBV. Tuttavia, si osservano di frequente fenomeni di integrazione [T. M. Block, 2003] noti per avere un importante ruolo nella trasformazione oncogenica degli epatociti infetti.

Il genoma di HBV presenta quattro *Open Reading Frame* (ORF) che si sovrappongono tra di loro, queste codificano per le proteine virali: preS/S (codificante per gli antigeni di superficie), preCore/Core (codificante le proteine del capsid), Pol (codificante per la polimerasi) ed X (codificante per la proteina X). L'ORF P racchiude al suo interno l'intera regione genica preS/S e alcune porzioni delle regioni PreC/C ed X [Figura 3].

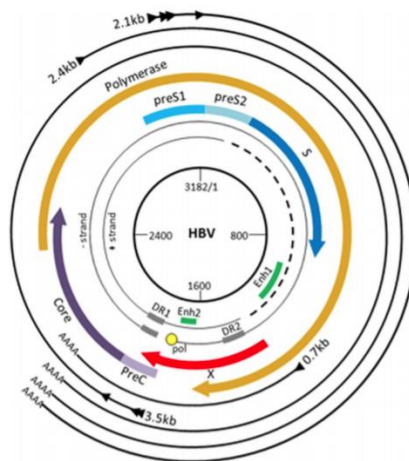


Figura 3. Organizzazione del genoma di HBV [Minor e Slagle, 2014]

## Proteine virali

PreS/S codifica per le tre proteine strutturali dell'envelope: Small (S), Medium (M) and Large (L). Sono generate da una traduzione alternativa partendo dal codone AUG, presentano la stessa estremità carbossi-terminale ma differente estremità ammino-terminale. Sono proteine glicosilate transmembrana di tipo II e formano multidimeri stabilizzati da ponti disolfuro [C. Seeger, 2000]. La proteina S definisce il dominio S (HbsAg) ed è costituita da 226 aminoacidi (aa), la proteina M è conosciuta come PreS2 e presenta un'estensione all'estremità N-terminale di 55 aa rispetto alla proteina S, invece la proteina L (PreS1) presenta un'estensione di 108 o 119 aa, dipendente dal genotipo virale [T. Pollicino, 2014]. Le proteine dell'envelope contengono i principali domini antigenici virali; in particolare è presente il dominio anti-HBs, chiamato *A-determinant*, mappato a livello degli aminoacidi 99-170 della proteina S [C. Seeger, 2000. H. Norder, 1992]. Gli antigeni di superficie prodotti dalla proteina S vengono secreti all'interno delle cellule infettate e la loro affinità per gli anticorpi causa la risposta immunitaria dell'ospite.

PreC/C codifica per due proteine: la proteina strutturale del nucleocapside chiamata antigene core, dell'epatite B, HBcAg, e per la proteina non strutturale conosciuta come antigene-e (HBeAg). Le due proteine derivano da traduzioni alternative a partire da due differenti codoni di start: dal codone AUG presente nella porzione del PreCore prende inizio la traduzione della proteina HBcAg, mentre dal codone AUG sito nella porzione genica del Core viene tradotta la proteina HBeAg [D. N. Standing, 1988]. Le subunità del core formano dei dimeri [C. Chang, 1994], successivamente un totale di 120 dimeri si assembleranno in un involucro di 36 nm di diametro [R. A. Crowther, 1994].



L'ORF Pol codifica un enzima necessario per la sintesi del DNA virale, DNA polimerasi DNA dipendente, la quale è localizzata all'interno del nucleocapside. Questa polimerasi è una proteina multifunzione, infatti presenta attività di DNA polimerasi DNA dipendente, di trascrittasi inversa (RT) e di RNaseH; inoltre include, all'estremità N-terminale, un dominio proteico terminale (TP) utilizzato come primer per la sintesi del genoma virale [T. Pollicino, 2014].

La regione X codifica per una piccola proteina regolatoria di 154 aa essenziale per incrementare la trascrizione del cccDNA e per la replicazione virale, essa può regolare direttamente e indirettamente l'espressione genica virale e dell'ospite [C. Seeger, 2000. M. Nassal, 2008].

## **CARCINOMA EPATOCELLULARE (HCC)**

Il carcinoma epatocellulare (HCC) è un tumore solido considerato la seconda causa di mortalità per cancro a livello mondiale [H. B. El-Searag, 2012]. Lo sviluppo di HCC è guidato da diversi fattori come la predisposizione genetica, fattori ambientali (come sindromi metaboliche, alcol, esposizione all'aflatossina B1) e infezioni virali (virus dell'epatite B e virus dell'epatite C). La trasformazione degli epatociti in cellule tumorali è guidata da diversi fattori tra cui infiammazione cronica, danno al DNA, modificazioni epigenetiche, riattivazione della telomerasi e l'instabilità cromosomica [M. Levrero, 2016]; questi fattori sembra che attivino i medesimi meccanismi oncogenetici, i quali possono provocare mutazioni puntiformi e aberrazioni cromosomiche.

È stato rilevato che la presenza di infezione cronica da HBV (CHB) aumenta la probabilità di sviluppare HCC. Infatti, i pazienti affetti da CHB mostrano un aumento nella possibilità di sviluppare tumore di circa 100 volte rispetto ai pazienti che non presentano infezione da HBV.

### **Meccanismi oncogenetici**

L'HCC è caratterizzato da un pattern di mutazioni che si accumulano nel tempo e che contribuiscono allo sviluppo e alla progressione del tumore [K. Schulze, 2016] [F. Guerrieri, L. Belloni, N. Pediconi, M. Levrero. Molecular mechanisms of HBV-associated hepatocarcinogenesis. 2013], queste mutazioni sono state riscontrate soprattutto in pazienti che presentano anche infezione cronica da HBV. La mutazione maggiormente riscontrata coinvolge la subunità catalitica dell'enzima telomerasi, chiamata *telomerase reverse transcriptase* (TERT); questa è stata rilevata nel 65% dei casi di HCC [J. C. Nault, 2014. J. Zucman-Rossi, 2015]. La suddetta mutazione è associata ad un incremento dell'espressione della telomerasi permettendo alla cellula di acquisire l'abilità di superare la senescenza [S. H. Ahn, 2005]. Mentre la mutazione nel promoter di TERT si riscontra nei primi stadi di sviluppo dell'HCC, altre mutazioni sono state riscontrate in step più tardivi della progressione tumorale. Queste coinvolgono geni appartenenti a pathway di segnalazione dell'oncogenesi come i

pathway WNT/ $\beta$ -catenine (mutazioni riscontrate a livello del gene CTNNB1 che codifica per la  $\beta$ -catenina) e il pathway del ciclo cellulare TP53. Queste mutazioni portano alla compromissione della degradazione della  $\beta$ -catenina e all'attivazione della segnalazione Wnt, il quale promuove la motilità cellulare, de-differenziazione e proliferazione [H. Clevers, 2012].

È stato evidenziato in diversi studi che in pazienti affetti da HBV, vi era la presenza di polimorfismi a singolo nucleotide (SNPs); questi, rispetto a dei casi controllo, aumentano il rischio di sviluppo di HCC [J. Zucman-Rossi, 2015].

### **Integrazione virale nel genoma umano ed epatocarcinogenesi**

Uno dei più importanti meccanismi grazie al quale HBV è in grado di aumentare il rischio di sviluppo dell'epatocarcinoma cellulare è la sua capacità di integrarsi all'interno del genoma umano. A seguito dell'espansione clonale degli epatociti, verrà trasmesso alle nuove cellule anche il genoma virale integrato. L'integrazione si rileva in circa l'80% dei casi di HCC [M. J. Bouchard, 2011] e la sua presenza in tumori che si sono sviluppati in fegati non cirrotici di bambini o giovani adulti supporta ulteriormente il ruolo dell'integrazione del DNA virale nel processo di epatocarcinogenesi [K. Yaginuma, 1987. H. Sezaki, 2004. Y. Murakami, 2005]. La maggior parte dei siti di *breakpoint* sono localizzati in prossimità di genoma codificante, all'interno di esoni o di regioni regolatorie la cui espressione è sovregolata nei tumori rispetto al tessuto non tumorale [W. K. Sung, 2012]. Altri siti comuni di integrazione del genoma virale si trovano in sequenze ripetitive non codificanti come LINE (*long interspersed nuclear elements*), SINE (*short interspersed nuclear elements*), LTR (*long terminal repeat*) ed elementi Alu (breve sequenza interspersa del genoma riconosciuta come sito di taglio dell'endonucleasi Alu). Inoltre, altri siti di integrazione conosciuti sono le *CpG islands*, siti fragili del genoma, introni o regioni *promoter* [R. Bonilla Guerrero, 2005. D. Ding, 2012. A. Fujimoto, 2012. W. K. Sung, 2012. C. C. Lau, 2014. L. H. Zhao, 2016].

L'integrazione non contribuisce alla replicazione virale ma può causare instabilità cromosomica, una de-regolazione dell'espressione dei geni dell'ospite e permette una continua espressione di proteine troncate dell'X (HBx), del core (HBc) e del preS; l'espressione non regolata di queste proteine è associata all'epatocarcinogenesi [F. Guerrieri, 2013. R. Bonilla Guerrero, 2005. M. Levrero, 2016. T. Tu, 2017].

Le proteine HBx wild-type e HBx troncate hanno entrambe funzioni oncogene e promuovono la tumorigenesi. HBx è in grado di legarsi ad un elevato numero di sequenze bersaglio, sia appartenenti a geni codificanti proteine coinvolte nel metabolismo cellulare, nelle modificazioni epigenetiche della cromatina e nella cancerogenesi, sia geni non codificanti (come promoter di lncRNA, snRNA e miRNA) noti per modulare la replicazione dell'HBV [F. Guerrieri, 2015]. HBx è quindi in grado di

agire sulla trascrizione dei geni (sia cellulari che virali) grazie alla sua capacità di intervenire nell'epigenetica e di interagire, attivandone o reprimendo la trascrizione, con molteplici fattori trascrizionali e con enzimi la cui funzione è quella di modificare la cromatina. Inoltre, HBx contribuisce a oltrepassare la senescenza attraverso vari meccanismi, tra cui la sovraespressione dell'espressione di TERT [M. Ozturk, 2009], diminuendo l'espressione di ASPP1 e ASPP2 (attivatori di TP53) [J. Zhao, 2010] e inattivando il soppressore tumorale RB [S.H. Park, 2011].

L'espressione e il conseguente accumulo di proteine S nel reticolo endoplasmatico cellulare può essere tossico per l'epatocita stesso e avviare una serie di eventi che progrediscono in una trasformazione in cellula tumorale [F.V. Chisari, 1989]. Proteine preS2 mutate (presentano l'estremità 3' deleta), rilevate in molti casi di HCC associato ad infezione da HBV, inducono la sovraespressione della ciclina A e della ciclossigenasi-2 che stimolano a loro volta la proliferazione cellulare [J.H. Hung, 2004. L.H. Wang, 2012]. Inoltre, preS2 mutato interagisce direttamente con una regione di DNA del promotore di TERT determinando l'aumento dell'attività della telomerasi e la promozione dello sviluppo di HCC [L. de La Coste, 1998].

La proteina HBc, oltre a svolgere la sua funzione di legame del minicromosoma di HBV, è in grado di legare diversi geni cellulari coinvolti nel controllo della proliferazione cellulare [Y. Guo, 2012].

Molti aspetti molecolari dell'integrazione dell'HBV rimangono, tuttavia, controversi o ancora poco chiari, come ad esempio i meccanismi molecolari implicati nell'integrazione virale. L'ipotetica forma virale di integrazione nel genoma umano è l'HBV DNA lineare a doppio filamento (dsDNA) [C. A. Bill, 2004], sebbene sia stato suggerito che anche altre forme molecolari di HBV potrebbero essere coinvolte nel meccanismo di integrazione virale [T. Tu, 2017]. Inoltre, non è chiaro se l'integrazione del DNA dell'HBV avvenga attraverso la giunzione non omologa indipendente dalla sequenza (NHEJ) o se possa avvenire anche attraverso giunzione mediata da micromologia (MMEJ) [C. A. Bill, 2004. R. Bonilla Guerrero, 2005. L. H. Zhao, 2016].

## **TECNICHE DI RILEVAZIONE DELL'INTEGRAZIONE VIRALE NEL DNA DELLA CELLULA OSPITE**

Esistono differenti tecniche che permettono di individuare i fenomeni di integrazione virale.

Inizialmente essa è stata studiata tramite l'utilizzo di metodi di ibridizzazione come il southern blot [J. C. Edman, 1980], utilizzando enzimi di restrizione e sonde di DNA specifiche per il genoma virale. Dopo aver effettuato l'estrazione del DNA dai tessuti di fegato, viene eseguita una digestione utilizzando l'enzima di restrizione EcoRI. In presenza di integrazione i frammenti di DNA contengono sia DNA virale che le sequenze umane; i prodotti della digestione verranno visualizzati

tramite gel di agarosio. Questa tecnica permette di distinguere il DNA virale libero dalle sequenze di HBV integrate nel genoma dell'ospite in quanto i frammenti chimerici sono più lunghi rispetto al DNA libero [S. Wirth, 1992]. Questa tecnica, nonostante sia poco dispendiosa e facile da eseguire, presenta alcuni limiti: non è possibile identificare le sequenze umane all'interno delle quali si è integrato l'HBV e non si può definire il numero di copie dei singoli integrati.

Un'altra tecnica utilizzata per rilevare la presenza di virus integrato nel genoma della cellula ospite è l'Alu-PCR. Questa metodica sfrutta la presenza, all'interno del genoma umano, di brevi tratti di DNA ripetuto di circa 300 bp definiti elementi Alu; questi elementi sono presenti nel genoma in percentuale maggiore rispetto ad altri elementi ripetuti, arrivando a contare oltre un milione di copie [K. Kitada, 2013]. Vengono utilizzate delle sequenze oligonucleotidiche (primer) specifiche per le sequenze virali e per le porzioni Alu in modo da poter individuare la presenza di chimere HBV/genoma. Uno dei limiti di questa tecnica è l'impossibilità di rilevare integrati di HBV lontani dalle sequenze Alu, portando alla necessità di sviluppare ulteriori metodiche.

La nested PCR inversa (invPCR) rappresenta un altro metodo per lo studio dell'integrazione di HBV nel DNA umano. È un metodo altamente sensibile, specifico e poco costoso [M.A. Budzinska, 2018], infatti è sfruttato in diversi campi scientifici tra cui la biologia molecolare e la genetica medica. Questa metodica richiede una quantità non elevata di DNA da estrarre e di conseguenza non sono necessari frammenti molto grandi di tessuto epatico [T. Tu, 2017]. La PCR inversa viene utilizzata per amplificare frammenti di DNA che fiancheggiano l'estremità di una sequenza nota di DNA e per la quale non sono disponibili primer. Il DNA contenente la sequenza nota e la sua regione fiancheggiante viene digerito da parte di un enzima di restrizione; il prodotto di tale digestione viene convertito in DNA circolare che sarà utilizzato come stampo nella PCR. La sequenza sconosciuta è amplificata da due primer che si legano alla sequenza nota e la dimensione del frammento amplificato dipende dalla distribuzione dei siti di restrizione all'interno di sequenze di DNA note e fiancheggianti [R. Michael, 2019].

Recentemente sono state sviluppate tecnologie di sequenziamento di nuova generazione (NGS) per rilevare la presenza di integrazione virale. Tra queste tecniche vi è il sequenziamento dell'intero genoma (*Whole genome sequencing*, WGS) che, oltre a identificare la presenza di integrazioni del genoma virale in quello umano, fornisce ulteriori informazioni tra cui presenza di mutazioni a singolo nucleotide (SNV), inserzioni o delezioni, variazioni nel numero di copie del DNA (CNV) e modifiche strutturali [Z. Kan, 2013]. Nonostante il WGS abbia reso possibile la rilevazione di chimere sull'intero genoma presenta dei limiti: necessità di utilizzare una quantità elevata di DNA iniziale, difficoltà di individuare integrazioni che hanno una bassa frequenza e i costi relativamente alti.

Dunque, per rilevare le integrazioni, è stata sviluppata una tecnica che va ad arricchire i frammenti di DNA che contengono giunzioni chimeriche [The Cancer Genome Atlas Research Network. 2017] favorendo l'amplificazione e l'identificazione dei siti unici di integrazione. In questo modo si sono abbassati i costi di sequenziamento dell'NGS a scapito del coverage delle reads di HBV che è diminuito.

## SCOPO DELLA TESI

Al fine di individuare i siti di integrazione di HBV nel genoma umano, sono stati studiati sia tessuti epatici tumorali (T) e non tumorali (NT) di pazienti affetti da epatocarcinoma cellulare correlato ad infezione da HBV, sia diverse linee cellulari tramite l'utilizzo di una metodica di *Next Generation Sequencing* (NGS) ad alta sensibilità. Questa metodologia applicata ci permette anche di quantificare i cloni cellulari recanti le stesse integrazioni.

# MATERIALI E METODI

## PAZIENTI E LINEE CELLULARI

Sono stati analizzati tessuti epatici tumorali (T) di 6 pazienti HBsAg positivi (5 maschi e 1 femmina, l'età media è di  $66.7 \pm 8$ ) con HCC e i relativi tessuti non tumorali (NT). Come controllo negativo sono stati analizzati tessuti epatici derivanti da 2 pazienti HBsAg negativi. Sia i pazienti HBsAg positivi che i pazienti HBsAg negativi, provenienti dalla Chirurgia Oncologica dell'Azienda Ospedaliera Universitaria di Messina, risultano essere negativi per le infezioni da HDV, HCV e HIV. Tutti i tessuti sono stati conservati a  $-80\text{ }^{\circ}\text{C}$  in RNAlater (Applied Biosystems/Ambion).

Per questo lavoro sono state studiate diverse linee cellulari come controlli positivi e negativi al fine di validare il nostro protocollo; le cellule utilizzate sono state le seguenti: PLC/PRF/5, HepG2, Vero ed HepAD38. Le cellule PLC/PRF/5 (SIGMA) appartengono ad una linea cellulare di epatoma umano e contengono DNA di HBV integrato nel loro genoma. Le cellule HepG2 (ATCC) sono state mantenute in coltura nel terreno *Dulbecco's modified Eagle medium* (DMEM) con GlutaMAX, 10% di FBS, 200 unità/ml di penicillina e 200  $\mu\text{g/ml}$  di streptomina. Le cellule Vero (fornite dalla Professoressa Maria Teresa Sciortino dell'Università di Messina) sono state mantenute in coltura nel terreno DMEM con GlutaMAX, 6% di FBS, 200 unità/ml di penicillina e 200  $\mu\text{g/ml}$  di streptomina. Le cellule HepAD38 (fornite dal Dottor Christopher Seeger), che derivano dalle cellule HepG2, contengono il genoma di HBV (sottotipo ayw) il cui *promoter* è sotto il controllo della tetraciclina; queste vengono mantenute in coltura in terreno DMEM/F12 con GlutaMAX, 10% di FBS, 200 unità/ml di penicillina, 200  $\mu\text{g/ml}$  di streptomina e 0.3  $\mu\text{g/ml}$  di tetraciclina. La rimozione della tetraciclina dal terreno di coltura delle HepAD38 induce la produzione di RNA di HBV e la secrezione nel terreno stesso di particelle simil-virali [S. K. Ladner, 1997].

Questo studio è stato approvato dalla commissione etica dell'Ospedale Universitario di Messina e i consensi informati sono stati ottenuti da tutti i pazienti.

## PREPARAZIONE DELLA LIBRERIA DA SEQUENZIARE

Il DNA genomico è stato ottenuto dai tessuti tramite lisi in proteinasi K utilizzando un buffer di lisi contenente 150 nM di NaCl, 10 nM di tris HCl a pH 8, 10 nM di EDTA, 0.5% di SDS e 1mg/ml di proteinasi K. Successivamente è stata effettuata un'estrazione fenolo/cloroformio.

Tutti i campioni sono stati normalizzati affinché si lavorasse con lo stesso numero di cellule, per cui è stata quantizzata la  $\beta$ -globina tramite lo strumento Light-Cycler (Roche Diagnostic). Per ogni campione è stato analizzato un numero di cellule pari a  $1 \times 10^7$ .

Il DNA genomico è stato frammentato attraverso sonicazione tramite lo strumento Sonopuls Ultrasonic Homogenizer (Bandelin); sono stati effettuati 3 cicli in cui il DNA è stato sonicato per 15 secondi e successivamente posto in ghiaccio per altrettanto tempo (15'' on/15'' off). Dalla sonicazione si ottiene un pool di frammenti di DNA con dimensioni che variano da 100bp a 1000bp.

Una volta ottenuti i frammenti di DNA ne vengono riparate le estremità attraverso una reazione di blunted, questa è eseguita utilizzando aliquote contenenti 5  $\mu$ g di DNA. La reazione è stata realizzata tramite End-it DNA Repair Kit (Epicentre) e successivamente purificata attraverso Qiaquick PCR Purification Kit (Qiagen).

Ai frammenti di DNA è stata quindi ligata all'estremità 3' una coda di poly(A) preparando una mix di reazione contenente 5  $\mu$ l di NEB buffer 10x, 1  $\mu$ l di dATP 10 mM e 2  $\mu$ l del frammento di Klenow 3'  $\rightarrow$  5' exo- 5000 U/ml (New England Biolabs Kit). La reazione è stata incubata a 37 °C per 1 ora. Segue una purificazione attraverso Qiaquick PCR Purification Kit (Qiagen).

Ad ogni aliquota di DNA sono stati ligati dei linker (Linker Top e Linker Bottom) preparando una mix di reazione contenente 4  $\mu$ l di linker, 5  $\mu$ l di T4 DNA ligase buffer (NEB) e 1  $\mu$ l di T4 DNA ligase ( $2 \times 10^6$  U/ml, New England Biolabs). I campioni così preparati sono stati incubati a 25 °C per 1 ora e a 16 °C overnight. La ligasi è stata poi inattivata a 70 °C per 20 minuti e i prodotti della ligazione sono stati purificati tramite Qiaquick PCR Purification Kit (Qiagen). Le aliquote sono state unite formando un unico pool.

È stata effettuata una reazione di PCR utilizzando dei primer (sia forward che reverse) biotinilati specifici per le regioni genomiche di HBV. Il pool è stato diviso in aliquote contenenti ciascuna 1  $\mu$ g di DNA e ad ognuna di queste è stata aggiunta la seguente mix di reazione: 20  $\mu$ l di Phusion HF buffer 5x, 3  $\mu$ l di dNTP 10mM, 1  $\mu$ l di primer biotinilato forward o reverse (20  $\mu$ M), 1  $\mu$ l di Phusion Taq 2000 U/ml (New England Biolabs) ed è stata portata ad un volume di 50  $\mu$ l con H<sub>2</sub>O. La reazione è stata posta in termocycler utilizzando il seguente programma: 1x (98 °C, 1 min) 12x (98 °C 15 s, 65 °C 30 s, 72 °C 45 s) 1x (72 °C 1 min) 1x (4 °C  $\infty$ ). In ogni tubo di reazione è stato successivamente aggiunto 1  $\mu$ l di pLinker 2.5  $\mu$ M ed è stato avviato un ulteriore ciclo di PCR: 1x (98 °C 1 min) 35x (98 °C 15 s, 65 °C 30 s, 72 °C 45 s) 1x (72 °C 5 min) 1x (4 °C  $\infty$ ). I prodotti della PCR sono stati poi purificati tramite Qiaquick PCR Purification Kit (Qiagen).



È stata eseguita una *size selection* dei prodotti purificati tramite corsa elettroforetica su gel di agarosio (2%), i frammenti sono stati selezionati in un range compreso tra 300 bp e 1000 bp. Il DNA è stato estratto da gel e purificato tramite Qiaquick Gel Extraction Kit (Qiagen). A ciascun prodotto di PCR sono stati aggiunti 100 µl di beads magnetiche funzionalizzate con la streptavidina (Invitrogen) e incubate 1 ora in agitatore rotante a temperatura ambiente. Dopo incubazione le beads sono state isolate tramite barra magnetica e sono stati effettuati 3 lavaggi con 500 µl di buffer B&W 1X (10 mM Tris pH7.5, 1 mM EDTA, 2 M NaCl) e un lavaggio con 500 µl di H<sub>2</sub>O. Dopo i lavaggi i campioni sono stati risospesi in 50 µl di H<sub>2</sub>O.

Viene effettuato un secondo round di PCR utilizzando primer che contengono all'estremità 5' degli adattatori Illumina fondamentali per il legame delle librerie alla superficie della flow cell del sequenziatore MiSeq. È stata preparata la seguente mix: 10 µl di Phusion HF buffer 5x, 1.5 µl di dNTP 10 mM, 1 µl di primer forward o reverse (20 µM), 1 µl di MiSeq-pLinker forward o reverse (20 µM), 0.5 µl di Phusion Taq 2000 U/ml ed è stata portata ad un volume di 50 µl con H<sub>2</sub>O. È poi stato avviato il seguente ciclo di PCR: 1× (98 °C 1 min) 35× (98 °C 10 s, 65 °C 40 s, 72 °C 40 s) 1× (72 °C 5 min) 1× (4 °C - ∞). È stato effettuato un ulteriore ciclo di PCR in modo da permettere l'arricchimento degli adattatori Illumina: 1x (98C° 1 min) 25x (98C° 30 sec, 55C° 30 sec, 72C° 30 sec) 1x (72C° 5 min). I prodotti della PCR sono stati poi purificati tramite Qiaquick Gel Extraction Kit (Qiagen).

Le librerie vengono quantizzate ed unite in modo equimolare in un'unica aliquota e successivamente sequenziate tramite sequenziamento 250 bp paired-end su piattaforma Illumina MiSeq. Sono state costruite un totale di 340 librerie [Figura 4].

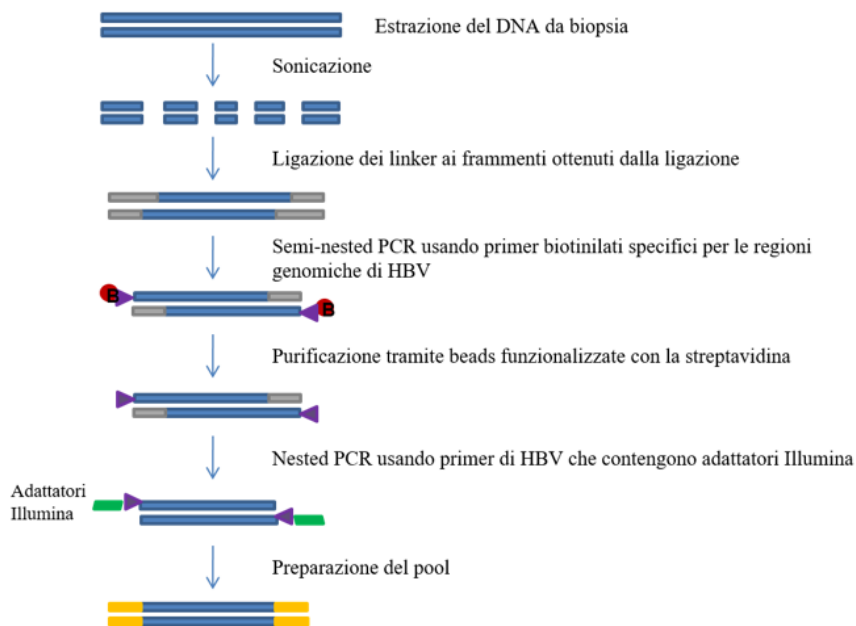


Figura 4. Rappresentazione schematica del protocollo di preparazione delle librerie

## VALIDAZIONE DEI SITI DI INTEGRAZIONE OTTENUTI IN NGS TRAMITE SEQUENZIAMENTO DI SANGER

Il DNA genomico estratto è stato sottoposto a nested-PCR con primer specifici per le regioni genomiche e virali delle chimere rilevate tramite metodica NGS utilizzando HotStart Taq Polymerase (Qiagen), ed è stato eseguito il seguente protocollo termico: 1x (98 °C 14 min) 40x (98 °C 30 s, 55 °C 30 s, 72 °C 30 s) 1x (72°C 5 min). I prodotti sono stati isolati tramite elettroforesi su gel di agarosio e successivamente è stato effettuato un sequenziamento diretto di Sanger tramite piattaforma Applied Biosystems (Foster City, CA) 3500 DNA analyzers.

## ANALISI COMPUTAZIONALI

### Determinazione dei siti di integrazione virale

I dati sono presentati sottoforma di media e deviazione standard (S.D.) e le variabili categoriche sottoforma di frequenza e percentuale assolute. Per esaminare la distribuzione delle variabili è stato utilizzato un approccio non parametrico applicando il test di *Kolmogorov-Smirnov*. Per confrontare diversi gruppi di variabili categoriche è stato utilizzato il test del  $\chi^2$  o test esatto di Fisher. Lo *Student's t test* è stato applicato per eseguire il confronto tra gruppi di variabili numeriche. Infine, il test NPC insieme al *Fisher combination function test* è stato utilizzato per combinare il p-value ottenuto da

diverse analisi. Le analisi statistiche sono state eseguite utilizzando il pacchetto Windows SPSS 22.0. Affinché un dato venga considerato statisticamente significativo il P-value deve risultare inferiore a 0,05.

I dati ottenuti dal sequenziamento Illumina vengono lavorati tramite il programma Trimmomatic (v.0.36) [A. M. Bolger, 2014] in modo da eliminare le reads di bassa qualità e le sequenze relative agli adattatori. Le reads così ottenute sono state mappate contro un genoma di riferimento per ricostruire un'unica sequenza in Fasta tramite il software BWA [H. Li, 2009]. I genomi di riferimento usati sono: GRCh38.p10 per quanto riguarda il genoma umano, e NC\_003977.2 per il genoma virale. Per ottenere il numero corretto di eventi di integrazione da parte di HBV all'interno del genoma umano è stato utilizzato il programma BEDtool [A. R. Quinlan, 2014] e successivamente sono state ricostruite le sequenze chimeriche utilizzando i software Cap3 [X. Huang, 1999] e cd-hit [L. Fu, 2012]. Le chimere assemblate sono state poi mappate sul genoma ibrido utilizzando l'algoritmo BLAST [C. Camacho, 2009]. La micromologia è stata determinata attraverso l'identificazione in ogni chimera delle regioni di DNA che mappano sia sul genoma umano che sul genoma virale.

I siti di integrazione virale sono stati considerati come clonali quando la maggior parte delle reads ( $\geq 92\%$ ) in cui è stato registrato l'evento di integrazione mostra la medesima sequenza nucleotidica al livello del sito di *breakpoint*.

### **Determinazione dei siti hotspot e dei motivi sovra rappresentati**

I siti hot-spot sono definiti come piccole regioni di DNA in cui si verifica molto frequentemente l'integrazione dell'HBV DNA. Per identificare i suddetti siti è stato utilizzato il software BEDtool ed è stato applicato il test Monte-Carlo. Utilizzando il test chi-quadrato, abbiamo confrontato il numero di integrazioni osservate con quelle ottenute dalle simulazioni, al fine di confermare come potenziali hot-spot solo quei siti che mostrano una frequenza superiore al background randomizzato.

Il software HOMER è stato utilizzato per determinare se specifiche sequenze umane o virali sono state trovate sovrarappresentate nelle sequenze chimeriche. L'analisi è stata eseguita unendo solo sequenze chimeriche umane, solo sequenze chimeriche virali e tutte le intere chimere trovate in questo studio.

## RISULTATI

È stato ottenuto un totale di 297 librerie di integrazione provenienti dal tessuto epatico tumorale di 6 pazienti con HCC affetti da HBV, e dai relativi tessuti epatici non tumorali. Per l'identificazione dell'integrazione del DNA virale nel genoma umano, è stato modificato un protocollo precedentemente descritto per studiare l'integrazione del virus dell'immunodeficienza umana (HIV) nelle cellule T CD4+ [L. Cohn, 2015]. Tramite questa metodologia (HBIS, *HBV Integration Sequencing*) sono state ottenute le integrazioni virali attraverso un semi-nested PCR, partendo dal DNA del paziente, utilizzando dei primer specifici per le differenti regioni genomiche dell'HBV. I risultati della PCR sono stati ulteriormente arricchiti tramite nested-PCR utilizzando primer che contengono degli adattatori Illumina, necessari per il legame alla flow cell Illumina. Le librerie ottenute sono state successivamente sequenziate tramite metodica NGS utilizzando la piattaforma Illumina MiSeq.

Dai risultati delle analisi è stato rilevato un totale di 2839 siti unici di integrazione di HBV, nello specifico 2476 siti unici di integrazione sono stati riscontrati nei T e 363 nei NT ( $P = 0.027$ , *Student's t test*). Le integrazioni di HBV sono state riscontrate in tutti i pazienti studiati e i siti di integrazione virale sono stati rilevati in tutti i tessuti studiati, sia T che NT, e ogni sito di integrazione presentava almeno 3 *paired-end reads*.

Al fine di confermare la veridicità degli eventi di integrazione, sono stati selezionati in modo casuale 10 sequenze chimeriche ottenute dal sequenziamento in NGS; queste sono state analizzate tramite sequenziamento diretto di Sanger. Dai risultati del sequenziamento si è ottenuta la conferma di tutti i siti di integrazione selezionati.

Nonostante l'elevato numero di integrazioni virali riscontrate nei tessuti epatici tumorali e i relativi tessuti non tumorali, solo 24 di queste sono in comune (rilevate in 3 pazienti). Questo risultato va a rinforzare il concetto che il pattern di integrazione risulta differente tra tessuti tumorali e non. [W. K. Sung *et al.*, 2012. L. H. Zhao *et al.* 2016].

È stato riscontrato che i siti unici di integrazione dell'HBV risultano distribuiti in maniera randomica lungo tutto il genoma umano. Tuttavia, l'analisi della distribuzione dei *breakpoint* di HBV nei singoli cromosomi ha rivelato che i campioni di tumore presentano una distribuzione statisticamente significativa dei siti di integrazione sul cromosoma 20 ( $P = 0,018$ , test NPC). Nei campioni di tessuto non tumorale, invece, si è riscontrato una distribuzione relativamente casuale dei siti di integrazione tra i vari cromosomi.

Al fine di validare il protocollo di sequenziamento HBIS, sono state costruite 60 librerie di DNA provenienti da 2 pazienti non infetti da HBV e da cellule Vero HBV-negative. Dall'analisi dei risultati non è stato rilevato nessun sito di integrazione virale, né nei campioni provenienti dai pazienti né nelle cellule Vero. Il protocollo è stato verificato anche analizzando l'integrazione di HBV nelle cellule PLC/PRF/5; in queste cellule sono stati rilevati 104 siti di integrazione virale, 12 dei quali erano siti unici di integrazione (coverage di almeno 3 reads). Tra questi sono stati riscontrati eventi di integrazione nella regione del promoter di TERT, nei geni MVK, CCDC57 e UNC5D e a valle del gene STARD13. La sequenza integrata di HBV nella regione del promoter del gene TERT includeva una sequenza (da nt 1260 a nt 1391) della regione ENH I/promoter X. Nei geni MVK e CCDC57 le sequenze virali integrate includevano due diverse sequenze del gene S (rispettivamente da nt 711 a nt 817 e da nt 456 a nt 491), nel gene UNC5D la porzione integrata del virus includeva una sequenza del gene core (da nt 2392 a 2436). Inoltre, sono state identificate nuove integrazioni di HBV nel gene NINL (Ninein Like), LOC105375716 ncRNA, TBC1D8 (TBC1 Domain Family Member 8) e in prossimità di RPS23P4 (pseudogene 4 della proteina ribosomiale S23), LOC105374110 e LOC101929814 ncRNAs [Figura 5]. Le integrazioni virali nel promoter di TERT e nel gene CCDC57 sono state confermate anche dalla PCR e dal sequenziamento di Sanger.

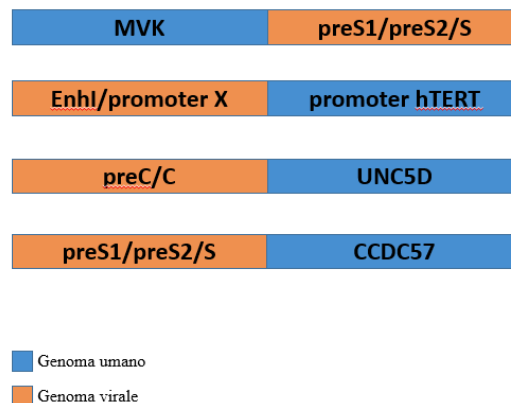


Figura 5. Rappresentazione schematica dei siti di *breakpoint* nelle cellule PLC/PRF/5.

## SITI DI INTEGRAZIONE

### Breakpoint nel genoma umano

È stata analizzata la distribuzione dei siti di integrazione dell'HBV in specifici elementi funzionali e strutturali del genoma umano. La maggior parte dei *breakpoint* nei tessuti tumorali si trovava in porzioni intergeniche (941/2476 rispetto a 100/363;  $P < 0,0001$ , test  $\chi^2$ ), mentre nei campioni non tumorali i siti di integrazione erano localizzati soprattutto nelle regioni geniche (1322/2476 rispetto a 218/363;  $P = 0,017$ ) e in particolare a livello degli introni (881/2476 rispetto a 151/363;  $P = 0,012$ ) e dei microsatelliti (1/2476 contro 99/363;  $P < 0,0001$ ). Dei 2476 siti di integrazione dell'HBV nei T, 1981 (80%,  $P < 0,0001$ ) erano localizzati in sequenze ripetute; nello specifico, questi siti di integrazione erano localizzati in maniera significativa nel DNA centromerico ( $P < 0,0001$ ), negli elementi SINE (*short interspersed nuclear elements*), che includono gli elementi Alu ( $P < 0,0001$ ), e nei retrotrasposoni LTR (*long terminal repeat*;  $P < 0,001$ ) [Figura 6]. Nel tessuto tumorale si è osservato, in media, un numero elevato di eventi di integrazione nei lncRNA rispetto ai campioni di tessuto non tumorale. Un elevato numero di integrazioni di HBV è stato riscontrato negli elementi LINE (*long interspersed nuclear element*; 315/2839; 11,1%), in questi ultimi non sono state osservate differenze significative tra campioni di T e NT (280/2476 rispetto a 36/363;  $P = 0,4$ , test  $\chi^2$ ). È interessante notare come nei tessuti di tumore è stato trovato un arricchimento negli integrati di HBV che presentano il gene X all'interno dell'elemento LINE (chimera HBx-LINE).

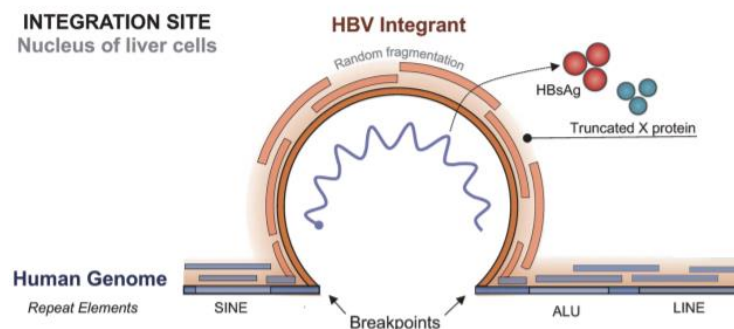


Figura 6. Schema dell'integrazione di HBV negli elementi ripetuti del genoma umano [Zhang et al., 2021]

## Breakpoint nel genoma di HBV

Per quanto riguarda i siti di integrazione del genoma di HBV, l'analisi dei dati ha rilevato che nei tessuti tumorali il 56% (1387/2476) dei *breakpoint* era localizzato nella regione genomica HBx/ENH II/BCP, il 24,3% (601/2476) nel gene preS/S e il 19,7% (488/2476) nella regione genomica preC/C. Nei campioni di tessuto non tumorale il 52,1% (189/363) dei siti di integrazione si trova nel gene che codifica per la proteina HBs, il 39,7% (144/363) nella porzione genomica HBx/ENH II/BCP e l'8,2% (30/363) nel gene preC/C [Figura 7].

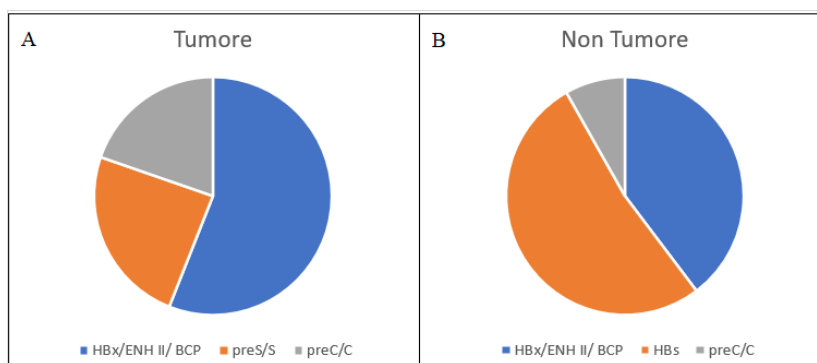


Figura 7. Rappresentazione grafica delle regioni di HBV riscontrate nei siti di integrazione, nei tessuti epatici tumorali (A) e nei tessuti non tumorali (B).

Nello specifico, nei campioni di T, rispetto ai tessuti NT, è stato riscontrato un significativo numero di eventi di integrazione in una regione di 450 bp del genoma di HBV (dal nt 1550 al nt 2000), che include la regione ENH II/BCP [1050/2476 (42,4%) rispetto a 86/363 (23,7%);  $P < 0,0001$ ]. Un altro sito particolarmente ricco di integrazione nei tessuti tumorali si trova tra il nt 1009 e il nt 1200, questa regione include il promotore ENH I/X [297/2476 (12%) rispetto a 18/363 (5%);  $P < 0,0001$ ]. Nei campioni di tessuto NT, i siti di integrazione di HBV sono stati trovati soprattutto in una regione che si estende dal nt 400 al nt 750 del gene S [189/363 (52,1%) rispetto a 602/2476 (24,3%);  $P < 0,0001$ ]. In particolare, un elevato numero di eventi di integrazione nei tessuti non tumorali si è riscontrata della porzione genica "A-determinant" del gene S [100/363 (27,5%) rispetto a 176/2476 (7,1%);  $P < 0,00000$ ].

Per fare luce sul meccanismo implicato nell'integrazione dell'HBV, è stata analizzata la presenza di sequenze di microomologia (MH) tra il DNA nucleare e la porzione di DNA di HBV integrato a livello dei siti di integrazione. È stato rilevato che la maggior parte dei siti unici di integrazione di HBV ha mostrato la presenza di MH sia in campioni di T che NT. In particolare nei T sono state trovate sequenze di MH di 5bp [74,4% (1842/2476;  $P < 0,0001$ ), 9bp [45,7% (1131/2476;  $P < 0,0001$ )],

13bp [22% (545/2476; P<0.0001)] e 15bp [16,8% (416/2476; P<0.0001)], mentre nei NT sono state riscontrate sequenze di MH di 5bp [69,2% (251/363; P<0,0001)], 9bp 48,6% [(176/363; P<0,0001)] e 13 bp [24,6% (89/363; P<0,0001)] [Tabella 1]. La presenza di un arricchimento di sequenze di micromologia nel sito di giunzione tra HBV integrato e il DNA cellulare suggerisce che il meccanismo di riparazione del DNA mediato da MH (MMEJ) potrebbe promuovere l'inserimento virale a livello dei *breakpoint* genomici.

Micromologia (MH)	MH di 5bp	MH di 9bp	MH di 13bp	MH di 15bp
Tumori	1842/2476 (74,4%)	1131/2476 (45,7%)	545/2476 (22%)	416/2476 (16,8%)
Micromologia (MH)	MH di 5bp	MH di 9bp	MH di 13bp	
Non Tumori	251/363 (69,24%)	176/363 (48,6%)	89/363 (24,6%)	

Tabella 1. Percentuale di micromologia (MH) in campioni di tessuti epatici T e NT

## IDENTIFICAZIONE DELLE INTEGRAZIONI NELLE ESPANSIONI CLONALI

Come dimostrato in precedenti studi [N. A. Gillet, 2011. C. C. Berry, 2012. L. B. Cohn, 2015], associando la frammentazione casuale del DNA (generando siti unici di legame ai linker) al sequenziamento paired-end (garantendo così l'identificazione precisa sia del sito di integrazione che dell'estremità frammentata) è possibile identificare le integrazioni soggette ad espansione clonale (siti di integrazione identici con estremità di frammentazione distinte, derivanti dall'espansione clonale di un singolo evento di integrazione) e le integrazioni singole (siti di integrazione unici con un singola estremità frammentata).

È stato stimato che su 2839 siti unici di integrazione virale, 1789 (63%) erano integrazioni soggette ad espansione clonale e la maggior parte di queste è stata trovata nei tessuti tumorali piuttosto che nei tessuti non tumorali [1542/1789 (86,2%) rispetto a 247/1789 (13,8%), P=0,02 *proportion test*]. Le integrazioni soggette ad espansione clonale sono state trovate in tutti i tessuti analizzati. Per valutare se la posizione delle integrazioni di HBV all'interno del genoma umano potesse essere associata all'espansione clonale, è stato effettuato un confronto tra le posizioni genomiche delle integrazioni clonali di HBV e le integrazioni uniche, sia nei tessuti T che NT. Nei tessuti tumorali, le integrazioni di HBV espanse clonalmente sono state trovate soprattutto nella porzione codificante dei geni, mentre le integrazioni uniche sono state rilevate nelle regioni intergeniche. Invece, nei tessuti NT si evince



una distribuzione comparabile delle integrazioni clonali e singole sia nei geni che nelle porzioni intergeniche.

Considerando che la maggior parte delle integrazioni soggette ad espansione clonale sono state individuate nei tessuti tumorali, si può ipotizzare che le integrazioni di HBV si riscontrano all'interno di geni coinvolti nel processo di cancerogenesi. Alla luce di questa informazione, i nostri dati sono stati esaminati al fine di individuare le integrazioni all'interno dei geni associati all'HCC [B. Vogelstein, 2013]. I geni espansi clonalmente recanti integrazione di HBV trovati nel nostro studio sono stati confrontati con quelli recentemente riportati nel Viral Integration Site Data Base (VISDB, disponibile su <https://bioinfo.uth.edu/VISDB>), la raccolta più completa di siti di integrazione virale (VIS) nel genoma umano [D. Tang, 2020]. Dei 5573 geni target dell'HBV associato all'epatocarcinoma cellulare (riportati in VISDB), nei campioni di tessuto provenienti dai 6 pazienti da noi studiati ne sono stati rilevati 323 (274 nei T e 49 nei NT). Dei 323 geni riscontrati, 35 si riscontrano soprattutto in campioni di tessuto tumorale [Tabella 2] e 3 in campioni di tessuto NT [Tabella 3]. Inoltre, singoli eventi di integrazione in campioni individuali sono stati osservati nei seguenti geni *cancer driver*: TERT, KMT2B (MLL4) [Tabella 4] [D. Ding, 2012. A. Y. Fujimoto, 2012. W. K. Sung, 2012. X. Li, 2014. L. H. Zhao, 2016. M. Zapatka, 2020]. Oltre agli eventi di integrazione nei geni precedentemente riportati, sono stati identificati 5 nuovi geni ricorrenti (HPCAL1, KCNQ2, LOXL4, MAL2, NTN5) sia nei campioni di tessuto T che NT. Inoltre, altri 4 geni sono stati target di integrazione virale in modo ricorrente solo nei campioni di tessuto T (ECE1, EVI5L, IL27RA, KIF13B) [Tabella 5].

**Geni target di integrazione virale nei campioni di tessuto tumorale**

Gene	Descrizione
AOAH	Acyloxyacyl hydrolase. Rimuove le catene degli acidi grassi secondari della regione del Lipide-A dei batteri e termina la risposta dell'ospite all'infezione batterica.
ANKS1B	Ankyrin repeat and sterile alpha motif domain-containing protein 1B. Quando over-espressa può avere un ruolo nella patogenesi della malattia di Alzheimer.
CAMTA1	Calmodulin-binding transcription activator 1. Fattore di trascrizione e possibile soppressore tumorale.
CDH4	Cadherin-4. Svolge un ruolo importante durante la segmentazione cerebrale e la crescita neuronale.
CYP2E1	Cytochrome P450 2E1. Monossigenasi coinvolta nella cirrosi epatica, nel cancro e nel metabolismo degli acidi grassi.
DSCAML1	Down syndrome cell adhesion molecule-like protein 1. Molecola di adesione cellulare coinvolta nella differenziazione neuronale.
EXOC4	Exocyst Complex Component 4. Componente di un complesso multiproteico essenziale per il <i>targeting</i> delle vescicole esocitiche.
FCHSD2	FCH And Double SH3 Domains 2. Contribuisce all'internalizzazione dei recettori di superficie e promuove l'endocitosi di EGFR nel tumore.
FSTL4	Follistatin Like 4.
HIVEP3	HIVEP Zinc Finger 3. Membro della famiglia delle proteine leganti l'enhancer del virus dell'immunodeficienza umana di tipo 1 e agisce come fattore di trascrizione.

HSPA12A	Heat Shock Protein Family A (Hsp70) Member 12A. Involved in cellular response to heat stress and cellular senescence.
IGH	Immunoglobulin Heavy Locus.
IQSEC1	IQ motif and SEC7 domain-containing protein 1. Proteina coinvolta nello sviluppo neuronale.
JARID2	Jumonji, AT Rich Interactive Domain 2. Partecipa alla regolazione negativa del segnale di proliferazione cellulare.
KCNQ3	Potassium Voltage-Gated Channel Subfamily Q Member 3. Proteina coinvolta nella regolazione dell'eccitabilità neuronale.
LMF1	Lipase maturation factor 1. Coinvolta nella maturazione di proteine specifiche del reticolo endoplasmatico e nel metabolismo delle lipoproteine.
MED26	Mediator Complex Subunit 26. Coinvolto regolazione/acetilazione della cromatina.
TGFBRAP1	Transforming Growth Factor Beta Receptor Associated Protein 1. Proteina che agisce come chaperone ed è coinvolta nell'associazione segnale-dipendente con SMAD4.
MLPH	Melanophilin. Coinvolta nel trasporto del melanosoma e nel cancro della vescica.
MPG	N-Methylpurine DNA Glycosylase. Coinvolto nel processo biologico del danno e della riparazione del DNA.
NFAT5	Nuclear Factor Of Activated T Cells 5. Ha un ruolo nella trascrizione genica durante la risposta immunitaria.
NCOR2	Nuclear Receptor Corepressor 2. Recettore nucleare co-repressore, un'espressione anomala di questo gene è associato ad alcune tipologie di tumori.
PLD5	Phospholipase D Family Member 5.
PTPRG	Protein Tyrosine Phosphatase Receptor Type G. Regola il ciclo cellulare, la differenziazione e la trasformazione oncogenica.
PTPRM	Protein Tyrosine Phosphatase Receptor Type M. Regola il ciclo cellulare, la differenziazione e la trasformazione oncogenica.
TRAPPC9	Trafficking Protein Particle Complex 9. Potrebbe essere coinvolta nella differenziazione delle cellule neuronali ea avere un ruolo nel trasporto vescicolare dal RE all'apparato del Golgi.
SLC29A3	Solute Carrier Family 29 Member 3. Svolge un ruolo nell'assorbimento cellulare di nucleosidi e dei loro analoghi.
SDCCAG8	Serologically Defined Colon Cancer Antigen 8. Coinvolto nell'organizzazione del centrosoma durante l'interfase e la mitosi.
SMOC1	SPARC Related Modular Calcium Binding 1. Svolge un ruolo essenziale nello sviluppo dell'occhio e degli arti, ed è un probabile regolatore della differenziazione degli osteoblasti.
SYNGR1	Synaptogyrin 1. Coinvolto nella regolazione della plasticità sinaptica a breve e lungo termine.
WDR66	Cilia And Flagella Associated Protein 251. Proteine coinvolte nella motilità degli spermatozoi e della regolazione della motilità ciliare.
WNK2	Protein kinase lysine-deficient 2. Regolazione dell'omeostasi degli elettroliti, segnalazione cellulare, sopravvivenza e proliferazione; regola negativamente la progressione del ciclo cellulare.
ZHX2	Zinc fingers and homeoboxes protein 2. Sopprime la differenziazione delle cellule progenitrici dei neuroni.
ZMIZ1	Zinc Finger MIZ-Type Containing 1. Regola l'attività di vari fattori di trascrizione.
ZNF536	Zinc Finger Protein 536. Regola negativamente la differenziazione neuronale.

Tabella 2. Geni target delle integrazioni virali riscontrate nei tessuti di campioni tumorale.

**Geni target di integrazione virale nei campioni di tessuto non tumorale**

Gene	Descrizione
IGH	Immunoglobulin Heavy Locus. Riconosce antigeni estranei e avvia le risposte immunitarie come la fagocitosi e il sistema del complemento.
MPG	N-Methylpurine DNA Glycosylase. Coinvolto nel processo biologico di danno e riparazione del DNA.
NRG3	Neuregulin 3. Influenza la proliferazione, la migrazione e la differenziazione dei neuroblasti.

Tabella 3. Geni target delle integrazioni virali riscontrate nei tessuti di campioni non tumorale.

**Geni cancer driver**

Gene	Descrizione
TERT	Telomerase Reverse Transcriptase. Subunità catalitica dell'enzima telomerase, è coinvolta nella senescenza cellulare e partecipa alla riparazione dei cromosomi.
KMT2B (MLL4)	Lysine Methyltransferase 2B. Proteina contenente multipli domini di legame (inclusi i domini di legame allo zinco).

Tabella 4. Geni target delle integrazioni virali coinvolti nel meccanismo di cancerogenesi.

**Nuovi geni target delle integrazioni virali**

Gene	Descrizione
HPCAL1	Hippocalcin Like 1. Può essere coinvolta nella segnalazione neuronale nel SNC.
MAL2	Mal, T Cell Differentiation Protein 2. Proteina transmembrana richiesta per la trascrizione.
ECE1	Endothelin Converting Enzyme 1. Coinvolto nel trattamento proteolitico dei precursori dell'endotelina in peptidi biologicamente attivi.
EVI5L	Ecotropic Viral Integration Site 5 Like. Funziona come una proteina attivante la GTPase.
KIF13B	Kinesin Family Member 13B. Coinvolto nella riorganizzazione del citoscheletro e regola la formazione degli assoni.
KCNQ2	Potassium Voltage-Gated Channel Subfamily Q Member 2. Svolge un ruolo critico nella regolazione dell'eccitabilità neuronale.
LOXL4	Lysyl Oxidase Like 4. Coinvolto nella regolazione dello sviluppo, nella senescenza, nella soppressione del tumore e nel controllo della crescita cellulare.
NTN5	Netrin 5. Svolge un ruolo nella neurogenesi.
IL27RA	Interleukin 27 Receptor Subunit Alpha. Media la trasduzione del segnale in risposta a IL27 ed è coinvolto nella regolazione delle risposte immunitarie di tipo Th1.

Tabella 5. Nuovi geni target delle integrazioni virali.

## DISCUSSIONI

Il metodo di sequenziamento da noi utilizzato (HBIS, *HBV Integration Sequencing*) ha permesso di eseguire un'analisi su larga scala dei siti di integrazione del DNA dell'HBV in campioni di tessuto epatico tumorale e non tumorale. Questo approccio ci ha consentito di rilevare un numero molto elevato di siti d'integrazioni HBV, e ci ha permesso di definire i siti di integrazione preferenziali all'interno del genoma umano. Inoltre, in accordo con altri studi [W. K. Sung, 2012. L. H. Zhao, 2016], abbiamo riscontrato un numero di eventi di integrazione più alto nei tessuti tumorali rispetto ai tessuti non tumorali.

È noto che gli eventi di integrazione di HBV avvengono in maniera randomica all'interno dell'intero genoma umano. Tuttavia, recenti studi hanno evidenziato regioni genomiche in cui le integrazioni avvengono con maggiore frequenza. Tant'è che si sta ampliando l'elenco dei geni umani coinvolti dall'evento d'integrazione.

Il nostro studio ha consentito di dimostrare che la maggior parte dei *breakpoint* di HBV nell'HCC sono siti a livello dei geni codificanti, suggerendo che gli eventi di integrazione sono verosimilmente associati ad uno stato conformazionale aperto della cromatina, all'interno della quale il virus può integrarsi in maniera più efficiente.

È noto che l'integrazione di HBV può modificare la produzione dei lncRNA, variando di conseguenza anche il loro ruolo sulla regolazione dell'espressione di geni codificanti proteine che possono avere un ruolo nella carcinogenesi [X. Yang, 2017. M. Ghidini, 2015]. Dall'analisi dei nostri risultati si è riscontrato un elevato numero di siti d'integrazione proprio a livello dei lncRNA in campioni di tessuti epatici tumorali. Inoltre, tali integrazioni sono risultate molto più abbondanti nei tessuti tumorali rispetto ai non tumorali. Come descritto in altri studi [Z. Kan, 2013. Y. Murakami, 2005. D. Ding, 2012. W. K. Sung, 2012] si è riscontrata la presenza di integrazioni virali in prossimità di elementi genetici ripetuti e non codificanti, come le sequenze LINE, SINE e LTR. In particolare abbiamo rilevato la presenza della chimera HBx-LINE, di cui è nota la capacità di produrre un RNA ibrido che funge da lncRNA. La presenza di tale lncRNA è associata ad una bassa sopravvivenza dei pazienti con HCC correlato ad infezione da HBV. L'espressione di HBx-LINE influisce sull'espressione della  $\beta$ -catenina sottolineando un suo ruolo nell'attivazione del pathway di segnalazione di Wtn. [C. Lau, 2014].

Tra le diverse regioni genomiche di HBV, le sequenze riscontrate più frequentemente negli integrati virali includono quelle corrispondenti alla regione PreS/S, all'ENH I/X promoter, e all'X deleto in -

3'. L'ENH I è noto per la sua capacità di indurre la produzione di trascritti HBx [M. Shamay, 2001]. Inoltre, le sequenze HBx nei casi di virus integrato sono spesso troncate e svolgono un ruolo importante nello sviluppo dell'HCC: possono infatti stimolare la proliferazione indipendente dal fattore di crescita, la resistenza all'inibizione della crescita, l'invasione e la metastasi dei tessuti, l'angiogenesi, la riprogrammazione del metabolismo energetico e la resistenza all'apoptosi.

Numerosi report sottolineano il ruolo delle sequenze di microomologia (MH) tra il DNA cellulare e quello dell'HBV a livello dei siti di integrazione. Esistono evidenze che confermano l'importanza di queste sequenze MH e il loro coinvolgimento nel processo di integrazione dell'HBV. Nel nostro studio abbiamo trovato un arricchimento delle sequenze di MH tra il genoma umano e il DNA virale integrato nei siti di *breakpoint*, questo suggerisce che la MH potrebbe aver giocato un ruolo importante nell'integrazione dell'HBV nei casi studiati.

Di fondamentale importanza, il metodo di sequenziamento HBIS ha permesso di caratterizzare le integrazioni in base alla loro clonalità, riuscendo in questo modo a discriminare con precisione gli eventi di integrazione clonale da quelli definiti unici. In accordo con gli studi che descrivono la presenza dell'espansione clonale delle cellule durante la carcinogenesi, abbiamo identificato più eventi di integrazione clonale in campioni di tessuto tumorale piuttosto che nei tessuti epatici non tumorali [W. Sung, 2012. L. Zhao, 2016. M. Furuta, 2018. L. Yang, 2018]. La presenza di integrazioni clonali nei tessuti non tumorali potrebbe essere spiegata da una pressione selettiva indotta dall'infiammazione cronica e dalla comparsa di noduli cirrotici nel fegato [M. A. Budzinska, 2018. W. S. Mason, 2016. T. Tu, 2015].

Il genoma di HBV, grazie alla sua capacità di integrarsi frequentemente in regioni centromeriche o telomeriche, è coinvolto in complessi riarrangiamenti strutturali che coinvolgono soprattutto i locus di TERT e di MLL4. Possiamo dunque affermare che, nei tessuti tumorali che presentano riarrangiamenti associati ad HBV, questi ultimi si trovano in prossimità di geni definiti "*cancer-driver*". Inoltre, le integrazioni di HBV possono guidare lo sviluppo tumorale anche alterando il gene più vicino al sito di *breakpoint* attraverso la mutagenesi inserzionale. Il nostro studio ha confermato che i geni TERT e MLL4 sono tra i principali siti *hotspot* d'integrazione dell'HBV nei casi di HCC, in quanto la maggior parte delle integrazioni in questi geni presentava espansione clonale.

In conclusione, i riarrangiamenti strutturali associati all'integrazione di HBV sono considerati dei meccanismi in grado di guidare la carcinogenesi alterando i geni *cancer-driver* sia a distanza sia attraverso la *cis*-attivazione della mutagenesi inserzionale. Questo sottolinea l'eterogeneità dell'HCC associato ad infezione da HBV e l'importanza della caratterizzazione molecolare al fine di identificare nuove opportunità terapeutiche specifiche.

## BIBLIOGRAFIA

- ❖ P. Karayiannis. Hepatitis B virus: virology, molecular biology, life cycle and intrahepatic spread. *Hepatol. Int.* 11, 500–508. 2017
- ❖ H. B. El-Serag. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* 142(6): 1264-1273 e1261. 2012
- ❖ T. Pollicino, I. Cacciola, F. Saffioti, G. Raimondo. Hepatitis B virus Pre/S gene variants: Pathobiology and clinical implication. *Journal of Hepatology.* vol. 61, 408–417. 2014
- ❖ A. Kramvis. Genotypes and Genetic Variability of Hepatitis B Virus. *Intervirology* 57, 141–150. 2014
- ❖ A. O. Santos et al. Characterization of Hepatitis B virus (HBV) genotypes in patients from Rondônia, Brazil. *Viol. J.* 7, 315. 2010
- ❖ S. Schaefer. Hepatitis B virus taxonomy and hepatitis B virus genotypes. *World J. Gastroenterol.* 13, 14. 2007
- ❖ K. Watashi, S. Urban, W. Li, T. Wakita. NTCP and Beyond: Opening the Door to Unveil Hepatitis B Virus Entry. *Int. J. Mol. Sci.* 15, 2892–2905. 2014
- ❖ H. Y. Kang, S. Lee, S.G. Park, J. Yu, Y. Kim, G. Jung. Phosphorylation of Hepatitis B virus Cp at Sert 87 facilitates core assembly. *Biochem J.* 311-317. 2006
- ❖ J. Kock, H. J. Schlicht. Analysis of the earliest steps of hepadnavirus replication: genome repair after infectious entry into hepatocytes does not depend on viral polymerase activity. *J Virol.* 67(8), 4867-4874. 1993
- ❖ P. Karayiannis. Hepatitis B virus: virology, molecular biology, life cycle and intrahepatic spread. *Hepatol. Int.* 11, 500–508. 2017
- ❖ J. R. Pollac, D. Ganem. An RNA stem-loop structure directs hepatitis B virus genomic RNA encapsidation. *J. Virol.* 67:3254-3263. 1993
- ❖ G. H. Wang, F. Zoulim, E. H. Leber, J. Kitson, C. Seeger. Role of RNA in enzymatic activity of the reverse transcriptase of hepatitis B viruses. *J. Virol.* 68:8437-8442. 1994
- ❖ J. Hu, D. O. Toft, C. Seeger. Hepadnavirus assembly and reverse transcription require a multi-component chaperone complex which is incorporated into nucleocapsids. *EMBO J.* 16:59–68. 1997
- ❖ G. Raimondo, T. Pollicino, I. Cacciola, G. Squadrito, Occult hepatitis B virus infection. *J. Hepatol.* 46, 160-170. 2007

- ❖ T. M. Block, A. S. Mehta, C. J. Fimmel, R. Jordan. Molecular viral oncology of hepatocellular carcinoma. *Molecular viral oncology* 22(33):5093-107. 2003
- ❖ M. M: Minor, B. L. Slagle. Hepatitis B virus interaction with the Ubiquitin Proteasome System. *Viruses*, 6, 4683—4702. 2014
- ❖ C. Seegeer, W. S. Mason. Hepatite B Virus Biology. *Microbiology and Molecular Biology Reviews*, Mar. p. 51-68. 2000
- ❖ H. Norder, A. M. Courouce, L. O. Magnius. Molecular basis of hepatitis B virus serotype variations within the four major subtypes. *J Gen Virol*; 73:3141–3145. 1992
- ❖ D. N. Standring, J. H. Ou, F. R. Masiarz, W. J. Rutter. A signal peptide encoded within the precore region of hepatitis B virus directs the secretion of a heterogeneous population of e antigen in *Xenopus* oocytes. *Proc Natl Acad Sci USA*, 85: 8405-8409. 1988
- ❖ C. Chang, S. Zhou, D. Ganem, D. N. Standring. Phenotyping mixing between different hepadnavirus nucleocapsid protein reveals C protein dimerization to be cis preferential. *J. Virol.* 68:5225-5231. 1994
- ❖ R. A. Crowther, N. A. Kiselev, B. Bottcher, J. A. Berriman, G. P. Borisova, V. Ose, P. Pumpens. Three dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell* 77:943-950. 1994
- ❖ Nassal M. Hepatitis B viruses: reverse transcription a different way. *Virus Res* 134:235–249 2008 2008
- ❖ H. B. El-Searag. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology*. 142: 1264-73. 2012
- ❖ M. Levrero, J. Zucman-Rossi. Mechanism of HBV-induced hepatocellular carcinoma. *Journal of hepatology*. 64:S84-S101. 2016
- ❖ K. Schulze, J. C. Nault, A. Villanueva. Genetic profiling of hepatocellular carcinoma using next-generation sequencing. *J Hepatol* 65: 1031-1042. 2016
- ❖ F. Guerrieri, L. Belloni, N. Pediconi, M. Levrero. Molecular mechanisms of HBV-associated hepatocarcinogenesis. *Semin Liver Dis.* 33:147–156. 2013
- ❖ J. C. Nault, J. Zucman-Rossi. Genetics of hepatocellular carcinoma: the next generation. *J Hepatol.* 60:224–226. 2014
- ❖ J. Zucman-Rossi, A. Villanueva, J. C. Nault, J. M. Llovet. Genetic landscape and biomarkers of hepatocellular carcinoma. *Gastroenterology*. 149:1226–1239. 2015
- ❖ S. H. Ahn, Y. N. Park, J. Y. Park, et al. Long-term clinical and histological outcomes in patients with spontaneous hepatitis B surface antigen seroclearance. *J Hepatol.* 42:188–194. 2005

- ❖ H. Clevers, R. Nusse. Wnt/beta-catenin signaling and disease. *Cell* 149:1192-205, 2012
- ❖ M. J. Bouchard, S. Navas-Martin. Hepatitis B and C virus hepatocarcinogenesis: lessons learned and future challenges. *Cancer Lett.* 305. 2011
- ❖ K. Yaginuma, H. Kobayashi, M. Kobayashi, T. Morishima, K. Matsuyama and K. Koike. Multiple integration site of hepatitis B virus DNA in hepatocellular carcinoma and chronic active hepatitis tissues from children. *J Virol* 61(6): 1808-1813. 1987
- ❖ H. Sezaki, M. Kobayashi, T. Hosaka, T. Someya, N. Akuta, F. Suzuki, A. Tsubota, Y. Suzuki, S. Saitoh, Y. Arase, K. Ikeda, M. Kobayashi, M. Matsuda, K. Takagi, J. Sato and H. Kumada. Hepatocellular carcinoma in noncirrhotic young adult patients with chronic hepatitis B viral infection. *J Gastroenterol* 39(6): 550-556. 2004
- ❖ Y. Murakami, K. Saigo, H. Takashima, M. Minami, T. Okanoue, C. Brechot and P. Paterlini-Brechot. Large scaled analysis of hepatitis B virus (HBV) DNA integration in HBV related hepatocellular carcinomas. *Gut* 54(8): 1162-1168. 2005
- ❖ W. K. Sung, H. Zheng, S. Li, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet.* 44:765–769. 2012
- ❖ R. Bonilla Guerrero, L. R. Roberts. The role of hepatitis B virus integrations in the pathogenesis of human hepatocellular carcinoma. *J Hepatol* 42(5): 760-777. 2005
- ❖ D. Ding, X. Lou, D. Hua, W. Yu, L. Li, J. Wang, F. Gao, N. Zhao, G. Ren, L. Li and B. Lin. Recurrent targeted genes of hepatitis B virus in the liver cancer genomes identified by a next-generation sequencing-based approach. *PLoS Genet* 8(12): e1003065. 2012
- ❖ A. Fujimoto, Y. Totoki *et al.* Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet* 44(7): 760-764. 2012
- ❖ W. K. Sung, H. Zheng *et al.* Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet* 44(7): 765-769. 2012
- ❖ C. C. Lau, T. Sun, A. K. Ching, M. He, J. W. Li, A. M. Wong, N. N. Co, A. W. Chan, P. S. Li, R. W. Lung, J. H. Tong, P. B. Lai, H. L. Chan, K. F. To, T. F. Chan, N. Wong. Viral-human chimeric transcript predisposes risk to liver cancer development and progression. *Cancer Cell* 25(3): 335-349. 2014
- ❖ L. H. Zhao, X. Liu, *et al.* Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nat Commun* 7: 12992. 2016
- ❖ T. Tu, M. A. Budzinska, N. A. Shackel, S. Urban. HBV DNA Integration: Molecular Mechanisms and Clinical Implications. *Viruses* 9(4). 2017



- ❖ F. Guerrieri, L. Belloni, D. D'Andrea. Genome-wide identification of direct HBx targets that control HBV replication. *Gastroenterology* 2015
- ❖ M. Ozturk, A. Arslan-Ergul, S. Bagislar, et al. Senescence and Immortality in Hepatocellular Carcinoma. *Cancer Lett.* 286:103–113. 2009]
- ❖ J. Zhao, G. Wu, F. Bu, et al. Epigenetic silence of ankyrin-repeat-containing, SH3-domain-containing, and proline-rich-region-containing protein 1 (ASPP1) and ASPP2 genes promotes tumor growth in hepatitis B viruspositive hepatocellular carcinoma. *Hepatology.* 51:142–153. 2010
- ❖ S.H. Park, J.K. Jung, J.S. Lim, I. Tiwari, K.L. Jang. Hepatitis B virus X protein overcomes all-trans retinoic acid-induced cellular senescence by downregulating levels of p16 and p21 via DNA methylation. *J Gen Virol.* 92:1309–1317. 2011
- ❖ F.V. Chisari, K. Klopchin, T. Moriyama, et al. Molecular pathogenesis of hepatocellular carcinoma in hepatitis B virus transgenic mice. *Cell.* 59:1145–1156. 1989
- ❖ J.H. Hung, I.J. Su, H.Y. Lei, H.C. Wang, W.C. Lin, W.T. Chang, et al. Endoplasmic reticulum stress stimulates the expression of cyclooxygenase-2 through activation of NF-kappaB and pp38 mitogen-activated protein kinase. *J Biol Chem.* 279:46384–46392. 2004
- ❖ L.H. Wang, W. Huang, M.D. Lai, I.J. Su. Aberrant cyclin A expression and centrosome overduplication induced by hepatitis B virus pre-S2 mutants and its implication in hepatocarcinogenesis. *Carcinogenesis.* 33:466–472. 2012
- ❖ L. de La Coste, B. Romagnolo, P. Billuart, et al. Somatic mutations of the betacatenin gene are frequent in mouse and human hepatocellular carcinomas. *Proc Natl Acad Sci USA.* 95:8847–8851. 1998
- ❖ Y. Guo, W. Kang, X. Lei, et al. Hepatitis B viral core protein disrupts human host gene expression by binding to promoter regions. *BMC Genomics.* 13:563. 2012
- ❖ C. A. Bill, J. Summers. Genomic DNA double-strand breaks are targets for hepadnaviral DNA integration. *Proc Natl Acad Sci U S A* 101(30): 11135-11140. 2004
- ❖ J. C. Edman, P. Gray, P. Valenzuela, L. B. Rall, W. J. Rutter. Integration of Hepatitis B Virus Sequences and Their Expression in a Human Hepatoma Cell. *Nature* 286 (5772):535-38. 1980
- ❖ S. Wirth, B. Zabel. Molecular hybridization techniques in current diagnosis of chronic hepatitis B in childhood. *Eur. J. Pediatr.* 151, 243–249. 1992
- ❖ K. Kitada, S. Aikawa, S. Aida. Alu-Alu Fusion Sequences Identified at Junction Sites of Copy Number Amplified Regions in Cancer Cell Lines. *Cytogenet. Genome Res.* 139, 1–8. 2013
- ❖ M.A. Budzinska, N.A. Shackel, S. Urban, T. Tu. Sequence analysis of integrated hepatitis B virus DNA during HBeAg-seroconversion. *Emerg. Microbes Infect.* 7, 1–12. 2018

- ❖ T. Tu, A.R. Jilbert. Detection of Hepatocyte Clones Containing Integrated Hepatitis B Virus DNA Using Inverse Nested PCR. *Hepatitis B Virus*. (eds. Guo, H. & Cuconati, A.) 1540, 97–118. 2017]
- ❖ R. Michael, J. Sambrook. *Inverse Polymerase Chain Reaction (PCR)*. Cold Spring Laboratory Press. 2019
- ❖ Z. Kan et al. Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res*. 23, 1422–1433. 2013
- ❖ The Cancer Genome Atlas Research Network. Integrated genomic and molecular characterization of cervical cancer. *Nature* 543, 378–384. 2017
- ❖ S. K. Ladner, M. J. Otto, C. S. Barker, K. Zaifert, G. H. Wang, J. T. Guo, C. Seeger, R. W. King. Inducible expression of human hepatitis B virus (HBV) in stably transfected hepatoblastoma cells: a novel system for screening potential inhibitors of HBV replication. *Antimicrob Agents Chemother* 41(8): 1715-1720. 1997
- ❖ A. M. Bolger, M. Lohse, B. Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15): 2114-2120. 2014
- ❖ H. Li, R. Durbin. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14): 1754-1760. 2009
- ❖ A. R. Quinlan. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* 47: 11.12.11-34. 2014
- ❖ X. Huang, A. Madan. CAP3: A DNA sequence assembly program. *Genome Res* 9(9): 868-877. 1999
- ❖ L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23): 3150-3152. 2012
- ❖ C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden. BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421. 2009
- ❖ L. Cohn, I. T. Silva, T. Y. Oliveira, R. A. Rosales, E. H. Parrish, G. H. Learn, B. H. Hahn, J. L. Czartoski, M. J. McElrath, C. Lehmann, F. Klein, M. Caskey, B. D. Walker, J. D. Siliciano, R. F. Siliciano, M. Jankovic and M. C. Nussenzweig. HIV-1 integration landscape during latent and active infection. *Cell* 160(3): 420-432. 2015
- ❖ W. K. Sung *et al.* Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet* 44(7): 765-769. 2012
- ❖ L. H. Zhao *et al.* Wang. Genomic and oncogenic preference of HBV integration in hepatocellular carcinoma. *Nat Commun* 7: 12992. 2016

- ❖ N. A. Gillet, N. Malani, A. Melamed, N. Gormley, R. Carter, D. Bentley, C. Berry, F. D. Bushman, G. P. Taylor, C. R. Bangham. The host genomic environment of the provirus determines the abundance of HTLV-1-infected T-cell clones. *Blood* 117(11): 3113-3122. 2011
- ❖ C. C. Berry, N. A. Gillet, A. Melamed, N. Gormley, C. R. Bangham and F. D. Bushman. Estimating abundances of retroviral insertion sites from DNA fragment length data. *Bioinformatics* 28(6): 755-762. 2012
- ❖ B. Vogelstein, N. Papadopoulos, V.E. Velculescu, S. Zhou, L. A. Diaz Jr, K. W. Kinzler. Cancer Genome Landscapes. *Science*. Mar 29;339(6127):1546-58. 2013
- ❖ D. Tang, B. Li, T. Xu, R. Hu, D. Tan, X. Song, P. Jia, Z. Zhao. VISDB: a manually curated database of viral integration sites in the human genome. *Nucleic Acids Res* 48(D1): D633-D641. 2020
- ❖ X. Li, J. Zhang, Z. Yang, J. Kang, S. Jiang, T. Zhang, T. Chen, M. Li, Q. Lv, X. Chen, M. A. McCrae, H. Zhuang, F. Lu. The function of targeted host genes determines the oncogenicity of HBV integration in hepatocellular carcinoma. *J Hepatol* 60(5): 975-984. 2014
- ❖ M. Zapatka, I. Borozan, D. S. Brewer, M. Iskar, A. Grundhoff, M. Alawi, N. Desai, H. Sultmann, H. Moch, P. Pathogens, C. S. Cooper, R. Eils, V. Ferretti, P. Lichter, P. Consortium. The landscape of viral associations in human cancers. *Nat Genet* 52(3): 320-330. 2020
- ❖ X. Yang *et al.* Distinct hepatitis B virus integration patterns in hepatocellular carcinoma and adjacent normal liver tissue. *Int. J. Cancer* **140**, 1324–1330. 2017
- ❖ M. Ghidini, C. Braconi. Non-Coding RNAs in Primary Liver Cancer. *Front. Med.* **2**. 2015
- ❖ D. Ding *et al.* Recurrent Targeted Genes of Hepatitis B Virus in the Liver Cancer Genomes Identified by a Next-Generation Sequencing–Based Approach. *PLoS Genet.* **8**, e1003065. 2012
- ❖ M. Shamay, R. Agami, Y. Shaul. HBV integrants of hepatocellular carcinoma cell lines contain an active enhancer. *Oncogene* **20**, 6811–6819. 2001
- ❖ M. Furuta, H. Tanaka, Y. Shiraishi, *et al.* Characterization of HBV integration patterns and timing in liver cancer and HBV-infected livers. *Oncotarget.* 9:25075–88. 2018
- ❖ L. Yang, S. Ye, X. Zhao, *et al.* Molecular characterization of HBV DNA integration in patients with hepatitis and hepatocellular carcinoma. *J Cancer.* 9:3225–35. 2018
- ❖ W. S. Mason, U. S. Gill, S. Litwin, *et al.* HBV DNA integration and clonal hepatocyte expansion in chronic hepatitis B patients considered immune tolerant. *Gastroenterology.* 151:986–98. 2016

- ❖ T. Tu, M. A. Budzinska, N. A. Shackel, *et al.* Conceptual models for the initiation of hepatitis B virus-associated hepatocellular carcinoma. *Liver Int.* 35:1786–800. 2015
- ❖ D. Zhang, K. Zhang, U. Protzer, C. Zeng. HBV Integration Induces Complex Interactions between Host and Viral Genomic Functions at the Insertion Site. *Journal of Clinical and Translational Hepatology.* vol. 9(3) | 399–408. 2021