



Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>



Universitat Autònoma de Barcelona

Programa de Doctorado en Medicina

Departamento de Medicina

Tesis Doctoral

**ESTUDIO DE LA QUASIESPECIES DEL VIRUS DE LA
HEPATITIS DELTA: ANÁLISIS DE LA REGIÓN DEL
ANTÍGENO DELTA Y DE LA RIBOZIMA**

Autora de la Tesis Doctoral

SARA SOPENA SANTISTEVE

Directores de la Tesis Doctoral

Dra. María Asunción Buti Ferret

Dr. David Tabernero Caellas

Tutora académica de la Tesis Doctoral

Dra. María Asunción Buti Ferret

Universitat Autònoma de Barcelona. Barcelona, 2020

AGRADECIMIENTOS

Hace ya cuatro años que empecé esta aventura de la tesis doctoral y nunca me hubiera imaginado que estaría escribiendo esto en medio de una pandemia. Durante este tiempo han pasado muchas personas por el camino, por eso en primer lugar, a todas aquellas personas que han aportado su granito de arena en este proyecto, aunque no salgan explícitamente sus nombres aquí escritos, les quería dar las gracias.

A continuación agradecer a mis directores, la Dra. Maria Buti y el Dr. David Tabernero. A Maria darle las gracias por hacerme sentir una más de su grupo, por depositar tanta confianza en mí y por aportar siempre su punto de vista clínico. A David darle las gracias por su apoyo durante todo este tiempo, por su gran dedicación en este trabajo y por estar al pié del cañón estos últimos meses con las correcciones, con las dudas, con los imprevistos de último momento... Sin su ayuda esto no hubiera sido posible. También agradecer al Dr. Francisco Rodríguez, por acogerme en su laboratorio para que realizara las prácticas del máster y así poder empezar en el camino de la investigación. Gracias a sus ideas, sus hipótesis, sus teorías..., a veces difíciles de entender, pero que siempre acaban siendo grandes descubrimientos para seguir avanzando en la ciencia. Asimismo agradecer al Dr. Rafael Esteban, sin su prestigio y reconocimiento sería imposible ser el grupo de investigación que somos hoy en día.

Además agradecer a Josep Gregori, bioinformático del grupo, por todos los análisis realizados en esta tesis, por su paciencia, por su dedicación y por todos los conocimientos aportados. También agradecer a la Dra. Maria Francesca Cortese, compañera de trabajo, por su aportación estadística en esta tesis, por sus ideas y por sus grandes ganas de trabajar día a día.

Continuando con mis compañeros de trabajo durante este tiempo, el "Hepatitis Team", gracias a Chari, a Irene, a Cristina, a Marçal, a Marta, a Selene, a Gerard, a Unai y a Adrián. Por hacer que el día a día se haga más ameno, por estar ahí en cualquier ocasión que lo he necesitado y porque sin su apoyo esto no hubiera sido posible. Además me gustaría agradecer de manera especial a Irene, Cristina y Marçal. A Irene porque además de compañera de trabajo se ha convertido en una gran amiga que siempre está ahí para todo lo que necesitas y que ha sido un gran apoyo. A Cristina, mi "partner in crime", otra de las grandes amistades que me llevo de todo esto, gracias por ayudarme siempre incondicionalmente. Y a Marçal por haber compartido esta experiencia de escribir la tesis a la misma vez y desahogarnos mutuamente.

Quiero también dedicar un agradecimiento a la Dra. Mar Riveiro, por su eficacia, por su capacidad de resolución y por su dedicación en el trabajo. Además agradecer a Adriana

Palom, por la facilidad a la hora de trabajar con ella, por su amabilidad y por ser tan buena compañera de trabajo.

Además, agradecer también al Dr. Josep Quer y a todo su grupo, en especial a Elena Vargas, Sofia Piriz, Meritxell Llorens y Mercedes Guerrero. Por hacer que los congresos siempre sean más divertidos y en general por todos esos momentos que hemos podido compartir gracias a la hepatitis.

Por último, pero por supuesto no menos importante, agradecer a toda mi familia. En especial a mi madre y a mi hermano, dos piezas clave en mi vida, que siempre han estado ahí, para lo bueno y para lo malo, por superar juntos todas las adversidades que a veces esta vida te pone por delante y también por siempre estar presentes en los mejores momentos de mi vida. Y también a Eloi, porque no podría tener a una mejor persona a mi lado, por ser un pilar esencial en mi vida ahora mismo, por su paciencia, por hacer que las dificultades se conviertan en soluciones en cuestión de segundos y por querer compartir la vida conmigo. Además también agradecer a todos mis amigos, en especial a Amalia, Jenni, Yolanda, Sheila, Raquel, Marta, Maria y Uxue. Porque durante estos cuatro años también hace falta una buena fiesta, una buena cerveza y por supuesto una muy buena compañía.

¡Muchas gracias a todos de corazón!

ABREVIACIONES

A

ADAR1	Adenosina desaminasa específica de ARN 1 (<i>Adenosine deaminases acting on RNA 1</i>)
ADF	Adefovir
ADN	Ácido desoxirribonucleico
ADNc	ADN complementario
ADNccc	ADN circular cerrado covalentemente
ADNds	ADN bicatenario
ARNsi	ARN de interferencia pequeño (<i>Small interfering RNA</i>)
ADNss	ADN de cadena simple
ALT	Alanina transaminasa
APOBEC	Enzimas catalíticas similares a polipéptidos de edición de ARNm apolipoproteína B (<i>Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like</i>)
APS	Adenosina-5'-fosfosulfato
ARM	Motivo rico en argininas (<i>Arginine rich motive</i>)
ARN	Ácido ribonucleico
ARNm	ARN mensajero
AST	Aspartato transaminasa
ATP	Adenosina trifosfato
AU	<i>Approximately Unbiased P-value</i>

B

BEA score	<i>Baseline-Event-Anticipation score</i>
BLV	Bulevirtide
BP	<i>Bootstrap Probability value</i>

C

Ca	Calcio
----	--------

CCD	Dispositivo de carga acoplada (Charge-coupled device)
CCS	<i>Coiled-Coil Sequence</i>
CI	Contenido de información
Cl	Cloruro
CLIA	Inmunoensayo de quimioluminiscencia indirecta
Cq	Ciclo de cuantificación
Cu	Cobre
C-terminal	Carboxilo terminal

D

dNTP	Desoxiribonucleótido trifosfato
ddNTP	Didesoxiribonucleótidos trifosfato
DTT	Ditiotreitol

E

EMA	Agencia Europea de Medicamentos (<i>European Medicines Agency</i>)
emPCR	PCR de emulsión
Enh	Secuencia potenciadora (<i>Enhancer</i>)
ETV	Entecavir

F

FDA	Administración Estadounidense de Alimentos y Medicamentos (U.S. Food and Drug Administration)
FRET	<i>Florescence Resonance Energy Transfer</i>
Fw	Forward

G

gr	Gramos
----	--------

H

HBeAg	Antígeno e de la hepatitis B
HBsAg	Antígeno de superficie de la hepatitis B
HCl	Ácido clorhídrico
HDAg	Antígeno de la hepatitis delta
HLH	<i>Helix-loop-helix</i>
HSPG	Proteoglicanos heparán sulfatos (<i>Heparan sulfate proteoglycan</i>)

I

IFN	Interferon
IFN- α	Interferon alfa
IFN- λ	Interferon lambda
IgG	Inmunoglobulina G
IgM	Inmunoglobulina M
INR	Índice Internacional Normalizado
ISG	Genes estimulados por interferón (<i>Interferon Stimulated Genes</i>)
IQR	Rango intercuartílico (<i>Interquartile Range</i>)

J

Jak	Janus quinasa
-----	---------------

K

kDA	Kilodalton
kpb	Kilo pares de bases

L

L	Litro
LB	Luria-Bertani
L-HBsAg	Proteína del antígeno de superficie de la hepatitis B grande

L-HDAg Antígeno largo del virus de la hepatitis delta

LNF Lonafarnib

log Logaritmo

M

M-HBsAg Proteína del antígeno de superficie de la hepatitis B mediana

MID Secuencias de Identificador Multiplex (Multiplex Identifier Sequences)

Mg Magnesio

mg Miligramo

µg Microgramo

MGP Partículas de vidrio magnéticas (*Magnetic Glass Particle*)

mL Mililitro

µL Microlitro

mM Milimolar

µM Micromolar

Mn Manganeseo

Mn(OAc)₂ Acetato de Manganeseo

N

N Número

NA Análogos de nucleó(t)idos (*Nucleo(t)side Analogue*)

NaCl Cloruro de sodio

NaOH Hidróxido de sodio

NAP Polímeros de ácido nucleico (*Nucleic Acid Polymer*)

ng Nanogramos

NGS Secuenciación masiva (*Next generation sequencing*)

NLS Secuencias de localización nuclear (*Nuclear localization sequence*)

nM Nanomolar

nm Nanometro

NTCP Receptor de membrana polipéptido cotransportador de taurocolato de sodio (*Sodium taurocholate cotransporting polypeptide*)

N-terminal Amino terminal

O

OH Grupo hidroxilo

ORF Marco de lectura abierto (*Open Reading Frame*)

P

pb Pares de bases

PCR Reacción en cadena de la polimerasa (Polymerase chain reaction)

pg Picogrammo

PEG Polietilenglicol

PEG-IFN- α Interferon alfa pegilado

PEG-IFN- λ Interferon lambda pegilado

pM Picomolar

pol Polimerasa

PPi Pirofosfato inorgánico

PTP Placa Picotituladora (PicoTiterPlate)

Q

qPCR PCR cuantitativa

qRT-PCR RT-PCR cuantitativa

R

RBD Dominio de unión (*Receptor-Binding Domain*)

RCA Replicación en Círculo Rodante

RE Retículo endoplasmático (*Endoplasmic reticulum*)

RNasa	Ribonucleasa
RNP	Ribonucleoproteína
RT	Transcripción inversa (Reverse transcription)
RTV	Ritonavir
rv	Reverse

S

SBS	Secuenciación por síntesis (Sequencing by synthesis)
S-HBsAg	Proteína del antígeno de superficie de la hepatitis B pequeña
S-HDAg	Antígeno corto del virus de la hepatitis delta
SRE	Factor de respuesta sérica (<i>Serum response Element</i>)
STAT	<i>Signal Transducer and Activator of Transcription</i>
SVP	Partícula subviral (<i>Subviral Particles</i>)

T

TAE	Tris-Acetato-EDTA
TBE	Tris-Borato-EDTA
TNF	Tenofovir
Tris	Tris(hidroximetil)aminometano

U

UI	Unidades internacionales
UPGMA	Método de grupo de pares no ponderados con media aritmética (Unweighted Pair Group Method with Arithmetic mean)

V

VAS	Secuencia de ensamblaje viral (<i>Viral Assembly Sequence</i>)
VHB	Virus de la Hepatitis B

VHC	Virus de la Hepatitis C
VHD	Virus de la Hepatitis Delta
VIH	Virus de la Inmunodeficiencia Humana

Aminoácidos

Aminoácido	Código de tres letras	Código de una letra	Aminoácido	Código de tres letras	Código de una letra
Alanina	Ala	A	Asparagina	Asn	N
Valina	Val	V	Glutamina	Gln	Q
Leucina	Leu	L	Serina	Ser	S
Isoleucina	Ile	I	Treonina	Thr	T
Prolina	Pro	P	Tirosina	Tyr	Y
Fenilalanina	Phe	F	Lisina	Lys	K
Triptófano	Trp	W	Arginina	Arg	R
Metionina	Met	M	Histidina	His	H
Cisteína	Cys	C	Aspartato	Asp	D
Glicina	Gly	G	Glutamato	Glu	E

Bases nitrogenadas presentes en el ADN y ARN

Base nitrogenada	Código
Adenina	A
Citosina	C
Inosina	I
Guanina	G
Timina	T
Uracilo	U

ÍNDICE

RESUMEN.....	17
ABSTRACT	19
1. INTRODUCCIÓN.....	23
1.1 Virus de la hepatitis delta	23
1.1.1 Descubrimiento del VHD.....	23
1.1.2 Transmisión	23
1.1.3 Prevalencia del VHD	24
1.1.4 Genotipos del VHD	24
1.2 Composición del VHD	25
1.2.1 Envuelta viral	26
1.2.2 Complejo ribonucleoproteico.....	27
1.3 Infección y replicación viral.....	27
1.3.1 Entrada del VHD al hepatocito.....	27
1.3.2 Transporte y entrada en el núcleo.....	28
1.3.3 Replicación en círculo rodante.....	28
1.3.4 Formación de nuevas partículas virales	31
1.4 El antígeno delta	32
1.4.1 Dominios funcionales del HDAg.....	34
1.4.2 Modificaciones postraduccionales del HDAg.....	35
1.5 La ribozima.....	36
1.6 Variabilidad genética del VHD	37
1.7 Patogenia del VHD.....	40
1.7.1 Coinfección y superinfección por VHD-VHB.....	40
1.7.2 Hepatitis delta aguda y crónica	41
1.8 Efecto inhibitorio del VHD sobre el VHB.....	42
1.9 Tratamiento antiviral.....	43

1.9.1 Tratamiento actual	44
1.9.2 Nuevos tratamientos	45
2. HIPÓTESIS.....	53
3. OBJETIVOS.....	57
4. MATERIAL Y MÉTODOS.....	61
4.1 Cuantificación del ADN del VHB.....	61
4.2 Determinación de los marcadores serológicos (HBsAg, HBeAg, anti-VHD).....	61
4.3 Extracción del ARN del VHD	62
4.4 Cuantificación del ARN del VHD.....	63
4.5 Amplificación del ARN del VHD para obtener librerías de amplicones para analizar por secuenciación masiva (NGS)	65
4.6 Comprobación de la amplificación del ARN del VHD.....	70
4.7 Purificación del producto de PCR.....	70
4.8 Comprobación de la calidad del producto de PCR purificado	71
4.9 Cuantificación del producto de PCR purificado.....	71
4.10 NGS del producto de PCR purificado	72
4.10.1 NGS mediante la plataforma Roche 454 GS Junior:	73
4.10.2 NGS mediante la plataforma MiSeq:	78
4.11 Análisis de los datos de NGS	86
4.11.1 Análisis de los datos obtenidos mediante el sistema Roche 454 GS Junior ..	86
4.11.2 Análisis de los datos obtenidos mediante el sistema MiSeq.....	87
4.12 Genotipaje.....	89
4.13 Clonación molecular.....	91
4.14 Purificación del producto de clonación	94
4.15 Reacción de secuenciación Sanger.....	94
4.16 Análisis del porcentaje de genomas no editados y editados.....	97
4.17 Cuantificación por NGS de las mezclas de clones con diferentes porcentajes de genomas no editados y editados	98
4.18 Análisis del tipo de cambios nucleotídicos.....	99

4.19	Análisis de la conservación y variabilidad de la región del HDAg y de la región de la ribozima.....	100
4.20	Estudio de los patrones de cambios de aminoácidos de la región del HDAg	102
5.	RESULTADOS	107
5.1	Primer estudio: Caracterización cuantitativa de la edición del genoma del VHD por secuenciación masiva (NGS)	107
5.1.1	Comparación entre las proporciones de genomas no editados y editados obtenidas por NGS y por secuenciación Sanger de los productos de la clonación molecular	108
5.1.2	Cuantificación por NGS de las mezclas de clones con diferentes proporciones de genomas no editados y editados.....	110
5.1.3	Tipos de cambios nucleotídicos en la región del HDAg analizada.....	114
5.2	Segundo estudio: Estudio longitudinal en pacientes con hepatitis crónica delta sobre la variabilidad y conservación de las regiones del HDAg y de la ribozima	115
5.2.1	Análisis de la conservación y variabilidad de la región del HDAg estudiada..	116
5.2.2	Estudio de los patrones de mutaciones de la región del HDAg analizada	120
5.2.3	Análisis de la conservación y variabilidad de la región de la ribozima estudiada	126
6.	DISCUSIÓN.....	133
6.1	Métodos de estudio de la quasiespecies	134
6.1.1	Cuantificación de la edición del codón stop del HDAg en el genoma del VHD mediante NGS	135
6.2	Análisis del patrón de cambios de nucleótidos	136
6.3	Conservación y variabilidad en la quasiespecies viral en la región codificante del HDAg y la ribozima.....	137
6.3.1	Análisis de la conservación y variabilidad en los aminoácidos del HDAg y en su región codificante.....	138
6.3.2	Análisis de la conservación y variabilidad en la región de la ribozima	144
6.4	Lecciones aprendidas sobre la quasiespecies del VHD.....	146
7.	CONCLUSIONES	151
8.	LÍNEAS DE FUTURO	155

8.1 Líneas clínicas.....	155
8.2 Líneas básicas	156
BIBLIOGRAFÍA.....	161
ANEXO.....	181

RESUMEN

El virus de la hepatitis delta (VHD) es un virus de ARN que requiere del antígeno de superficie del virus de la hepatitis B (HBsAg) para completar su ciclo replicativo. Este virus es el agente etiológico de la hepatitis delta, la forma más grave de hepatitis viral en humanos, que se estima que afecta un total de 15 a 20 millones de personas en todo el mundo. El genoma del VHD es una molécula de ARN de una sola hebra, circular, de alrededor de 1700 nucleótidos y de polaridad negativa. Este genoma codifica tan solo una proteína, el antígeno delta (HDAg), que tiene dos isoformas (la larga y la corta) generadas gracias a una modificación postranscripcional (desaminación) durante la replicación viral, que cambia el codón 196 de stop (TAG) al aminoácido W (TGG). Este proceso también se conoce como edición y está catalizado por la adenosina desaminasa que actúa sobre ARN 1 (ADAR-1). Además el genoma contiene un dominio con actividad enzimática capaz de procesar al propio genoma (autoescisión), esencial para la replicación viral, la ribozima. El objetivo principal de esta tesis ha sido estudiar la quasispecies del VHD mediante secuenciación masiva (NGS), analizando la conservación o variabilidad de las regiones funcionales del virus, correspondientes al HDAg (cuantificando su edición) y a la ribozima.

En el primer estudio se incluyó una muestra de plasma de 4 pacientes con hepatitis crónica delta, para validar la cuantificación de porcentajes de poblaciones de genomas editados y no editados por NGS, comparándola con el método de secuenciación Sanger de los productos de clonación. Además, se estudió la precisión y sensibilidad de la NGS, utilizando mezclas de diferentes proporciones de productos de clonación editados y no editados, y se observaron los patrones de cambios nucleotídicos para analizar si la edición es un fenómeno general que tiene lugar en toda la región codificante del HDAg analizada. Por otro lado, en el segundo estudio se incluyeron 36 muestras de plasma de 12 pacientes con hepatitis crónica delta y se analizó el grado de conservación o variabilidad de la quasispecies en las regiones analizadas del HDAg y de la ribozima. Además se determinó la existencia de patrones de mutaciones a nivel aminoacídico en el HDAg.

Mediante los resultados obtenidos se ha demostrado que la NGS es un método útil para detectar proporciones de genomas del VHD editados y no editados, más sensible y preciso que la secuenciación Sanger de los productos de clonación. También se ha observado que el fenómeno de edición en la región codificante del HDAg podría no ser

exclusivo de la posición de edición. Además, los resultados mostraron que la región del HDAg es variable y que los patrones de mutaciones aminoacídicas generalmente podrían estar relacionados con el escape al sistema inmunitario, posiblemente sin afectar significativamente la funcionalidad de las regiones del HDAg donde se encuentran. Por último, se ha observado que la región de la ribozima contiene un elevado nivel de conservación en su secuencia nucleotídica.

ABSTRACT

Hepatitis delta virus (HDV) is an RNA virus that requires the hepatitis B virus surface antigen (HBsAg) to complete its replicative cycle. This virus is the etiological agent of delta hepatitis, the most severe form of viral hepatitis in humans, which is estimated to affect a total of 15 to 20 million people worldwide. The HDV genome is around 1700 nucleotides, single-stranded, circular and negative polarity RNA molecule. This genome encodes just one protein, the delta antigen (HDAg), which has two isoforms (the large one and the short one) generated by post-transcriptional modification (deamination) during viral replication. This modification changes the stop codon 196 (TAG) into the W amino acid (TGG). This process is also known as editing and is catalyzed by adenosine deaminase that acts on RNA 1 (ADAR-1). Furthermore, the genome contains a domain with enzymatic activity capable of processing the genome itself (self-cleaving), essential for viral replication: the ribozyme. The main objective of this thesis has been to study the HDV quasispecies by next generation sequencing (NGS), analyzing the conservation or variability of the functional regions of the virus, corresponding to HDAg (quantifying its edition) and the ribozyme.

In the first study, a plasma sample from 4 patients with chronic hepatitis delta was included to validate the quantification of percentages of edited and unedited genome populations by NGS and it was compared with the Sanger sequencing method of cloning products. In addition, the accuracy and sensitivity of NGS were studied, using mixtures of different percentages of edited and unedited cloning products, and nucleotide change patterns were also analyzed to confirm whether editing is a general phenomenon taking place along the entire coding region of the analyzed HDAg. On the other hand, 36 plasma samples from 12 patients with chronic hepatitis delta were included in the second study, and the conservation or variability of quasispecies in the analyzed regions of HDAg and ribozyme was analyzed. Furthermore, the existence of amino acid mutation patterns in HDAg was determined.

The obtained results showed that NGS is a useful method to detect proportions of edited and unedited HDV genomes, more sensitive and accurate than Sanger sequencing of cloning products. It has also been observed that the editing phenomenon taking place along the entire HDAg coding region analyzed may not be unique to the editing position. Furthermore, the results showed that the HDAg region is variable and that amino acid mutation patterns could generally be related to escape from immune system, possibly

without significantly affecting the functionality of the HDAg regions where they are found. Finally, it has been observed that the ribozyme region contains a high level of conservation in its nucleotide sequence.

INTRODUCCIÓN

1. INTRODUCCIÓN

1.1 Virus de la hepatitis delta

El virus de la hepatitis delta (VHD) es un virus de ARN defectivo, ya que requiere del antígeno de superficie del virus de la hepatitis B (HBsAg) para completar su ciclo replicativo. Perteneció a la familia *Deltaviridae* y al género *Deltavirus*. Este virus es el agente etiológico de la hepatitis delta, la forma más grave de hepatitis viral en humanos.^{1,2}

1.1.1 Descubrimiento del VHD

El VHD fue descrito por primera vez en 1977 por un grupo italiano liderado por el Dr. Rizzetto. Se descubrió un nuevo sistema antígeno-anticuerpo al que llamaron antígeno delta y anticuerpo anti-delta. Estaba presente en el hígado y en la sangre de pacientes con enfermedad hepática crónica por el virus de la hepatitis B (VHB) con HBsAg positivo, que padecían una hepatitis muy severa. En un principio se hipotetizó que dicha hepatitis era causada por una variante del VHB que expresaba un nuevo antígeno.³ Posteriormente, en 1980, estudios de experimentación animal con chimpancés demostraron que ese antígeno era en realidad un nuevo virus de hepatitis dependiente del VHB y fue nombrado virus de la hepatitis delta o VHD. Este virus era el agente infeccioso responsable de la exacerbación de la enfermedad hepática en dichos pacientes.⁴ Años más tarde, varios laboratorios clonaron y secuenciaron el genoma del VHD, y determinaron que estaba formado por un ARN circular de cadena sencilla de alrededor de 1700 nucleótidos.⁵⁻⁸

1.1.2 Transmisión

El VHD se transmite por vía parenteral a través de la exposición a sangre infectada o fluidos corporales infectados, igual que el virus de la inmunodeficiencia humana (VIH) o el virus de la hepatitis C (VHC), con los que suele coinfectar. El riesgo de infección es alto en usuarios de drogas inyectables, pacientes con múltiples parejas sexuales, hombres que tienen sexo con hombres y personas de áreas endémicas. En poblaciones altamente endémicas, la transmisión se produce principalmente a través de la propagación intrafamiliar e iatrogénica en asociación con malas condiciones de higiene. En cambio, la transmisión perinatal es poco común. Debido al cribado en los productos sanguíneos, las

nuevas infecciones en hemofílicos, receptores de transfusiones de sangre y pacientes que reciben hemodiálisis, ya son residuales en los países desarrollados.⁹⁻¹¹

1.1.3 Prevalencia del VHD

Al menos el 5% de las personas con infección crónica por el VHB están coinfectadas con el VHD, lo que da como resultado un total de 15 a 20 millones de personas infectadas en todo el mundo,¹² aunque hay estudios que estiman que la prevalencia global podría ser superior.¹³⁻¹⁵ Las áreas de alta prevalencia incluyen África (África central y occidental), Asia (Asia central), Islas del Pacífico (Kiribati y Nauru), Oriente Medio, Europa Oriental (regiones del Mediterráneo Oriental y Turquía) y América del Sur (cuenca amazónica) (Figura 1).¹²

A nivel mundial, el número total de infecciones ha disminuido desde 1980. Esta tendencia se debe principalmente a la disminución en el número de portadores del HBsAg, debido a mejores estándares de salud pública, programas de vacunación contra el VHB, utilización de jeringas desechables y medidas para controlar la propagación del VIH.¹² Sin embargo, sigue habiendo regiones endémicas como las que se han citado anteriormente y en algunas áreas no se ha observado disminución, probablemente debido a la migración desde éstas regiones (por ejemplo, en Europa Central y Norte América).¹⁶⁻¹⁷ Como se puede observar en la Figura 1, la estimación global y la información geográfica están incompletas porque muchos países no informan de la prevalencia de este virus. Además, otro factor importante para subestimar la prevalencia de la infección por VHD es la falta de pruebas rutinarias de anticuerpos específicos del HDAg en individuos con el HBsAg positivo.¹²

1.1.4 Genotipos del VHD

Hasta la fecha, se han descrito 8 genotipos en base a la divergencia genética de distintos genomas virales (como se explicará en el apartado *1.6 Variabilidad genética del VHD*) que dan lugar a grupos filogenéticamente diferentes, numerados del 1 al 8. Estos genotipos están distribuidos por todo el mundo, presentando una clara prevalencia en regiones geográficas muy específicas, a excepción del genotipo 1 que es frecuente en todo el mundo. El genotipo 2 se encuentra en Japón, China, Taiwán, sudeste asiático y Rusia, el genotipo 3 en la región del Amazonas, el genotipo 4 en Taiwán y Japón, los genotipos 5, 6 y 7 en África y el genotipo 8 en África y en Europa, en personas de origen africano (Figura 1).¹⁸⁻²⁰ Los diferentes genotipos pueden dar lugar a diferentes manifestaciones clínicas. En comparación con el genotipo 1, los genotipos 2 y 4 parecen

causar una enfermedad hepática más leve. El genotipo 2 en particular, se ha asociado con una menor incidencia de cirrosis, hepatocarcinoma y una disminución de la mortalidad en comparación con el genotipo 1. El genotipo 3, por otro lado, se asocia con un curso más grave de infecciones agudas y un mayor riesgo de insuficiencia hepática aguda.^{21,22}

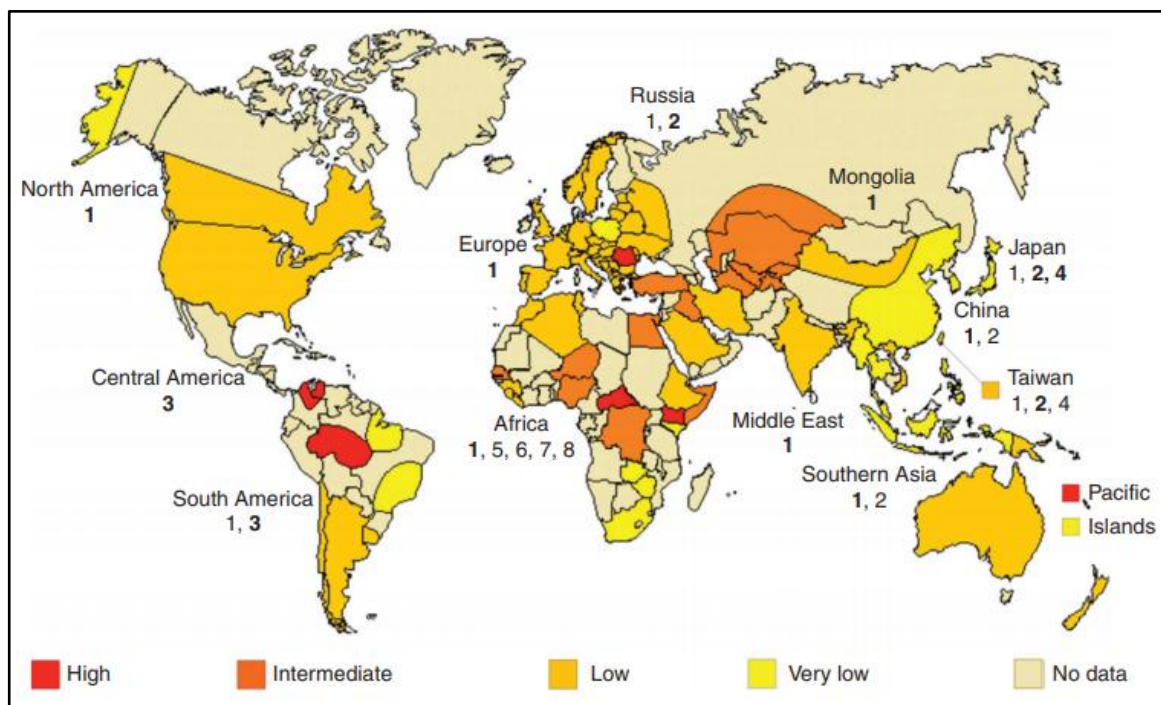


Figura 1. Prevalencia y distribución geográfica de los diferentes genotipos del VHD alrededor del mundo.²³

1.2 Composición del VHD

El virión delta es una pequeña partícula esférica de aproximadamente 36 nm de diámetro, el más pequeño entre los virus de mamíferos.⁴ Se compone de una envuelta viral y un complejo ribonucleoproteico (RNP) (Figura 2).

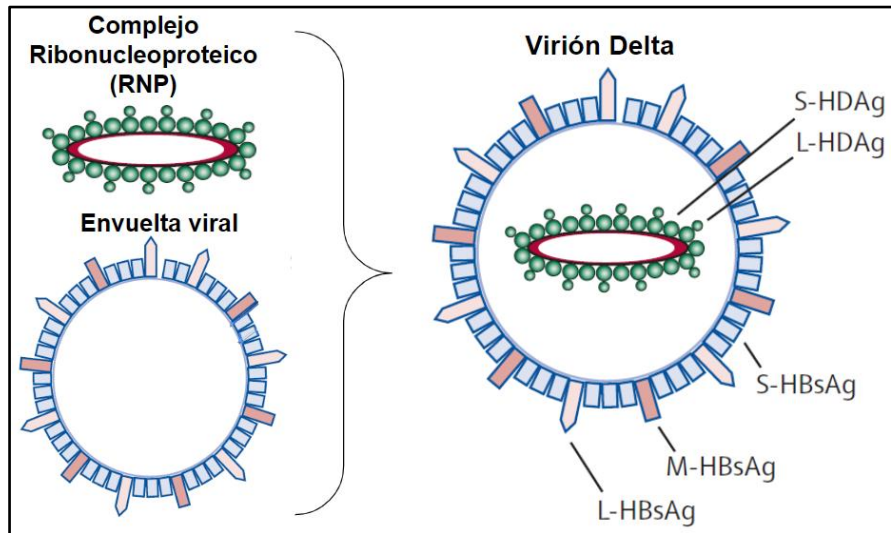


Figura 2. Representación esquemática del virión delta. Figura modificada.²⁴

1.2.1 Envuelta viral

La envuelta viral contiene una membrana lipídica externa, proveniente del retículo endoplasmático (RE) de la célula huésped, en la que están ancladas las tres proteínas del HBsAg (pequeña (S-HBsAg), mediana (M-HBsAg) y grande (L-HBsAg)).⁴ Un solo marco de lectura abierto en el genoma del VHB codifica estas tres proteínas, pero la traducción se inicia a partir de tres codones AUG distintos. Por este motivo, estas tres proteínas son idénticas en su región carboxilo terminal (C-terminal), de 226 aminoácidos, que corresponde a la secuencia de la S-HBsAg. Respecto a esta secuencia común, la M-HBsAg presenta 55 aminoácidos en el extremo amino terminal (N-terminal) que corresponden al dominio PreS-2 y la L-HBsAg presenta, además de los 55 aminoácidos del dominio PreS-2, 108 o 119 aminoácidos adicionales (según el genotipo del VHB) en el extremo N-terminal que corresponden al dominio PreS-1. La S-HBsAg es más abundante que las otras dos proteínas y es necesaria tanto para la entrada del VHD en el hepatocito como para la liberación de las partículas.^{25,26,30} La región PreS-1 presente en la L-HBsAg es el sitio de unión de alta afinidad al receptor de membrana polipéptido cotransportador de taurocolato de sodio (NTCP) presente en los hepatocitos. El extremo N-terminal de la región PreS-1 está miristoilado y esta modificación juega un papel esencial en la entrada del VHD al hepatocito y en su infectividad.^{27,28} En cambio, la M-HBsAg no parece ser necesaria para la infectividad.^{29,30} Curiosamente, el VHD puede ensamblarse solo con la S-HBsAg, pero sin la L-HBsAg, la partícula no es infecciosa.³¹

1.2.2 Complejo ribonucleoproteico

El complejo RNP está constituido por una copia del genoma del VHD y alrededor de 200 moléculas de las dos diferentes isoformas del HDAg, la larga (L-HDAg) y la corta (S-HDAg). El genoma del VHD es una molécula de ARN de una sola hebra, circular, de entre 1.672-1.697 nucleótidos (dependiendo del genotipo) y de polaridad negativa. Este genoma presenta un alto contenido en G/C (Guanosina/Citosina) y su secuencia es autocomplementaria en un 70%. Por este motivo, el genoma puede plegarse sobre sí mismo por apareamiento de bases intramolecular, resultando en una estructura secundaria no ramificada de gran estabilidad, prácticamente constituyendo un genoma de ARN de doble cadena cerrado por los extremos.^{5,32} Estas características del genoma del VHD hacen que se parezca más a un viroide, un patógeno de plantas, que a cualquier virus que infecta a animales. No obstante, su genoma es más largo que el de cualquier viroide (que habitualmente tienen una longitud entre 200 y 450 nucleótidos), además codifica el L-HDAg y S-HDAg, mientras que los viroides no codifican ninguna proteína. Finalmente los viroides no necesitan la colaboración de ningún virus para completar su ciclo replicativo.³³

1.3 Infección y replicación viral

1.3.1 Entrada del VHD al hepatocito

La entrada del VHD empieza con la unión de la partícula viral a la superficie del hepatocito. Los proteoglicanos heparán sulfatos (HSPGs) presentes en la superficie del hepatocito se unen a la asa antigénica presente en el dominio S.^{30,34,35} En particular, el glicano-5, un grupo de seis HSPGs altamente expresados en el hígado, ha sido identificado recientemente como el principal factor de entrada del VHD, siendo responsable de la unión inicial del virus a la superficie del hepatocito.³⁶ A esta unión le sigue la interacción de la región N-terminal miristoilada del dominio PreS-1 de la proteína L-HBsAg con el receptor de membrana NTCP, el cual se distribuye en la membrana basolateral de los hepatocitos y participa en la circulación enterohepática de sales biliares. Los pasos posteriores de entrada del virus están menos caracterizados, se cree que la interacción con el receptor de membrana NTCP permite la fusión de la membrana viral con la membrana celular del hepatocito o se cree que se somete a un proceso de fusión mediado por la endocitosis para liberar el complejo RNP al citoplasma celular.³⁷⁻³⁹

1.3.2 Transporte y entrada en el núcleo

Una vez en el citoplasma del hepatocito, las secuencias de localización nuclear (NLS) presentes en los HDAG, permiten que el complejo RNP sea dirigido al núcleo y las importinas celulares, que reconocen secuencias de aminoácidos específicas para promover el transporte a través de los poros nucleares, hacen que la RNP entre al interior del núcleo del hepatocito para iniciar el proceso de replicación viral.^{39,40}

1.3.3 Replicación en círculo rodante

La replicación del genoma del VHD ocurre en el núcleo del hepatocito y es completamente independiente del VHB. Como ninguna de las dos formas del HDAG, que son las únicas proteínas que codifica el VHD, tiene actividad ARN polimerasa, el virus secuestra a las ARN polimerasas del huésped [ARN polimerasa I (ARN pol I), ARN polimerasa II (ARN pol II) y ARN polimerasa III (ARN pol III)].⁴¹ Este proceso se basa en tres partes: Replicación en círculo rodante, auto-escisión y ligación (Figura 3).

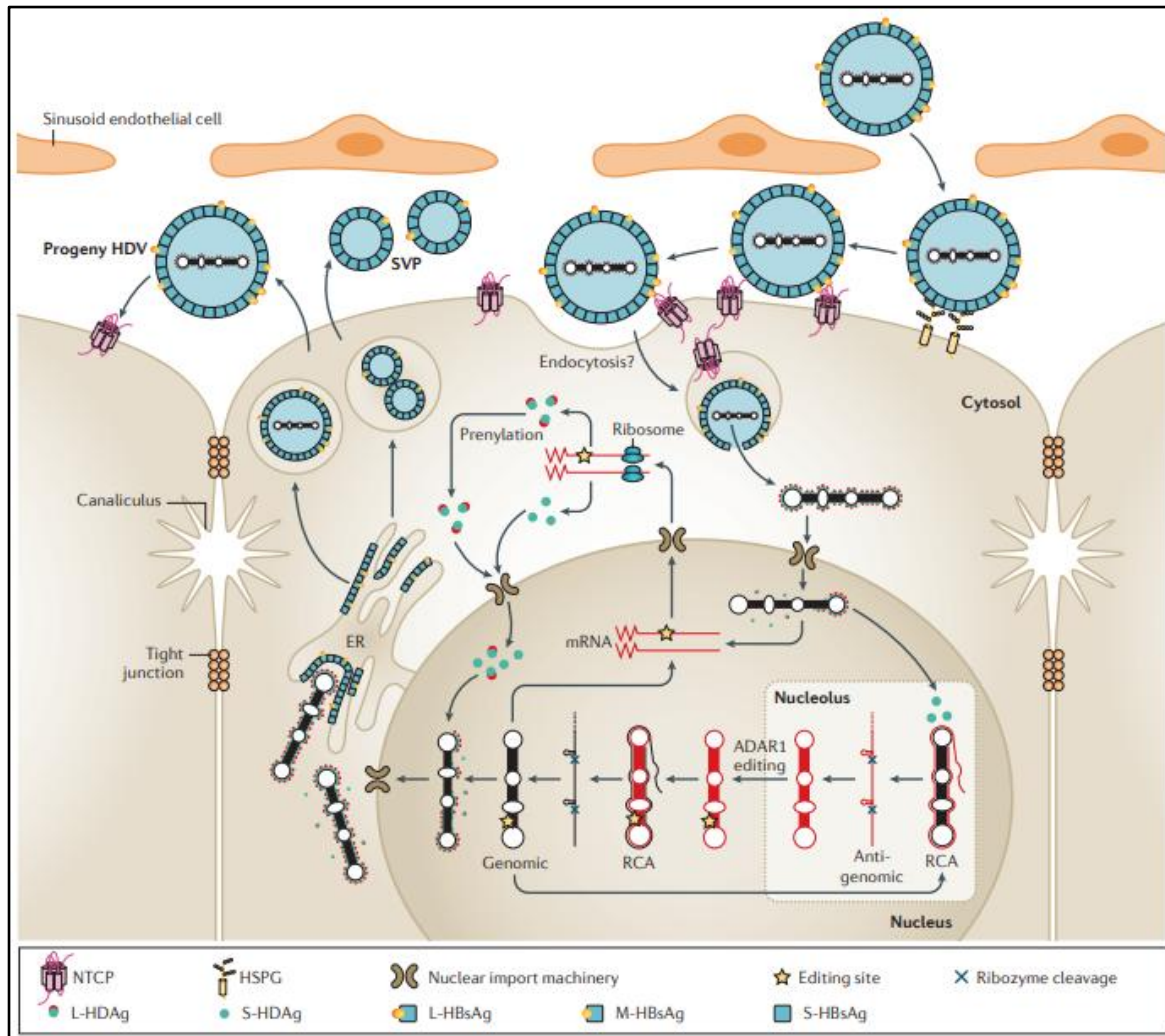


Figura 3. Ciclo replicativo del VHD. El VHD interactúa con los HSPGs, se une al NTCP y se fusiona con la membrana plasmática o se somete a un proceso de fusión mediado por endocitosis para liberar el complejo RNP al citosol. El complejo RNP se transporta al núcleo para iniciar la replicación del ARN. El ARN genómico con polaridad negativa sirve como plantilla para el primer paso de la replicación en círculo rodante (RCA). El ARN antigenómico resultante, se escinde por la actividad de la ribozima, se liga originando el ARN antigenómico circular y este se somete a una segunda RCA. El ARN genómico resultante, se escinde por la actividad de la ribozima, se liga originando el ARN genómico circular y este se usa como plantilla para la transcripción del (ARN mensajero) ARNm de manera dependiente a la ARN pol II. Los ARNm se traducen en S-HDAg y L-HDAg y se transportan al núcleo para regular la replicación del virus o unirse al genoma circular del virus para formar RNP, que se exportan al citosol. L-HDAg es responsable de las interacciones de la RNP con las proteínas del HBsAg. El VHD se libera a través de la vía secretora clásica y los virus de la progenie infectan a los hepatocitos vecinos. SVP, partícula subviral.⁴²

1.3.3.1 Replicación en círculo rodante

La replicación del genoma del VHD tiene lugar mediante un mecanismo denominado replicación en círculo rodante (RCA), de manera similar a los viroides. Este mecanismo consiste en un proceso de replicación unidireccional del ARN genómico que permite la síntesis de transcripciones lineales multiméricas complementarias al genoma, que corresponden al ARN antigenómico. A partir del ARN antigenoma, siguiendo el mismo

proceso, se generan nuevas moléculas de ARN genómico y así se van generando sucesivamente nuevos ARN multiméricos de polaridad genómica o antigenómica. Este proceso tiene lugar mediante la ARN pol II, aunque hay estudios que han demostrado que la ARN pol I y III también interactúan con el ARN-VHD.⁴³ Debido a la estructura plegada del genoma, por apareamiento de bases intramolecular, como se ha comentado anteriormente, las ARN polimerasas de la célula tratan el genoma viral de ARN como un ADN de doble hebra e inician su replicación.^{41,43,44} El antigenoma tiene polaridad positiva y por lo tanto contiene la secuencia codificante del HDAg. Sin embargo, su localización es exclusivamente nuclear y le falta el elemento interno del sitio de entrada al ribosoma, por lo que no puede traducirse directamente a S-HDAg y L-HDAg. Por tanto, se produce una tercera forma de ARN viral en el huésped durante la replicación. Además del ARN genómico circular y el ARN antigenómico circular complementario, se produce un ARN antigenómico 5'-capped y 3'-poliadenilado lineal de 800 nucleótidos y de la misma polaridad que el antigenómico (por lo tanto producido a partir del ARN genómico), que es el ARN mensajero (ARNm) a partir del cual se producirá la traducción del HDAg (Figura 4).^{24,45-47} Además, la evidencia sugiere que la síntesis de las diferentes especies de ARN se produce en diferentes ubicaciones subcelulares, mediada por distintas polimerasas celulares: la síntesis del ARN antigenómico se produce en el nucléolo mediada por la ARN pol I, mientras que la síntesis del ARN genómico se produce de manera más difusa en el nucleoplasma por la ARN pol II.⁴⁸

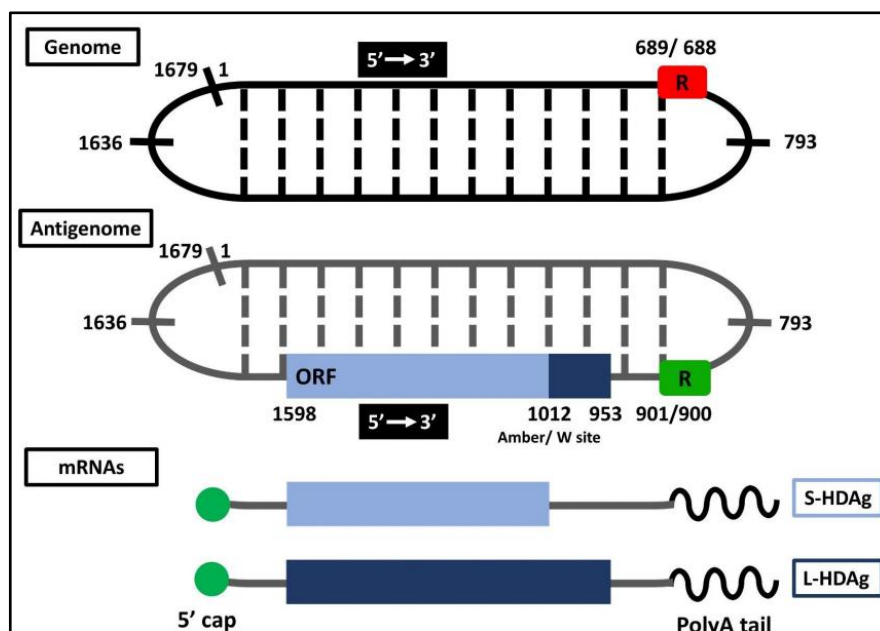


Figura 4. Estructura de las tres formas de ARN-VHD: ARN genómico circular, ARN antigenómico circular complementario y ARNm. Aunque el ARN antigenómico contiene el marco de lectura abierta (ORF) para el HDAg en la posición 1598, el antígeno se traduce del ARNm. R: Ribozima.⁴⁹

1.3.3.2 Auto-escisión

Tras la replicación del genoma viral ocurre el proceso de autoescisión del ARN-VHD lineal, en el cual los ARN genómicos y antigenómicos multiméricos generados darán lugar a unidades monoméricas. Este proceso lo lleva a cabo una secuencia de 85 nucleótidos llamada ribozima, en más detalle en el apartado *1.5 La ribozima*, presente tanto en el ARN genómico (sitio de escisión entre los nucleótidos 689/688) como en el ARN antigenómico (sitio de escisión entre los nucleótidos 901/900) (Figura 4).^{50,51}

1.3.3.3 Ligación

Tras el proceso de autoescisión los monómeros del VHD se ligan para formar estructuras de ARN circular.⁵² El mecanismo molecular involucrado en este proceso sigue sin estar claramente definido. Algunos estudios involucran a una ARN ligasa del huésped,⁵³ mientras que otros estudios creen que la propia ribozima posee propiedades de autoligado.⁵⁴

1.3.4 Formación de nuevas partículas virales

Las diferentes formas del ARN-VHD producidas durante su replicación se encuentran en diferentes proporciones. En primer lugar, el ARN genómico circular es el más abundante, pudiendo detectarse hasta 300.000 copias por célula durante el proceso de infección, entre el citoplasma y el núcleo. El ARN genómico es el único que se encapsida en los viriones. En segundo lugar, el ARN antigenómico complementario circular se encuentra entre 5-22 veces en menor abundancia que el ARN genómico. Y, en tercer lugar, el ARNm que es el menos abundante con aproximadamente 600 copias por célula.⁵⁵⁻⁵⁷

Para comenzar la transcripción del ARNm a partir del ARN circular, se requiere del S-HDAg pero solo en cantidades bajas.⁵⁸

El ARNm se exporta al citoplasma y se traduce en el RE para formar nuevas moléculas de las isoformas del HDAg. El S-HDAg entra en el núcleo donde promueve la replicación del VHD. Por otro lado, el L-HDAg se somete a una prenilación en el citoplasma antes de entrar en el núcleo, donde inhibe la replicación viral. Ambas formas del HDAg se asocian con nuevas transcripciones de ARN genómico para formar nuevas RNP. Las RNP se exportan al citoplasma y se dirigen a las membranas del RE por una señal en el extremo C-terminal del L-HDAg. Una vez allí, el ensamblaje del VHD depende de la interacción específica entre el grupo farnesilo unido a la L-HDAg y la región S del HBsAg, por lo tanto, solo puede tener lugar en células coinfectadas con el VHB. Es probable que el

grupo farnesilo unido a la L-HDAg ancla la RNP a la membrana del RE donde se sintetizan las proteínas del HBsAg⁵⁹ y que la unión de la RNP a las proteínas del HBsAg involucra a una secuencia rica en prolina ubicada cerca del sitio de farnesilación.^{60,61} Además, la interacción del extremo C-terminal del L-HDAg con la cadena pesada de clatrina en la red trans-Golgi es esencial para el ensamblaje viral. Por último, las partículas se liberan del hepatocito usando la vía de secreción de Golgi para infectar nuevas células.⁶²⁻⁶⁴

El VHD puede entrar en los hepatocitos que no expresan el HBsAg, replicar eficientemente su genoma y expresar el HDAg; sin embargo, no se produce secreción de partículas infecciosas.⁶⁵

1.4 El antígeno delta

El HDAg tiene dos isoformas, el S-HDAg y el L-HDAg. El S-HDAg consta de 195 aminoácidos y tiene un peso molecular de 24 kDa, en cambio, el L-HDAg tiene la misma secuencia N-terminal que el S-HDAg pero con 19 aminoácidos más, por lo tanto con un total de 214 aminoácidos y un peso molecular de 27 kDa.⁶⁶ Estas dos formas del HDAg se forman a partir de la región codificante que contiene el ARN antígenómico. Esto ocurre a partir de la única pauta de lectura, gracias a una modificación postranscripcional (desaminación) en el ARN antígenómico durante la replicación viral, que cambia el codón 196 de stop (ámbar) al aminoácido triptófano (W). Este proceso también se conoce como edición y está catalizado por la adenosina desaminasa que actúa sobre ARN 1 (ADAR1), que actúa sobre ARN de doble cadena dentro del núcleo. La adenosina (A) en la posición nucleotídica 1012 se desamina a una inosina (I), que al replicarse el ARN antígenómico es reconocida como una guanosina (G), dando lugar a la adición de una citosina (C) en el ARN genómico. Esto da lugar a una transición en el ARNm y antígenómico resultantes, que transforma el codón de parada original UAG en un codón de W UGG (Figura 5). En este proceso se obtienen genomas no editados y editados. Los genomas no editados contienen el codón de parada original UAG, en las posiciones nucleotídicas 1011-1013, y codifican el S-HDAg. En cambio, los genomas editados contienen el codón de W UGG, en las posiciones nucleotídicas 1011-1013, y codifican el L-HDAg.⁶⁷⁻⁷⁰

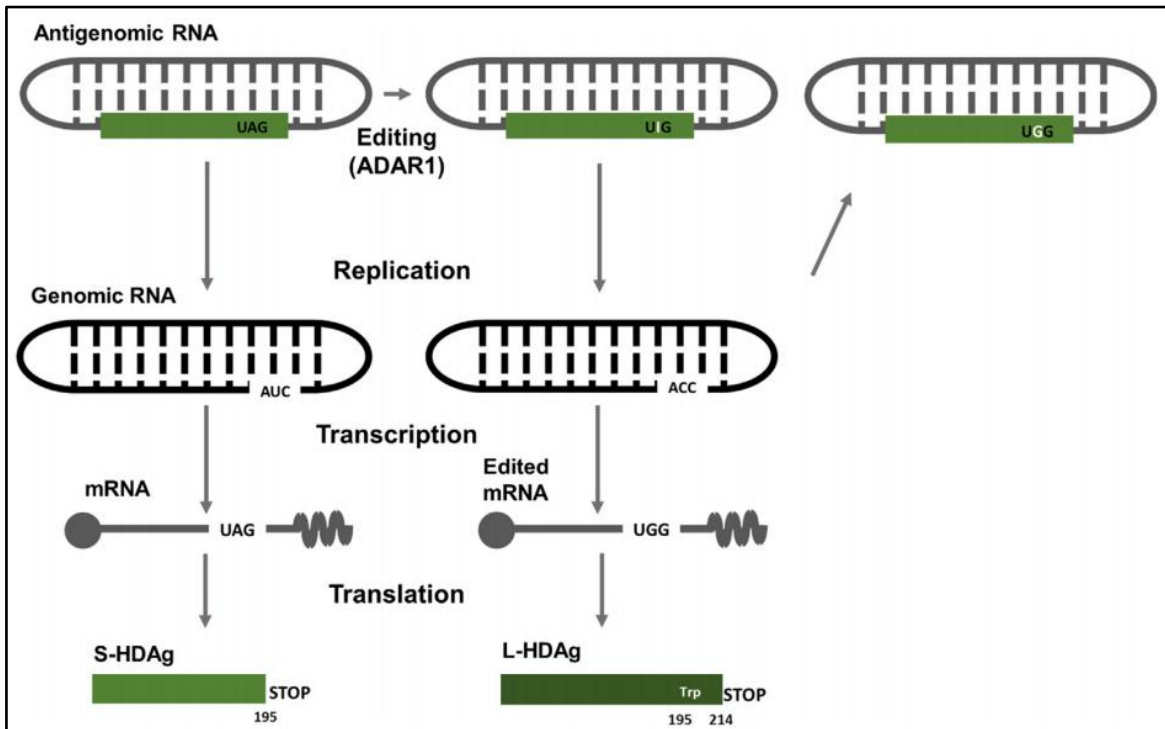


Figura 5. Mecanismo de edición del ARN-VHD. ADAR-1 cataliza la edición del sitio ámbar/W (codón 196) en el ARN antígenómico del VHD, convirtiendo la adenosina 1012 en inosina. Después de la replicación y de la transcripción, el codón de parada original (AUG, que determina la síntesis del S-HDAg) se convierte en UGG en el ARNm y el antígenómico. Este codón codifica un W y permite que la traducción continúe hasta el siguiente codón stop, lo que da como resultado la adición de 19 aminoácidos (L-HDAg).⁶⁵

ADAR-1 tiene dos isoformas: pequeña (ADAR-1-S) y grande (ADAR-1-L), con un extremo C-terminal común. ADAR-1-S es la forma más abundante, expresada constitutivamente, y tiene una localización nuclear, mientras que ADAR-1-L se encuentra principalmente en el citoplasma y su expresión es estimulada por el interferón tipo I. Como la plantilla antígenómica se localiza exclusivamente en el núcleo, se cree que ADAR-1-S media la mayoría de los eventos de edición.⁷⁰ Sin embargo, algunos informes también especulan sobre el papel de ADAR-1-L en la edición.^{71,72}

El S-HDAg y el L-HDAg tienen diferentes funciones en el ciclo viral.⁷³ El S-HDAg se produce en las primeras etapas del ciclo viral y, como se ha dicho anteriormente, promueve la replicación del VHD: es necesario para la iniciación de la replicación del genoma viral, para el reclutamiento de la maquinaria de transcripción, para promover el alargamiento del ARN al eliminar los factores inhibitorios del mismo y para la acumulación del ARN-VHD durante el ciclo.⁷⁴⁻⁷⁶ También se ha demostrado que el S-HDAg interactúa con varias subunidades de la ARN pol I y II.⁷⁷ En cambio, el L-HDAg se sintetiza más tarde durante la infección y sus 19 aminoácidos adicionales le confieren propiedades funcionales únicas. El L-HDAg se encarga del ensamblaje viral, siendo

esencial para la producción de nuevas partículas del VHD.⁷⁸ Pero, tal como se ha dicho anteriormente, el L-HDAg también inhibe la replicación viral, más concretamente, bajas cantidades de L-HDAg inhiben la síntesis del ARN genómico; en cambio, el L-HDAg en exceso puede inhibir la síntesis del ARN antígenómico y del ARNm.⁷⁹

Por lo tanto, la edición del ARN, mediada por ADAR-1, es fundamental para el ciclo de replicación del VHD porque controla los niveles de cada isoforma del HDAg y, en consecuencia, el equilibrio entre la síntesis viral y el ensamblaje de partículas.²⁴

1.4.1 Dominios funcionales del HDAg

El S-HDAg y el L-HDAg comparten varios dominios funcionales dentro de la secuencia de aminoácidos común (Figura 6). Ambas isoformas, entre los aminoácidos 2-27, contienen un dominio crítico de unión al ARN (RBD crítico); entre los aminoácidos 12-60, contienen un dominio de oligodimerización; entre los aminoácidos 31-52 tienen una *coiled-coil sequence* (CCS) y entre los aminoácidos 66-75, contienen una señal de localización nuclear (NLS). Esta NLS es la responsable de transportar los HDAg y el complejo RNP, del citoplasma al núcleo. Además, entre los aminoácidos 97-146 hay un segundo dominio de unión a ARN (RBD), constituido por dos motivos ricos en argininas (ARM), comprendidos entre los aminoácidos 97-107 y 136-146, y una región *helix-loop-helix* (HLH) comprendida entre los motivos ricos en argininas. Y, además, la unión de la ribozima se produce entre los aminoácidos 24-50 y la unión de la ARN pol II entre los aminoácidos 150-195.^{41,80} La unión de la ARN pol II y el S-HDAg estimula la transcripción de ADN y ARN mejorando el paso de elongación *in vitro*, afecta en la decisión de qué nucleótido se incorpora, acelera la translocación hacia adelante de la ARN pol II y puede facilitar la inusual síntesis de ARN dependiente de la ARN pol II.⁸¹

El L-HDAg contiene dominios funcionales únicos, como la secuencia de ensamblaje viral (VAS) comprendida entre los aminoácidos 195-214 que comprende la señal de exportación nuclear (NES) entre los aminoácidos 198-210 y la señal de isoprenilación entre los aminoácidos 211-214. Este dominio es de vital importancia para la interacción del complejo RNP con los HBsAg y para la formación de las partículas virales.^{41,80}

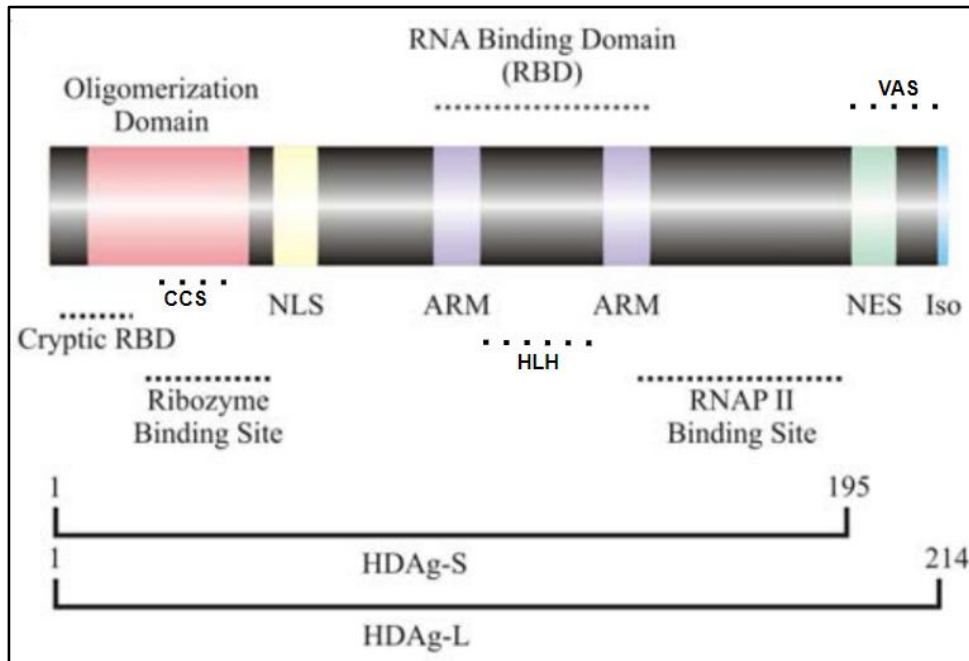


Figura 6. Representación de los dominios funcionales del HDAg. RBD: Dominio de unión al ARN; CCS: *Coiled-coil sequence*; NLS: señal de localización nuclear; ARM: motivos ricos en arginina; HLH: *Helix-loop-helix*; VAS: secuencia de ensamblaje viral; NES: señal de exportación nuclear; Iso: señal de isoprenilación para la adición de farnesilo. Figura modificada.⁸⁰

1.4.2 Modificaciones postraduccionales del HDAg

Ambos antígenos delta sufren modificaciones postraduccionales catalizadas por las enzimas del huésped, tales como fosforilación, acetilación, metilación, sumoilación e isoprenilación (Figura 7).⁸²⁻⁸⁴

La fosforilación en la serina (S) 177 y en la S-123 del S-HDAg aumenta la replicación del ARN antigenómico al interactuar con la ARN pol II.⁸⁵⁻⁸⁷ También está fosforilada la S-2 que es necesaria para la actividad del S-HDAg en la replicación del VHD. Además también se ha demostrado la fosforilación en la S-180, la T-182 y la S-210 pero no se conoce del cierto su funcionalidad.^{88,89} La acetilación regula el ciclo del VHD, en concreto, en la lisina (K) 72, que se encuentra dentro de la secuencia de localización nuclear, y se cree que regula el tráfico del ARN viral entre el núcleo y el citoplasma. Además, la acetilación del S-HDAg se ha sugerido como esencial para la síntesis del ARN genómico y del ARNm viral, pero prescindible para la síntesis del ARN antigenómico.⁹⁰ La metilación de la arginina (R) 13, que se encuentra dentro de un dominio de unión a ARN en el S-HDAg, es esencial para la translocación del S-HDAg al núcleo y para la replicación del ARN antigenómico que formará ARN genómico.⁹¹ La sumoilación se ha observado en múltiples residuos de K, pero sólo ha sido detectada en la isoforma S-

HDAg y mejora la síntesis del ARN genómico y del ARNm.⁹² La isoprenilación en la cisteína (C) 211 del L-HDAg es necesaria para que el virus interactúe con el HBsAg. A este sitio de prenilación, la farnesil transferasa celular le une covalentemente un farnesilo. La farnesilación del L-HDAg es importante para inhibir la replicación viral y al mismo tiempo, promover la encapsidación del complejo RNP y la liberación de las partículas virales.⁹³⁻⁹⁵

Por lo tanto, las modificaciones postraduccionales regulan el ciclo viral y podrían ser atractivos objetivos terapéuticos para el desarrollo de fármacos.

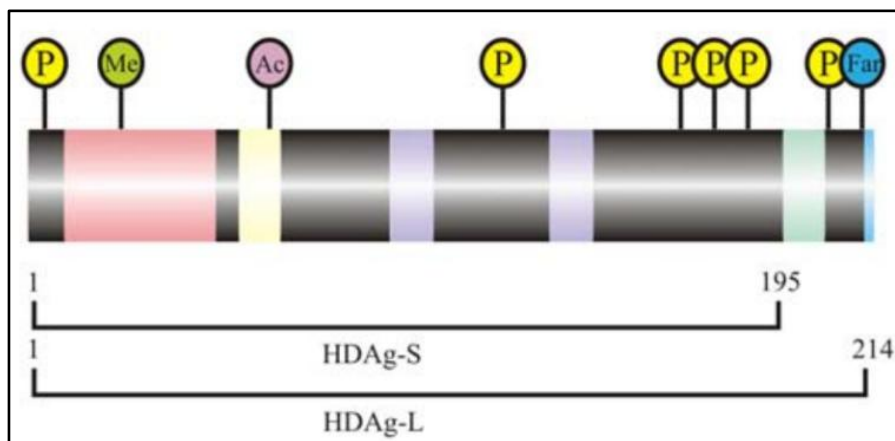


Figura 7. Modificaciones postraduccionales del HDAg. P: Fosforilación; Me: Metilación; Ac: Acetilación; Far: Farnesilación.⁸⁰

1.5 La ribozima

La ribozima es un dominio del genoma del VHD de 85 nucleótidos con actividad enzimática capaz de procesar al propio genoma (autoescisión) tanto *in vitro* como *in vivo* en ausencia de cualquier factor proteico.⁹⁶ La estructura cristalina de la ribozima del VHD se obtuvo en 1998. Su secuencia está presente en el ARN genómico (nucleótidos 688-771) y antigenómico (nucleótidos 900-984) pero no es exactamente igual aunque adopta la misma estructura y está muy conservada en todos los genotipos del VHD (Figura 8). En el ARN genómico el sitio de escisión se encuentra entre los nucleótidos 689/688 y en el ARN antigenómico el sitio de escisión se encuentra entre los nucleótidos 901/900.⁴⁹ La estructura secundaria de la ribozima consta de regiones de doble cadena (P1, P1.1, P2, P3 y P4), regiones monocatenarias (J1/2, J1.1/4 y J4/2) y regiones en forma de bucle (L3 y L4).⁹⁷

La ribozima del VHD promueve la autoescisión a través de una reacción de transesterificación, promoviendo un ataque nucleofílico por un grupo hidroxilo 2' del nucleótido anterior al primero de la ribozima en el fosfato adyacente. La reacción produce un fosfato cíclico 2'-3' en el nucleótido anterior al primero de la ribozima y un grupo hidroxilo libre en 5' en el primer nucleótido de la ribozima.⁵⁰ Ambas ribozimas requieren de iones metálicos divalentes, como magnesio (Mg^{2+}), manganeso (Mn^{2+}) o calcio (Ca^{2+}), para una catálisis eficiente y exhiben un actividad drásticamente menor con iones monovalentes.^{51,98}

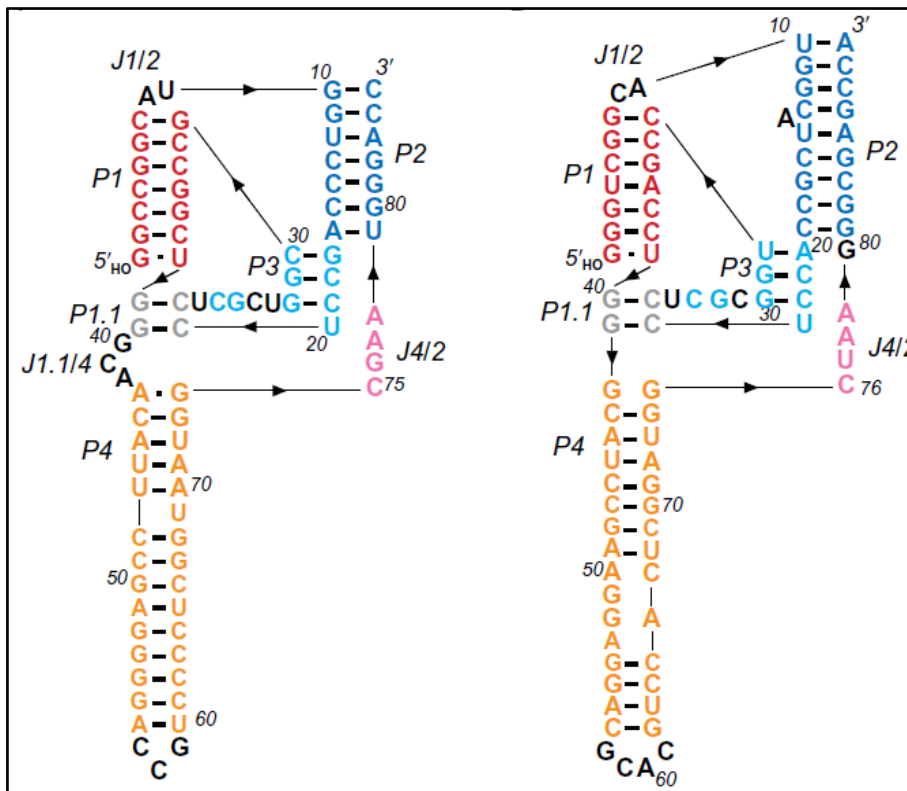


Figura 8: Estructuras secundarias de la ribozima del VHD del ARN genómico (izquierda) y del ARN antígenómico (derecha). Regiones de doble cadena: en rojo P1, en azul oscuro P2, en azul claro P3, en gris P1.1 y en naranja P4. Regiones monocatenaria: en rosa J4/2, sin colorear J1/2 y J1.1/4. Las flechas indican la dirección de las hebras y el 5' OH marca el extremo 5' de la ribozima después de la autoescisión.⁹⁶

1.6 Variabilidad genética del VHD

Tal como se ha explicado en la sección 1.3.3.1 *Replicación en círculo rodante* la principal enzima involucrada en la replicación del VHD es la ARN pol II. Dicha polimerasa es una enzima con actividad de corrección de errores (*proof reading*), con una baja tasa de error de transcripción ($\sim 1,00E-06$ sustituciones/nucleótido).⁹⁹ Sin embargo, el cambio de molde de ADN a ARN puede estar asociado a pequeñas deleciones o inserciones de nucleótidos e incluso recombinación intermolecular, que pueden desempeñar un papel

importante en el aumento de la diversidad genética.^{100,101} Además, la unión del S-HDAg a la ARN pol II, acelera la translocación directa de esta polimerasa a coste de su fidelidad, tanto en la adición de nucleótidos entrantes como el reconocimiento del molde.⁸¹ Además, la posibilidad de que las proteínas de edición genética del huésped también puedan influir en la variabilidad genética del VHD no puede descartarse. Un ejemplo de estas proteínas serían ciertos miembros de la familia de las enzimas catalíticas similares a polipéptidos de edición de ARNm apolipoproteína B (APOBEC), cuya síntesis resulta significativamente inducida en la coinfección VHB-VHD respecto a la mono infección con VHB en modelos *in vivo*.¹⁰²

La elevada diversidad genética resultante de todas estas fuentes de variabilidad ha permitido clasificar el VHD en 8 genotipos, como se ha comentado en el apartado 1.1.4 *Genotipos del VHD*, debido a su alta diversidad genética. La secuencia del genoma del ARN-VHD es muy variable, existe una divergencia de hasta un 16% dentro de un mismo genotipo y de un 20-40% entre diferentes genotipos.¹⁰³ Otra consecuencia de esta elevada diversidad genética es la alta tasa de evolución observada para el VHD por secuenciación masiva, de 9,5E-03 a 1,2E-03 sustituciones/sitio/año.¹⁰⁴ La replicación activa y la alta tasa evolutiva de los genomas virales son las dos características importantes del ciclo de vida de un virus ARN, por las cuales se genera una gran heterogeneidad genética del VHD, que da lugar a espectros de diferentes mutantes en un mismo individuo conocidos como quasiespecies virales. Las quasiespecies son conjuntos de poblaciones de moléculas de ARN o ADN con secuencias de nucleótidos estrechamente relacionadas pero ligeramente diferentes (variantes genéticas) (Figura 9).^{105,106} Sin embargo, la heterogeneidad no es uniforme para todo el genoma; la secuencia de la ribozima y el dominio de unión a ARN del HDAg están altamente conservados, mientras que la región C-terminal del L-HDAg es muy divergente.¹⁰⁷

La distribución de las poblaciones de mutantes en la quasiespecies es dinámica y se ve influenciada por el ambiente en el que se multiplican los virus en un organismo infectado (por ejemplo, respuesta inmune, alteraciones metabólicas...) y por las intervenciones externas (por ejemplo, tratamientos antivirales e inmunoterapia), que hacen variar la capacidad replicativa de los mutantes (aptitud biológica o *fitness*). Por un lado, esta dinámica es una consecuencia de los principios darwinianos de la selección de variantes genéticas de un virus, es decir, la competición y el dominio de las variantes más adaptadas a cada ambiente particular. Además de esta selección positiva, existe la deriva genética, la cual juega un papel importante, ya que, se acentúa durante la transmisión del virus, proceso que a menudo constituye un cuello de botella poblacional, cuya intensidad

depende del modo de transmisión (Figura 10).¹⁰⁸ A lo largo del tiempo, el proceso de los sucesivos cuellos de botella poblacionales y los episodios tanto de selección positiva como de selección negativa (eliminación de las variantes con menor *fitness*), conducen a las formas dominantes que se observan durante los cursos clínicos de la hepatitis delta crónica y pueden desempeñar un papel importante en el escape viral del ataque inmune del huésped y en la recaída clínica.^{101,109} La diversidad genética de una quasispecies se puede estudiar mediante la propiedad que cuantifica la diversidad y la frecuencia de haplotipos (conjunto de genomas con la misma secuencia nucleotídica, que se pueden considerar la misma variante genética), independientemente del tamaño de la población viral, que se define como complejidad de una quasispecies.¹⁰⁸

Además, existe un fenómeno llamado transcomplementación, que ocurre cuando el ciclo infeccioso de un virus o de una variante genética del virus se ve reforzado o respaldado por las proteínas expresadas por otro virus o por otra variante. Por lo tanto, el virus o variante auxiliar puede mejorar o modificar las propiedades del virus o variante dependiente.¹¹⁰

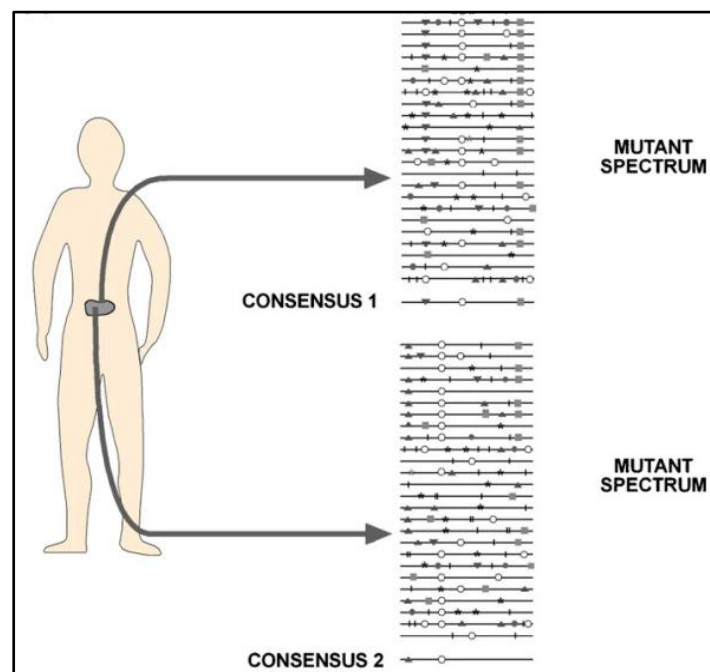


Figura 9. Quasispecies virales. Individuo infectado con múltiples poblaciones virales replicantes, incluso dentro del mismo órgano, que forman una quasispecies. Se representan dos espectros mutantes con sus secuencias consenso. Cada línea horizontal representa un genoma y cada símbolo en una línea representa un tipo de mutación. Una quasispecies viral real puede contener miles de genomas continuamente mutando que conforman enormes nubes de mutantes. Figura modificada.¹⁰⁸

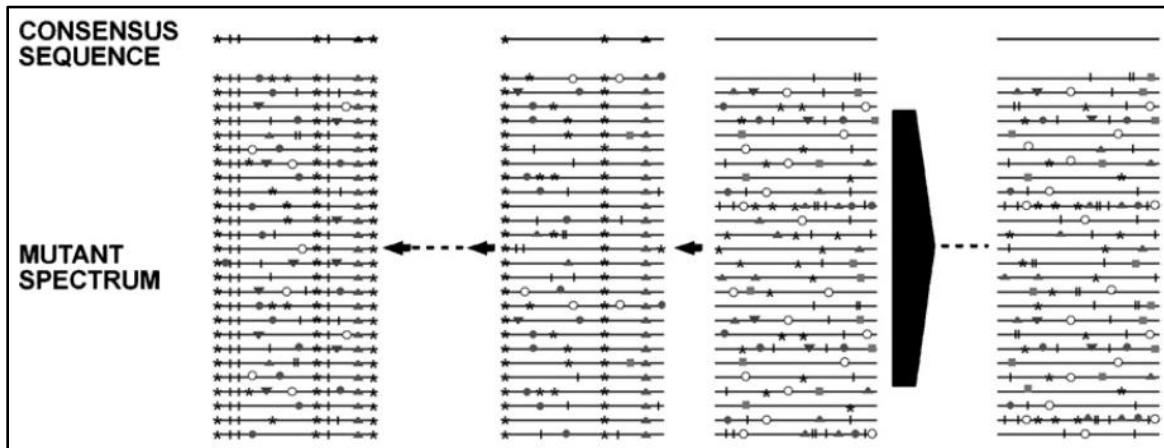


Figura 10. Variación del *fitness*. La evolución de la quasispecie está guiada por el gradiente de *fitness* del ambiente, es decir, los mutantes con mayor *fitness* pueden dominar la quasispecie por replicación competitiva. Las flechas pequeñas representan pasos repetidos del cuello de botella que resultan en una disminución del *fitness*. La flecha grande representa pasos a poblaciones más grandes que generalmente resultan en un aumento del *fitness*. Figura modificada.¹⁰⁸

1.7 Patogenia del VHD

El VHD replica solo en el hígado. Su infección se asocia con una amplia gama de manifestaciones clínicas, desde hepatitis asintomáticas hasta hepatitis fulminantes.¹¹¹

1.7.1 Coinfección y superinfección por VHD-VHB

Los pacientes infectados por el VHD y el VHB presentan la forma más grave de hepatitis viral con mayor daño hepático. El riesgo de desarrollar cirrosis y hepatocarcinoma es tres veces mayor, el riesgo de presentar una descompensación hepática es dos veces mayor, y aumenta también el riesgo de presentar fibrosis y/o hepatitis fulminante y, por último, el riesgo de mortalidad es doble en comparación con la monoinfección por el VHB.^{112,113}

Esta infección se puede dar de dos formas distintas: de manera simultánea con la infección del VHB (coinfección) o en individuos ya infectados anteriormente por el VHB (sobreinfección). En la Figura 11 se muestra la evolución típica de los marcadores serológicos y virológicos en pacientes coinfectados versus pacientes sobreinfectados. En la coinfección se observa un patrón de hepatitis aguda, que varía de leve a severa, o incluso fulminante, y que se autolimita en más del 90% de los adultos inmunocompetentes. Esto significa que la tasa de progresión a la cronicidad, en adultos inmunocompetentes, es baja. Los niveles de alanina aminotransferasa (ALT) y aspartato aminotransferasa (AST) pueden seguir un curso bifásico típico, a menudo se observan dos picos, con varias semanas de diferencia, ya que la infección por el VHB debe primero establecerse antes de que el VHD pueda comenzar a replicarse. En cambio, la

sobreinfección se cronifica en la práctica totalidad de los casos.¹¹³ Una vez que se establece la infección crónica por el VHD, la enfermedad hepática preexistente empeora a pesar de que la replicación del VHB se suprime en muchos casos. Los niveles de ALT y AST se elevan persistentemente en la mayoría de los casos.^{114,115}

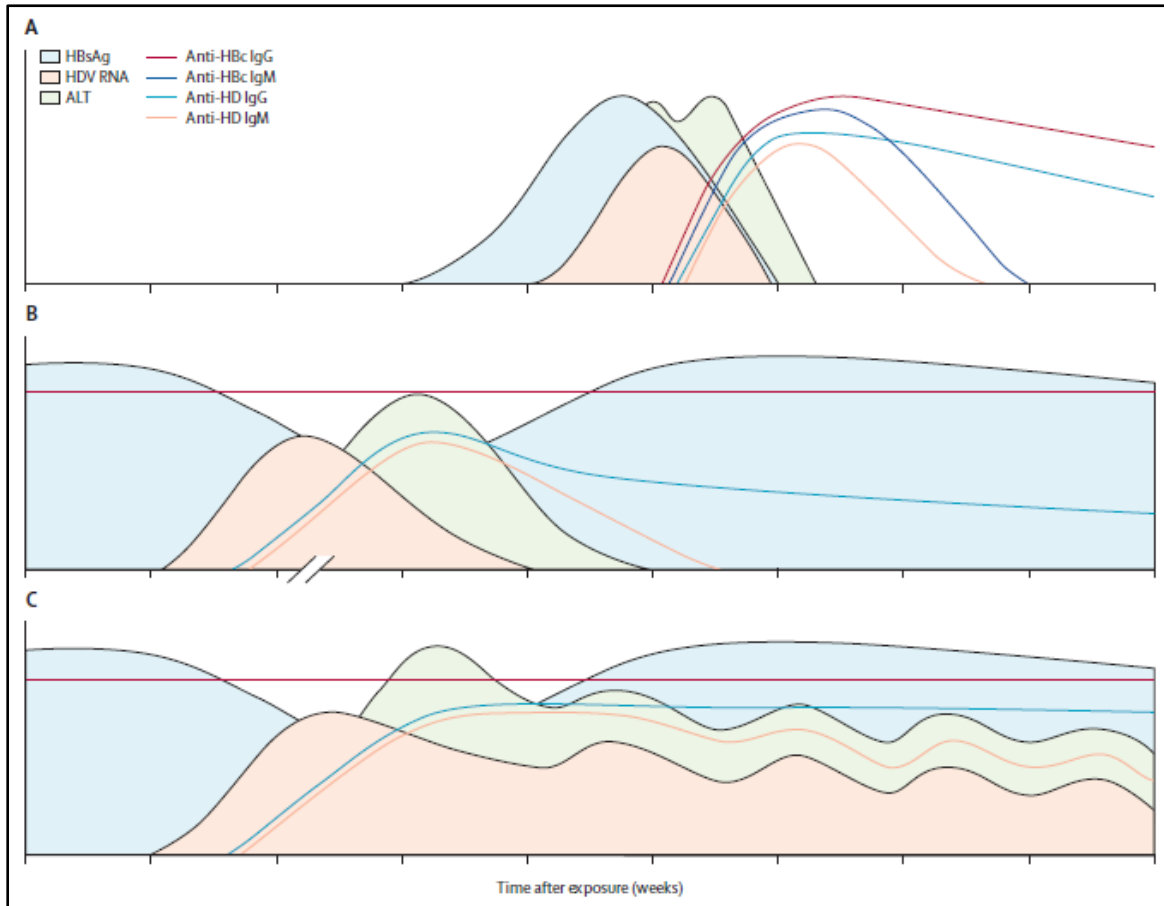


Figura 11. Evolución típica de los marcadores serológicos y virológicos en la infección por el VHD. A) Coïnfección del VHB-VHD, lo que resulta en la eliminación de ambos virus en casi todos los pacientes. B) Sobreinfección por el VHD en un portador del VHB con resultado autolimitado. El aclaramiento espontáneo del ARN-VHD puede tardar años en ocurrir (indicado por la ruptura en el eje x) y, en algunos casos, puede ocurrir la pérdida del HBsAg (no representado). C) Sobreinfección por el VHD en un portador del VHB con replicación viral persistente crónica, el resultado más común después de la sobreinfección.²⁴

1.7.2 Hepatitis delta aguda y crónica

La infección delta aguda puede ocurrir a través de la coïnfección o sobreinfección. El curso clínico de una coïnfección aguda VHB-VHD es similar al de la infección aguda por el VHB. Sin embargo, la sobreinfección aguda por VHD tiene un curso clínico más grave en comparación con la coïnfección aguda.^{113,116} Puede complicarse en pacientes con hepatitis B crónica, con necrosis masiva de hepatocitos que conduce a la muerte en la gran mayoría de los pacientes a menos que se realice un trasplante de hígado. Durante

la hepatitis delta aguda, los marcadores del VHD no permiten una distinción confiable entre coinfección y superinfección. La diferencia entre coinfección y superinfección se establece en función de la presencia de anticuerpos IgM anti-HBc que serán positivos en el contexto de una infección aguda por VHB-VHD (como es el caso de la coinfección) y seguirán siendo negativos o muy bajos en pacientes sobreinfectados (con una hepatitis crónica B de base).^{49,113}

La hepatitis delta se considera crónica cuando se detecta el anticuerpo anti-VHD más de 6 meses después de la infección, con el ARN-VHD detectable en suero o con el HDAg detectable por inmunohistoquímica.^{117,118} Esta es la hepatitis viral crónica con una progresión más rápida hacia la cirrosis y fibrosis, y un mayor riesgo de descompensación y mortalidad.¹¹² De hecho, del 10-15% de los pacientes con infección crónica pueden desarrollar cirrosis dentro de los 2 años posteriores a la infección y hasta el 70-80% después de los 5-10 años.¹¹⁹ Existe una amplia gama de manifestaciones clínicas, desde casos asintomáticos hasta hepatitis rápidamente progresiva. Es clínicamente indistinguible de la infección crónica por el VHB, aunque la replicación persistente del VHD agrava la progresión de la enfermedad hepática.^{49,113}

1.8 Efecto inhibitorio del VHD sobre el VHB

El VHD parece suprimir fuertemente la replicación del VHB en muchos casos, aunque se sabe poco sobre el mecanismo de esta interacción.¹²⁰ Los estudios realizados en modelos experimentales han demostrado que el VHD, que necesita el HBsAg para su empaquetado y liberación, puede reprimir la replicación del VHB manteniendo la producción de HBsAg. Se ha observado un aumento de la proporción de HBsAg/ADN-VHB en las células coinfectadas con VHB-VHD, en comparación con células mono infectadas con VHB. Se cree que este fenómeno es independiente del sistema inmune adaptativo ya que es reproducible *in vitro*.¹²¹⁻¹²² Esta observación ha sido confirmada en pacientes, aunque los patrones de dominancia viral parecen fluctuar a lo largo del tiempo. Cabe decir, que la mayoría de estos pacientes muestran un patrón de dominancia del VHD, con una disminución significativa en el ADN-VHB en comparación con la de los pacientes mono infectados por el VHB.^{123,124}

La inhibición de la replicación del VHB puede deberse a una interacción directa entre los dos virus dentro de las células. El L-HDAg podría inhibir la transcripción del ADN circular cerrado covalentemente (ADNccc) al interactuar con la ARN pol II.^{79,81,120} Además, la

posibilidad de una regulación epigenética de la actividad transcripcional del ADNccc por parte del HDAg se ha sugerido tanto *in vitro* como en muestras de pacientes, lo que aumenta la posibilidad de una transcripción diferencial del ARNm de Pre-S/S frente a ARNm pregenómico (usado como molde para la transcripción reversa durante la replicación del VHB).¹²⁶ Por otro lado, se ha demostrado que ambas isoformas del HDAg interactúan y reprimen fuertemente ambas secuencias potenciadoras (*enhancers*) del VHB (Enh1 y Enh2), con un impacto directo en su replicación.¹²⁷ También, el HDAg, al ser una proteína de unión a ARN, interactúa con ARN celulares específicos y puede unirse al ARNm del VHB y afectar selectivamente su estabilidad.¹²⁸⁻¹³⁰ Asimismo, en pacientes infectados por el VHB, el ADN-VHB integrado es una fuente abundante de HBsAg, incluso en ausencia de replicación del VHB.¹³¹⁻¹³³ Se ha demostrado que el HBsAg derivado del ADN-VHB integrado puede promover el ensamblaje y la liberación de partículas infecciosas del VHD, aunque todavía no *in vivo*.¹³⁴ Por último, el VHD puede dificultar la replicación del VHB de manera indirecta, a través de la respuesta inmune innata, estimulando la expresión de los genes estimulados por interferón (ISG) y la vía del factor de respuesta sérica (SRE).¹³⁵ Además, el L-HDAg puede activar el promotor MxA, inducible por interferón alfa (IFN- α), y la replicación del VHB puede ser inhibida por la actividad antiviral de la proteína MxA.¹²⁷

1.9 Tratamiento antiviral

Actualmente no hay tratamientos antivirales específicos contra el VHD. Esto se debe principalmente al hecho de que el virus no codifica actividades enzimáticas y depende completamente de la maquinaria de la célula huésped para su replicación. Por tanto, aún se requiere una mejor comprensión de la biología molecular y celular del VHD para encontrar un tratamiento eficaz.

El objetivo del tratamiento contra el VHD es prevenir la progresión de la enfermedad hepática y la muerte. Esto se conseguiría con la pérdida del HBsAg que llevaría a una respuesta virológica sostenida del VHD (ARN-VHD negativo 6 meses después de suspender el tratamiento), es decir, una cura funcional, que se asocia con un pronóstico excelente.¹³⁶ En la práctica, una disminución del ARN-VHD de al menos 2 log con normalización de las ALT al final del tratamiento se considera un punto final de eficacia valioso. El mantenimiento de la respuesta virológica sostenida o alcanzar la indetectabilidad del ARN-VHD en la semana 24 o 48 después del final del tratamiento son puntos finales adicionales; sin embargo, pueden ocurrir recaídas tardías.¹³⁷

1.9.1 Tratamiento actual

El IFN- α sigue siendo el único fármaco recomendado por las directrices internacionales para el tratamiento de la hepatitis delta crónica, el cual se introdujo en la década de 1980.¹³⁸ Además, en 2006 se introdujo el interferón pegilado (PEG-IFN- α). Por tanto, actualmente el tratamiento para el VHD es PEG-IFN- α durante 48 semanas sin otras opciones terapéuticas en pacientes que no responden a este tratamiento. Los análogos de nucleós(t)idos (NA) utilizados para tratar la hepatitis B crónica, principalmente Entecavir (ETV) y Tenofovir (TDF), no han sido aprobados para la infección por VHD.¹³⁶ Tanto el IFN- α como el PEG-IFN- α suprimen la viremia del VHD en algunos pacientes pero no seroconvierten el HBsAg. Las tasas de respuesta virológica sostenida varían entre el 14-50% para el IFN- α y entre el 17-44% para el PEG-IFN- α .^{24,139,140} Los niveles de ALT y AST se normalizan en solo el 40-70% de los pacientes tratados y la recaída ocurre en el 60-97%.¹⁴¹

Un ensayo clínico multicéntrico en 91 pacientes (HIDIT-1) probó el PEG-IFN- α solo, el PEG-IFN- α en combinación con adefovir (ADF) o el ADF solo, durante 48 semanas. Se observó el ARN-VHD indetectable en el 31%, 26% y 0% de los pacientes, respectivamente, 24 semanas después del final del tratamiento.¹⁴² Sin embargo, el 56% de estos pacientes con ARN-VHD indetectable inicialmente recayeron durante el seguimiento a largo plazo, lo que indica que 24 semanas después del tratamiento es demasiado temprano como punto final para la definición de una respuesta virológica sostenida.¹⁴³

En otro estudio sucesivo (HIDIT-2), el tratamiento con PEG-IFN- α se prolongó a 96 semanas y se administró solo o en combinación con TDF. Los resultados provisionales en la semana 24 después del tratamiento mostraron indetectabilidad del ARN-VHD en el 23% de los pacientes tratados con PEG-IFN- α y en el 30% de los pacientes tratados con PEG-IFN- α y TDF, lo que indica que la prolongación del tratamiento no mejora el resultado en comparación con el HIDIT-1.¹⁴⁴ Además, el tratamiento prolongado se asoció con una alta frecuencia de eventos adversos graves incluyendo sintomatología gripal, eventos neuropsiquiátricos, anemia y trombocitopenia.^{142,145-147} Sin embargo, un seguimiento de 5 años del estudio HIDIT-1 reveló que más del 50% de los pacientes con ARN-VHD indetectable a los 6 meses posteriores al tratamiento, desarrollaron ARN-VHD detectable al menos una vez durante el seguimiento.¹⁴³

La ineficacia del IFN- α podría estar asociada con la capacidad recientemente descrita del L-HDAg para bloquear la señalización del IFN- α mediante la inhibición de la cascada

Janus quinasa / Signal Transducer and Activator of Transcription (JAK/STAT). El L-HDAg es capaz de inhibir la fosforilación de los factores STAT-1 y STAT-2, así como su translocación al núcleo, necesaria para la activación/expresión de los ISG, esenciales para combatir la infección viral.^{148,149}

Por último, debido a que el VHD usa el HBsAg del VHB para propagarse, los programas de vacunación contra el VHB también son considerados preventivos y efectivos para combatir la infección por el VHD. A partir de 1990, tras iniciar la campaña de vacunación contra el VHB, la incidencia del VHB y del VHD declinó a nivel mundial, sobre todo en los países desarrollados.^{150,151}

1.9.2 Nuevos tratamientos

Las nuevas terapias antivirales en desarrollo se dirigen a factores virales o celulares, tratando de inhibir la unión y entrada del virus, su replicación o el ensamblaje y liberación de viriones (Figura 12).

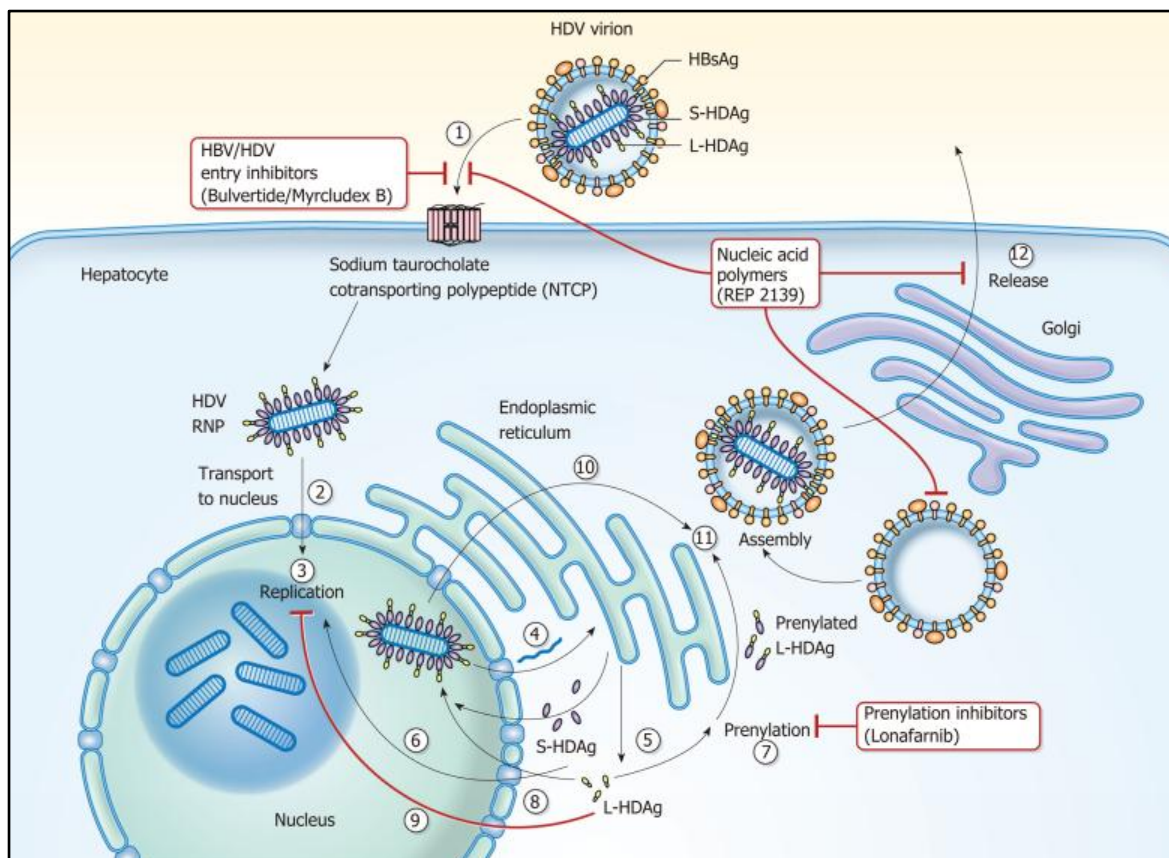


Figura 12. Ciclo replicativo del VHD y nuevos tratamientos. 1) El virión del VHD se adhiere al hepatocito. Proceso inhibido por Bulevirdite y REP-2139. 2) La RNP se transloca al núcleo. 3) Replicación en círculo rodante del genoma del VHD. 4) El ARNm se transporta fuera del núcleo al RE. 5) El ARNm se traduce en el RE a S-HDAg y L-HDAg. 6) El S-HDAg se transporta al núcleo. 7) El L-HDAg se somete a una prenilación antes del ensamblaje de las partículas virales. Proceso inhibido por Lonafarnib. 8) El S-HDAg promueve la replicación del VHD en el núcleo. 9) El L-HDAg inhibe la replicación del VHD en el núcleo. 10) Los HDAG se asocian con nuevas transcripciones de ARN genómico para formar nuevas RNP que se exportan al citoplasma. 11) Las RNP se asocian con el HBsAg y se ensamblan en viriones. Proceso inhibido por REP 2139. 12) Los viriones VHD se liberan del hepatocito a través de la red trans-Golgi. Proceso inhibido por REP 2139.¹⁵²

1.9.2.1 Inhibidores de la entrada viral

Bulevirdite (BLV), anteriormente conocido como Myrcludex-B, es un lipopéptido lineal sintético miristoilado de 47 aminoácidos derivado del dominio Pre-S1 de la L-HBsAg, que compite con éste para unirse al NTCP, lo que inhibe la entrada viral, evitando así la reinfección de los hepatocitos y, con el tiempo, conduciendo a la eliminación del virus.¹⁵³ Esta molécula mostró eficiencia *in vitro*¹⁵⁴ y los estudios preclínicos mostraron que obstaculiza eficazmente el establecimiento de la infección en ratones quiméricos humanizados pretratados.^{122,155}

Estudios en pacientes con hepatitis delta crónica tratados con BLV han demostrado que BLV es capaz de bloquear el NTCP como receptor del VHD sin que este pierda su

capacidad transportadora de ácidos biliares. De hecho, aunque en algunos pacientes se ha observado un ligero aumento de las sales biliares, es un tratamiento bien tolerado y con pocos efectos adversos. Su administración, limitada en el tiempo, con o sin combinación con PEG-IFN- α o TDF es eficaz para reducir significativamente el ARN-VHD al final del tratamiento. La mayoría de pacientes mejora la actividad necroinflamatoria del hígado al normalizar las ALT cuando reciben BLV en la dosis adecuada. Sin embargo, muchos pacientes no mantienen una respuesta virológica sostenida al finalizar el tratamiento.¹⁵⁶⁻¹⁶⁰

Datos recientes sugieren que los hepatocitos positivos para el HDAg podrían replicarse permitiendo la persistencia del VHD incluso en presencia del inhibidor de entrada. El recambio celular en un contexto inflamatorio y la persistencia del ADN-VHB integrado como la principal fuente de HBsAg podrían explicar la pérdida limitada del ARN-VHD y posiblemente el rebote viral después del tratamiento.¹⁶¹

Dados estos primeros resultados, BLV recibió la designación de *Orphan Drug Designation* para el tratamiento de la infección por VHD de la Agencia Europea de Medicamentos (EMA) y de la Administración Estadounidense de Alimentos y Medicamentos (FDA). Además de la *PRIME Eligibility* por la EMA. En octubre de 2018, fue nombrado *Breakthrough Therapy Designation* por la FDA. Y por último, destacar que en Francia, 2 mg de BLV al día están disponibles para pacientes con hepatitis delta crónica (*Autorisation temporaire d'utilisation (ATU)*) desde septiembre de 2019.^{136,162}

1.9.2.2 Inhibidores de la isoprenilación

El Lonafarnib (LNF), utilizado previamente como un tratamiento contra el cáncer, se encuentra actualmente en estudio en pacientes con hepatitis delta crónica. Es un inhibidor de la farnesil transferasa, una enzima clave en la prenilación que promueve la farnesilación del residuo C-211 del L-HDAg y, en consecuencia, el ensamblaje y la liberación de partículas virales. Por lo tanto, la reducción de los niveles de ARN-VHD en suero está directamente relacionada con el ensamblaje de las partículas virales del VHD (similar al efecto de los NA para la liberación del VHB que evitan la transcripción inversa y bloquean la producción del virus).^{163,164} El bloqueo del ensamblaje de las partículas virales da como resultado la acumulación intrahepatocelular de intermediarios replicativos del VHD pero no conduce principalmente a una reducción de los hepatocitos infectados en el hígado. Sin embargo, podría causar la renovación acelerada de los hepatocitos infectados a través de los efectos citotóxicos directos de los intermediarios replicativos acumulados o la destrucción de células mediada por el sistema inmunitario.¹⁶⁵

Estudios en pacientes con hepatitis delta crónica tratados con LNF han demostrado que se consigue reducir significativamente los niveles de ARN-VHD de una manera dosis-dependiente pero se observan rebotes virológicos al suspender el tratamiento. LNF presenta efectos secundarios gastrointestinales, que se reducen cuando se administra juntamente con ritonavir (RTV), ya que inhibe la metabolización de LNF lo que permite administraciones a dosis más bajas. LNF se ha administrado en combinación con PEG-IFN- α o PEG-IFN- λ y TDF o ETV para la supresión del VHB, pero continúa teniendo los mismos efectos secundarios.^{163,166-168}

LNF ha sido concedido como *Orphan Drug Designation* por la FDA y la EMA, *Fast Track Designation and Breakthrough Therapy Designation* por la FDA y *PRIME Eligibility Designation* por la EMA.^{136,162}

1.9.2.3 Polímeros de ácido nucleico

Los polímeros de ácido nucleico (NAP) son oligonucleótidos fosforotioados que tienen efectos antivirales a través de varios mecanismos.¹⁶⁹ Pueden bloquear la entrada viral gracias a la fosforotioación o anfipaticidad de los NAP que les permite interactuar con las superficies hidrofóbicas de las glicoproteínas, gracias a su estructura similar a los HSPGs.^{170,171} También pueden intervenir en la inhibición de la liberación del HBsAg,¹⁷²⁻¹⁷⁴ en la reducción del HBsAg intracelular a través de la inhibición del ensamblaje de partículas subvirales,¹⁷⁵ y posiblemente, en la interacción con el S-HDAg y el L-HDAg que conduce a la inhibición del ciclo de replicación del VHD.¹⁷⁶

Entre ellos, REP-2139 es un compuesto que inhibe la unión del virus a los HSPGs, bloquea el ensamblaje de partículas subvirales, evita la liberación del HBsAg y permite su eliminación y la restauración del control funcional de la infección cuando se combina con diversas inmunoterapias.¹⁷⁷⁻¹⁷⁹

Estudios en pacientes con hepatitis delta crónica tratados con REP-2139-Ca en combinación con PEG-IFN- α demuestran una reducción significativa de los niveles de ARN-VHD. Además, REP-2139-Ca parece ser la única terapia en investigación para reducir el HBsAg rápidamente. Los efectos secundarios incluyeron pirexia, escalofríos, trombocitopenia y leucopenia.^{180,181}

1.9.2.4 Interferon lambda

Interferon lambda (IFN- λ) es un IFN tipo III con características estructurales, características del receptor y actividades biológicas distintas del IFN- α . IFN- λ induce la

expresión de ISG para inducir una respuesta inmune antiviral de amplio espectro. La unión de IFN- λ a los receptores de IFN tipo III da como resultado la dimerización del receptor, lo que conduce a la activación de las vías de transducción de señales mediadas por JAK y STAT, similar a las vías estimuladas por IFN- α .^{182,183} La señalización de IFN- λ se localiza en el hígado y puede tener un papel protector.¹⁸⁴⁻¹⁸⁶ Se ha demostrado que IFN- λ tiene un efecto antiviral contra el VHD, reduciendo los marcadores de infección intrahepática del VHD, comparable con IFN- α , en ratones con hígados humanizados infectados con VHB-VHD.¹⁸⁷

Estudios en pacientes con hepatitis delta crónica tratados con IFN- λ y TDF o ETV, para la supresión del VHB, demostraron que este tratamiento fue bien tolerado y los pacientes que recibieron tratamiento previo con IFN- α informaron menos efectos secundarios con IFN- λ . Además, la respuesta virológica sostenida y la normalización de las ALT se logró en algunos pacientes pero era dosis-dependiente.¹⁸⁸

1.9.2.5 Otros tratamientos

También se ha planteado el uso de la ribozima viral como diana para el desarrollo de antivirales específicos, ya que, es el único elemento viral con actividad catalítica y de vital importancia para la replicación y procesamiento del genoma viral. La inhibición específica de la actividad de la ribozima se ha logrado *in vitro* con pequeños ARN de interferencia (ARNsi), evitando la escisión del ARN multimérico y la producción funcional del ARN genómico y antigénico. Sin embargo esta opción presenta limitaciones debido a la propia biología del sistema y la rápida replicación del material genético del VHD.¹⁸⁹⁻¹⁹¹ Algunos antibióticos y sus complejos de cobre (Cu) (II) podrían interferir con la ribozima.^{192,193}

También se han propuesto oligonucleótidos antisentido complementarios a los sitios de autoescisión de la ribozima genómica como un posible enfoque terapéutico.^{194,195} La ventaja de estos oligonucleótidos antisentido en comparación con los ARNsi es que son capaces de entrar en el núcleo e inhibir la ribozima. En cambio, los ARNsi solo pueden interactuar y promover la degradación de los ARNm, lo que impediría la expresión de los HDAg.

HIPÓTESIS

2. HIPÓTESIS

El estudio de las quasiespecies virales se ha basado tradicionalmente en la clonación y la secuenciación directa, y más recientemente en una tecnología más sensible y reproducible conocida como secuenciación masiva (NGS).¹⁹⁶⁻¹⁹⁹

Esta tesis se ha centrado en el estudio de la quasiespecies del VHD, sobre la cual existen pocos estudios y por tanto, todavía no se dispone de datos muy concluyentes sobre sus implicaciones.

La quasiespecies del VHD circulante muestra dos poblaciones virales, como resultado de la edición por la enzima ADAR-1: una población con genomas no editados (que codifican el S-HDAg, el cual se requiere para la replicación viral) y otra población con genomas editados (que codifican el L-HDAg, que inhibe la replicación viral).^{33,78} Esto sugiere que podría ser importante analizar el porcentaje de ambos genomas al monitorizar pacientes con hepatitis delta crónica. En un estudio preliminar reciente de nuestro grupo,¹⁰⁴ se utilizó la NGS para cuantificar porcentajes de genomas del VHD no editados y editados en las quasiespecies circulantes de 3 pacientes. No obstante, se necesitan más estudios para consolidar este tipo de aproximaciones metodológicas, para cuantificar de manera fiable genomas no editados y editados.

Además de la región del genoma del VHD que codifica el antígeno de la hepatitis delta (HDAg), el cual puede ser editado por ADAR-1, este genoma tiene una segunda región fundamental para la replicación viral, la región de la ribozima. La ribozima del VHD es el único ejemplo de un ARN catalítico asociado con un virus animal, el cual cataliza la autoescisión de las copias de su genoma. La secuencia de la ribozima no es codificante y es activa *in vitro* en ausencia de proteínas.⁹⁷ El análisis de la conservación o variabilidad de esta región de forma profunda, como permite la NGS, no se ha estudiado hasta la actualidad.

En este sentido, es importante estudiar la conservación o variabilidad de las dos regiones mencionadas. Por tanto, este estudio puede ser de utilidad a nivel clínico, tanto para monitorizar pacientes con hepatitis delta crónica como para encontrar nuevas estrategias terapéuticas.

Debido a ello, se plantean las siguientes hipótesis:

- La metodología de NGS es más precisa que la clonación y la secuenciación Sanger para cuantificar porcentajes de poblaciones de genomas editados y no editados en el VHD.
- Puesto que los genomas no editados codifican para el S-HDAg, que activa la replicación viral, la proporción de estos genomas en la quasiespecie circulante del VHD es superior a la de genomas editados.
- El fenómeno de edición podría no ser exclusivo en una sola posición. Con lo cual, los cambios de nucleótidos de tipo transición, en la misma región del genoma del VHD, son más frecuentes que las transversiones.
- Dado que el HDAg es una proteína multifuncional con varios dominios que experimentan modificaciones postraduccionales, la conservación de su secuencia codificante no es homogénea a lo largo de toda su longitud.
- La región de la ribozima del VHD tiene una función esencial para la replicación viral, por lo cual, su secuencia debería ser poco variable.

OBJETIVOS

3. OBJETIVOS

El objetivo principal de esta tesis ha sido estudiar la quasiespecies del VHD mediante NGS, cuantificando la edición de su genoma, que da lugar a los antígenos S-HDAg y L-HDAg, y la conservación o variabilidad de las regiones funcionales del virus, correspondientes al HDAg y a la ribozima, que es el único dominio de este virus con actividad enzimática propia y esencial para la replicación viral.

Para ello se plantean como objetivos secundarios:

- Validar el método de NGS para cuantificar porcentajes de poblaciones de genomas editados y no editados en el VHD, comparándolo con el método de clonación molecular y secuenciación Sanger.
- Confirmar si la edición es un fenómeno general que tiene lugar en toda la región codificante del HDAg analizada.
- Analizar el grado de conservación o variabilidad de la quasiespecies del VHD en las regiones analizadas del HDAg y de la ribozima, compatible con su papel esencial en la replicación viral.
- Analizar si existen patrones de mutaciones a nivel aminoacídico en la región del HDAg analizada.

MATERIAL Y MÉTODOS

4. MATERIAL Y MÉTODOS

4.1 Cuantificación del ADN del VHB

El ADN del VHB se cuantificó por PCR cuantitativa (qPCR) en el sistema COBAS 6800, mediante el ensayo *cobas® HBV Test* (Roche Diagnostics, Mannheim, Alemania). El sistema utiliza un proceso universal de preparación de muestras y reactivos para aislar, purificar y extraer los ácidos nucleicos totales a partir de 650 µL de plasma extraído en tubo de EDTA. La qPCR se utiliza para amplificar y simultáneamente cuantificar de forma absoluta el producto de la amplificación. Esta técnica tiene un intervalo de linealidad desde 10 hasta 1,00E+09 UI/mL y un límite de detección de 2,7 UI/mL. En esta cuantificación no afecta que la muestra esté lipémica, icterica o hemolizada.²⁰⁰

4.2 Determinación de los marcadores serológicos (HBsAg, HBeAg, anti-VHD)

El HBsAg y el HBeAg se determinaron a partir de 150 µL de plasma extraído en tubo de EDTA o suero, usando inmunoensayos electroquimioluminiscentes comerciales (*Elecsys® HBsAg II kit* y el *Elecsys® HBe Ag kit*, respectivamente), en el sistema COBAS 8000 (Roche Diagnostics, Rotkreuz, Suiza). En estas determinaciones no afecta que la muestra esté lipémica, icterica o hemolizada. A partir de sustratos estables, se generan productos capaces de emitir fotones, cuyo origen es electroquímico. El anticuerpo utilizado recubre unas micropartículas imantadas, que tras la formación del complejo antígeno-anticuerpo, se fijan a un electrodo por magnetismo. Dicho anticuerpo está conjugado con un marcador capaz de emitir fotones cuando se aplica una pequeña diferencia de potencial sobre el electrodo. La energía lumínica se detecta en un fotomultiplicador.^{201,202} En estas determinaciones no afecta que la muestra esté lipémica, icterica o hemolizada.

Los anticuerpos anti-VHD se determinaron cualitativamente a partir de 170 µL de plasma extraído en tubo de EDTA o suero, con el *HDV Ab kit* (DIA. PRO. Diagnostics Bioprobes Srl, Sesto San Giovanni, Italia) utilizando el sistema LIAISON® XL MUREX (DiaSorin, Saluggia, Italia) mediante la tecnología de inmunoensayo de quimioluminiscencia indirecta (CLIA). El antígeno recombinante específico para el VHD se usa para recubrir las partículas magnéticas (fase sólida). Durante la primera incubación, los anticuerpos

anti-VHD presentes se unen a la fase sólida a través del antígeno VHD recombinante. Durante la segunda incubación, los anticuerpos monoclonales de ratón para la inmunoglobulina G (IgG) e inmunoglobulina M (IgM) humanas, ambos unidos a un derivado de isoluminol (conjugado de isoluminol-anticuerpo), reaccionan con los anticuerpos anti-VHD ya unidos a la fase sólida. Después de cada incubación, el material no unido se elimina con un ciclo de lavado. Posteriormente, se añaden unos reactivos y se induce así una reacción de quimioluminiscencia instantánea. La señal luminosa, y por lo tanto, la cantidad de conjugado isoluminol-anticuerpo, se mide mediante un fotomultiplicador y es indicativa de la presencia de anticuerpos anti-VHD. En esta técnica no afecta que la muestra esté lipémica, icterica o hemolizada.²⁰³

4.3 Extracción del ARN del VHD

En el primer estudio, la extracción del ARN del VHD se realizó con el *Total Nucleic Acid Isolation Kit* (TNAI) (Roche Diagnostics, Mannheim, Alemania) utilizando el sistema automatizado COBAS AmpliPrep (Roche Diagnostics, Rotkreuz, Suiza). En cambio, en el segundo estudio, se realizó con el *MagNA Pure LC Total Nucleic Acid Isolation Kit* (Roche Diagnostics, Mannheim, Alemania) utilizando el sistema automatizado MagNA Pure LC (Roche Applied Science, Indianapolis, Estados Unidos).

Ambos kits se basan en el mismo principio para la extracción de ácidos nucleicos (Figura 13). A partir de 250 µL de plasma extraído en tubo de EDTA, se lleva a cabo la lisis de las partículas víricas mediante incubación a alta temperatura con una proteinasa K y un tampón caotrópico de unión/lisis. Esto libera los ácidos nucleicos y protege al ARN liberado de las ribonucleasas (RNasa) presentes en el plasma. Después, se añaden las partículas de vidrio magnéticas (MGP), y los ácidos nucleicos presentes en la muestra se unen a su superficie. Las sustancias no unidas, tales como sales, proteínas y otras impurezas, se eliminan al lavar las MGP. Por último, se eluyen los ácidos nucleicos adsorbidos a temperatura elevada con un tampón bajo en sal.²⁰⁴

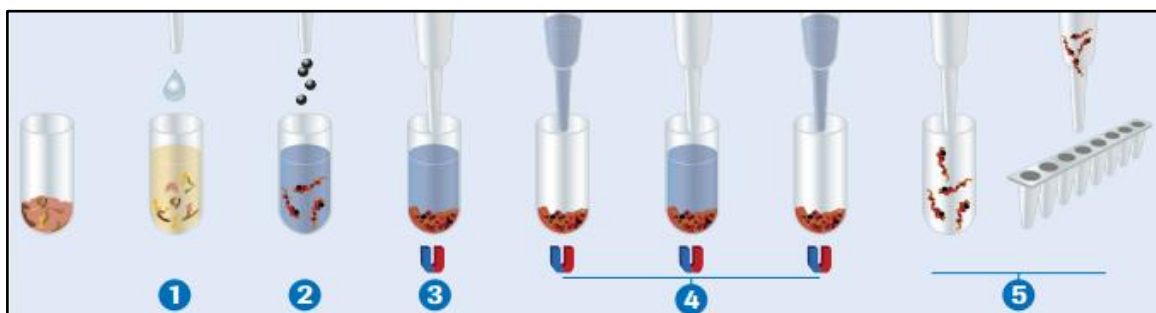


Figura 13. Extracción de ácidos nucleicos. 1) Lisis de la muestra, inactivación de nucleasas y liberación de ácidos nucleicos. 2) Adición de partículas de vidrio magnéticas (MGP) y unión de los ácidos nucleicos a las MGP. 3) Separación magnética de los complejos MGP/ácido nucleico de la muestra lisada. 4) Múltiples pasos de lavado. 5) Elución de los ácidos nucleicos purificados.²⁰⁵

4.4 Cuantificación del ARN del VHD

En el primer estudio, el ARN del VHD se cuantificó por un método *in-house* utilizando un estándar genómico completo del ARN del VHD, que tiene un intervalo de linealidad desde $1,00E+02$ hasta $1,00E+08$ copias/ μL .²⁰⁶ En cambio, en el segundo estudio, el ARN del VHD se cuantificó utilizando el estándar internacional del ARN del VHD de la Organización Mundial de la Salud (*1st World Health Organization International Standard for Hepatitis D Virus RNA for Nucleic Acid Amplification Techniques-based assays; PEI code number: 7657/12*). Este estándar internacional tiene un intervalo de linealidad desde $5,75E+02$ hasta $5,75E+05$ UI/mL y un límite de detección de $5,75E+01$ UI/mL.²⁰⁷

En ambos estudios, debido a la alta tasa de emparejamiento de bases internas (un 70%) del genoma del VHD,^{5,32} es esencial maximizar la ruptura de los puentes de hidrógeno para facilitar el cebado en el paso de transcripción inversa (RT). Por esta razón, se consideró la aplicación de un fuerte choque térmico al ARN del VHD, obtenido de plasma con EDTA tal como se explica en el apartado anterior. Este choque térmico consistió en una desnaturalización con $10 \mu\text{L}$ de ARN del VHD a 95°C durante 10 minutos, seguido inmediatamente de un enfriamiento a -80°C durante 5 minutos, para minimizar la posible recuperación del emparejamiento de bases internas naturales (renaturalización).²⁰⁶ Seguidamente, se realizó la cuantificación del ARN del VHD desnaturalizado utilizando la tecnología de *Florescence Resonance Energy Transfer* (FRET) con los cebadores y las sondas reportados por Schaper *et al.* (2010)¹²⁵ (Tabla 1), mediante RT-qPCR en un solo paso utilizando el kit *LightCycler RNA Master HybProbe* (Roche Diagnostics, Mannheim, Alemania) en el sistema LightCycler 2.0 (Roche Diagnostics, Mannheim, Alemania). Los componentes utilizados en la RT-qPCR y el protocolo seguido se resumen en la Tabla 2 y Tabla 3, respectivamente.

Cebadores y sondas	Posiciones	Secuencia (5'- 3')
Cebador DP15	815 -836	CGTCCTTCTTTCTCTTCGGGT
Cebador DP2	900-919	AGTGGCTCTCCCTTAGCCAT
Sonda fluorescente HDV-FL	848-870	TCCTCCTTCGGATGCCAGGTCG - fluoresceína
Sonda fluorescente HDV-LC	873-895	LC640 - CCGCGAGGAGGTGGAGATGCCAT - fosfato

Tabla 1. Cebadores y sondas utilizados en la cuantificación del ARN del VHD.¹²⁵

Componentes	Concentración	Volumen (µL)
Agua estéril	-	0,6
Acetato de manganeso [Mn(OAc) ₂]	50 mM	1,3
Cebador DP15	100 µM	0,1
Cebador DP2	100 µM	0,1
Sonda fluorescente HDV-FL	20 µM	0,2
Sonda fluorescente HDV-LC	20 µM	0,2
<i>LightCycler® RNA Master HybProbe</i> (ADN polimerasa Tth, tampón de reacción y dNTPs)	2,7 x	7,5

Tabla 2. Componentes, concentración y volumen utilizados en la cuantificación del ARN del VHD. dNTPs, desoxinucleótidos trifosfato.

	Temperatura	Tiempo		
Choque térmico	98°C	5 minutos		
	-80°C	5 minutos		
LightCycler 2.0				
Programa	Temperatura	Tiempo	Modo de adquisición de fluorescencia	Modo de análisis
RT	61°C	20 minutos	Ninguno	Ninguno
Desnaturalización	95°C	10 minutos	Ninguno	Cuantificación
Amplificación (45 ciclos)	95°C	2 segundos	Ninguno	Cuantificación
	55°C	15 segundos	Único	
	72°C	15 segundos	Ninguno	
Enfriamiento	40°C	30 segundos	Ninguno	Ninguno

Tabla 3. Tiempos y temperaturas del protocolo de cuantificación del ARN del VHD. RT, transcripción inversa.

4.5 Amplificación del ARN del VHD para obtener librerías de amplicones para analizar por secuenciación masiva (NGS)

En el primer estudio, la región seleccionada para el análisis cubrió un fragmento de 360 pares de bases (pb) del genoma del VHD, desde las posiciones nucleotídicas 910 a 1270. Este fragmento incluía la región codificante del antígeno de la hepatitis delta (HDAg) que contiene el sitio de edición (nucleótido 1012 dentro del codón 196). En el segundo estudio, las regiones seleccionadas fueron la misma que en el primer estudio y la región de la ribozima, que consta de un fragmento de 198 pb, desde las posiciones nucleotídicas 684 a 881.

En ambas regiones, en primer lugar, el ARN del VHD extraído se desnaturaliza a 98°C durante 5 minutos y se transfirió inmediatamente a -80°C, por las razones explicadas en el apartado anterior.²⁰⁶ Después se realizó la reacción de RT utilizando la transcriptasa inversa Accuscript HiFi (Agilent Technologies, Santa Clara, Estados Unidos) de acuerdo con las instrucciones del fabricante. El ADN complementario (ADNc) resultante se amplifica en una primera reacción de PCR usando un cebador directo (*forward* (fw)) y un

cebador inverso (*reverse* (rv)). La PCR anidada (PCR M13) se realizó con los cebadores M13-fw y M13-rv, que incluían cada uno una secuencia adaptadora universal M13 diferente en el extremo 5'. Para el primer estudio, en el que la NGS se llevó a cabo en la plataforma Roche 454 GS Junior (Roche Diagnostics, Branford, Estados Unidos), en la última PCR (PCR MID) se utilizó un par específico de cebadores para cada muestra que contenía la secuencia adaptadora de NGS en 5', seguido de una secuencia no palindrómica "llave" (*Key sequence*), seguido de un identificador único para cada muestra (secuencias de identificador multiplex [MID]) y las mismas secuencias universales M13 que en los cebadores utilizados en la PCR anterior. Hay dos clases de adaptadores, A y B, que se diferencian en la secuencia de nucleótidos. La secuencia MID es un identificador único establecido por el fabricante, que se utiliza durante el análisis bioinformático para identificar la muestra a la que pertenece cada lectura de secuencia (*read*) obtenida.¹⁰⁴ Para el segundo estudio, en el que la NGS se llevó a cabo en la plataforma MiSeq (Illumina, San Diego, Estados Unidos), la PCR MID utilizó un par de cebadores para cada muestra que contenían la secuencia MID y las secuencias M13. Todos los pasos de PCR se realizaron utilizando la ADN polimerasa Pfu Ultra II de alta fidelidad (Agilent Technologies, Santa Clara, Estados Unidos). El límite de sensibilidad para la amplificación del HDAg, es de muestras con ARN-VHD superior a 1,00E+03; en cambio, para la ribozima es de muestras con ARN-VHD superior a 1,00E+02. Los cebadores utilizados para amplificar las secuencias del primer y segundo estudio se muestran en la Tabla 4, los componentes utilizados se muestran en la Tabla 5 y los protocolos que se han seguido se muestran en la Tabla 6.

Primer estudio – Amplificación de la región del HDAg			
PCR	Cebadores	Posiciones	Secuencia 5' - 3'
RT	RT rv	728-747	CGGTCCCCTCGGAATGTTG
Primer a PCR	1a fw	865-884	AGGTCGGACCGCGAGGAGGT
	1a rv	306-328	GCTGAAGGGGTCCTCTGGAGGTG
PCR M13	M13-fw	886-909	GTTGTA AACGACGGCCAGT GAGATGCCATGCCGACCCGAAGAG
	M13-rv	1272-1295	CACAGGAAACAGCTATGACC CGACGAAGGAAGGCCCTCGAGAAC
PCR MID	MID A	-	CGTATCGCCTCCCTCGCGCCATCAG-MID-GTTGTA AACGACGGCCAGT
	MID B	-	CTATGCGCCTTGCCAGCCCGTCAG-MID-CACAGGAAACAGCTATGACC
Segundo estudio – Amplificación de la región del HDAg			
PCR	Cebadores	Posiciones	Secuencia 5' - 3'
PCR MID	MID fw	-	MID-GTTGTA AACGACGGCCAGT
	MID rv	-	MID-CACAGGAAACAGCTATGACC
Segundo estudio – Amplificación de la región de la Ribozima			
PCR	Cebadores	Posiciones	Secuencia 5' - 3'
RT	RT rv	1435-1454	TGGCTGGGAAACATCAAAGG
Primer a PCR	1a fw	1435-1454	TGGCTGGGAAACATCAAAGG
	1a rv	308-326	CCTCCAGAGGACCCCTTCA
PCR M13	M13-fw	883-900	CACAGGAAACAGCTATGACC TCGGCATGGCATCTCCAC
	M13-rv	663-683	GTTGTA AACGACGGCCAGT CGCGTTCCATCCTTTCTTACC
PCR MID	MID fw	-	MID-GTTGTA AACGACGGCCAGT
	MID rv	-	MID-CACAGGAAACAGCTATGACC

Tabla 4. Cebadores utilizados en la transcripción inversa (RT) y en las PCRs para amplificar la región del HDAg y la región de la ribozima. En la región del HDAg, del primer al segundo estudio, sólo varían los cebadores de la PCR MID. Los nucleótidos coloreados en **rosa** corresponden a la cola M13, los coloreados en **azul** a la secuencia adaptadora de NGS A o B y los coloreados en **amarillo** a la *Key sequence* utilizada en la plataforma Roche 454 Roche 454 GS Junior.

	Componentes	Concentración	Volumen (µL)
Desnaturalización y choque térmico	RT rv	10 µM	2
	ARN-VHD extraído	-	10
RT	Agua estéril	-	1
	Ditiotreitol (DTT)	-	2
	Tampón de reacción	10 x	2
	dNTPs	10 mM	2
	Transcriptasa inversa Accuscript HiFi	-	1
	ARN-VHD extraído + RT rv	-	12
Primera PCR	Agua estéril	-	28,75
	Tampón de reacción	10 x	5
	dNTPs	10 mM	1,25
	1a fw	10 µM	2
	1a rv	10 µM	2
	ADN polimerasa Pfu Ultra II	-	1
	Producto de la RT	-	10
PCR M13	Agua estéril	-	16,875
	Tampón de reacción	10 x	2,5
	dNTPs	10 mM	0,625
	M13-fw	10 µM	1
	M13-rv	10 µM	1
	ADN polimerasa Pfu Ultra II	-	0,5
	Producto de la 1a PCR	-	2,5
PCR MID	Agua estéril	-	38,75
	Tampón de reacción	10 x	5
	dNTPs	10 mM	1,25
	M13-fw	20 µM	1
	M13-rv	20 µM	1
	ADN polimerasa Pfu Ultra II	-	1
	Producto de la PCR M13	-	2

Tabla 5. Componentes, concentración y volumen utilizados en la amplificación del HDAg y de la ribozima.

		Temperatura	Tiempo
RT	Transcripción inversa	42°C	60 minutos
	Inactivación	70°C	10 minutos
	Enfriamiento	20°C	∞
	Desnaturalización	95°C	1 minuto
Primera PCR	Amplificación (40 ciclos)	94°C	20 segundos
		54°C	20 segundos
		72°C	45 segundos
	Extensión	72°C	3 minutos
	Enfriamiento	4°C	∞
	Desnaturalización	95°C	2 minutos
PCR M13	Amplificación (35 ciclos)	94°C	20 segundos
		63°C (para el HDAg) 60°C (para la ribozima)	20 segundos
		72°C	30 segundos
		Extensión	72°C
	Enfriamiento	4°C	∞
	Desnaturalización	95°C	2 minutos
PCR MID	Amplificación (25 ciclos)	94°C	20 segundos
		60°C	20 segundos
		72°C	30 segundos (Roche 454 GS Junior) / 45 segundos (MiSeq)
		Extensión	72°C
	Enfriamiento	4°C	∞
	Desnaturalización	95°C	2 minutos

Tabla 6. Tiempos y temperaturas del protocolo de la transcripción inversa (RT) y de los diversos pasos de PCR para amplificar la región del HDAg y la región de la ribozima. Todos los tiempos y temperaturas son los mismos en los diferentes programas de PCR para las dos regiones excepto en la PCR M13, en el segundo paso de amplificación y en la PCR MID en el tercer paso de amplificación.

4.6 Comprobación de la amplificación del ARN del VHD

Los productos obtenidos de las PCR de amplificación se visualizaron como bandas individuales en gel de electroforesis de agarosa al 1,5% diluida en el tampón 1x Tris-Borato-EDTA (TBE) (Thermo Fisher Scientific, Geel, Bélgica) para el primer estudio y el tampón 1x Tris-Acetato-EDTA (TAE) (Corning Mediatech Inc., Manassas, Estados Unidos) para el segundo estudio. Esta diferencia en el tampón se debe al método de NGS que se utilizará después. La electroforesis es un método de separación de biomoléculas de acuerdo a su tamaño que permite una observación directa de cada fragmento separado directamente en el gel. Cuando una corriente eléctrica se aplica sobre el gel de agarosa, el ADN tiene una carga negativa y migra hacia el ánodo (polo positivo). Para visualizar las bandas de ADN se utilizó el reactivo *SybrSafe DNA Gel Stain* (Thermo Fisher Scientific, Carlsbad, Estados Unidos), el cual tiñe el ADN, haciéndolo visible mediante la emisión de luz ultravioleta.²⁰⁸ Además, esta técnica permite comprobar la longitud de los fragmentos amplificados añadiendo un marcador de peso molecular. En la separación electroforética, este marcador, da como resultado un patrón regular de bandas donde se conoce el peso molecular de cada una de ellas. Depende de la longitud esperada del fragmento se ha utilizado el marcador *DNA Molecular Weight Marker XIV* (100–1500 pb) o el marcador *DNA Molecular Weight Marker II* (0.12–23.1 kpb) (Roche Diagnostics, Mannheim, Alemania). Para visualizar los productos de PCR obtenidos se cargaron 10 µL en el gel de agarosa si solo se quería detectar si había o no producto de PCR amplificado. En cambio, si después se realiza el corte de banda y su purificación, se carga todo el producto de PCR.

4.7 Purificación del producto de PCR

Una vez separadas las bandas correspondientes al producto de PCR en el gel de agarosa, se cortaron y se purificaron usando el *QIAquick Gel Extraction Kit* (QIAGEN®, Hilden, Alemania) siguiendo las instrucciones del fabricante. Este kit de extracción permite la eliminación de nucleótidos, enzimas, sales, agarosa y otras impurezas de las muestras, asegurando una recuperación del ADN de hasta el 80%. El ADN se purificó utilizando un procedimiento simple de unión, lavado y elución. Las bandas de gel se disuelven en un tampón que contiene un indicador de pH, lo que permite una fácil determinación del pH óptimo para la unión del ADN. La mezcla se aplica a la columna de centrifugación *QIAquick* y los ácidos nucleicos se adsorben a la membrana de sílice, en

las condiciones de alto contenido de sal proporcionadas por el tampón. Las impurezas se eliminan y el ADN puro se eluye en 50 μL de tampón 10 mM Tris-Cl, pH 8,5.²⁰⁹

4.8 Comprobación de la calidad del producto de PCR purificado

Una vez purificados los productos de PCR, la calidad del ADN, es decir la pureza en la que se han obtenido los fragmentos de interés en las librerías de NGS, se analizó utilizando el sistema automatizado Agilent 2200 TapeStation con el kit D1000 *ScreenTape* (Agilent Technologies, Waldbronn, Alemania). El sistema Agilent 2200 TapeStation es una plataforma automatizada que utiliza *ScreenTapes* con 16 carriles independientes. A partir de 1 μL de muestra, se mezcla con un tinte intercalante que recubre el ADN y se realiza una electroforesis automatizada. Los resultados se muestran como una imagen de gel (gel de electroforesis virtual), un electroferograma y una tabla con la estimación de la concentración de ADN ($\text{ng}/\mu\text{L}$) en cada pico del electroferograma. El tiempo necesario es de 1 o 2 minutos/muestra.²¹⁰

Este análisis de la calidad es más sensible que la electroforesis en gel de agarosa e indicará si existen fragmentos diferentes a nuestro fragmento de interés. Es importante que no haya fragmentos cortos, ya que, pueden causar interferencias importantes en la NGS (reducen significativamente las lecturas de la secuencia de interés) y en el clonaje (los fragmentos cortos se insertan con más eficiencia en el plásmido).

4.9 Cuantificación del producto de PCR purificado

Para garantizar que cada amplicón se representa adecuadamente en el análisis de NGS, se cuantificó el ADN purificado de cada muestra de manera automatizada mediante el robot Freedom EVO® (Tecan, Mannedorf, Suiza) acoplado al fluorímetro Infinite 200 Pro (Tecan, Mannedorf, Suiza), utilizando el reactivo *Quant-iT™ PicoGreen® dsDNA Assay Kit* (Thermo Fisher Scientific, Camarillo, Estados Unidos) según las instrucciones del fabricante. Este reactivo es una tinción de ácidos nucleicos fluorescente ultrasensible para cuantificar específicamente el ADN bicatenario (ADNs) en solución. Tiene un intervalo de linealidad desde 50 pg/mL hasta 2 $\mu\text{g}/\text{mL}$ y un límite de detección de 25 pg/mL . Se utiliza un espectro fluorimétrico estándar y longitudes de onda de excitación y emisión de fluoresceína.²¹¹

Una vez se conoce la concentración de cada amplicón, se realiza una normalización, es decir, se ajustan todos los amplicones a la misma concentración. Para las muestras que se van a secuenciar por la plataforma Roche 454 GS Junior esta concentración será de $1,00E+09$ moléculas/ μ L y para las muestras que se vayan a secuenciar en la plataforma MiSeq esta concentración será de $1,00E+10$ moléculas/ μ L. Este paso se realiza diluyendo los amplicones con tampón 10 mM Tris-Cl, pH 8,5. Una vez tenemos todos los amplicones a la misma concentración, se mezclan todos y se crea un *pool*. De esta manera cada amplicón estará representado en cantidades equimoleculares en la librería que será analizada por NGS.

Otro tipo de cuantificación, utilizado para cuantificar la concentración de los productos de PCR antes de la clonación y para cuantificar el producto purificado obtenido después de la clonación, es mediante el NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, Estados Unidos). Este espectrofotómetro cuantifica a partir de 1 μ L de muestra con alta precisión y reproducibilidad. Tiene un intervalo de linealidad desde 2 hasta 3700 ng/ μ L. La muestra se pipetea en el extremo de un cable de fibra óptica (la fibra receptora). Luego se pone en contacto un segundo cable de fibra óptica (la fibra fuente) con la muestra líquida, haciendo que el líquido cierre el espacio entre los extremos de la fibra óptica. Una lámpara de flash de xenón pulsado proporciona la fuente de luz y un espectrofotómetro que utiliza un dispositivo de carga acoplada (CCD) lineal se utiliza para analizar la luz después de pasar a través de la muestra.²¹²

4.10 NGS del producto de PCR purificado

En el primer estudio, el *pool* fue secuenciado por NGS en el sistema Roche 454 GS Junior,^{213,214} mientras que en el segundo estudio, el *pool* fue secuenciado por NGS en la plataforma MiSeq.²¹⁵ Ambas plataformas usan sistemas de NGS diferentes que se resumen a continuación.

4.10.1 NGS mediante la plataforma Roche 454 GS Junior:

1. Amplificación clonal en la PCR en emulsión (emPCR).

Antes de la NGS se necesita la amplificación clonal de cada una de las cadenas de los amplicones incluidos en cada librería mediante una emPCR (Figura 14). En este proceso, los amplicones son desnaturalizados para obtener ADN de cadena simple (ADNss). Estas moléculas se mezclan en una emulsión de “agua en aceite” con unas esferas microscópicas llamadas *DNA Capture Beads*.²¹⁶ Esta mezcla tiene lugar en condiciones que favorecen que en una gota de emulsión entre solamente una esfera y una sola molécula de ADNss. Esta molécula se hibrida con las esferas recubiertas de una secuencia complementaria al adaptador (A o B) mencionado en el apartado 4.5 *Amplificación del ARN del VHD para obtener librerías de amplicones para analizar por secuenciación masiva (NGS)*. Una vez que las moléculas de ADNss están inmovilizadas en las esferas, este complejo es emulsionado con los reactivos necesarios para la amplificación [desoxinucleótidos trifosfato (dNTPs), cebadores y la ADN polimerasa]. De esta manera se forman microreactores donde una única cadena de cada amplicón será amplificada millones de veces y las copias cubrirán toda la esfera.

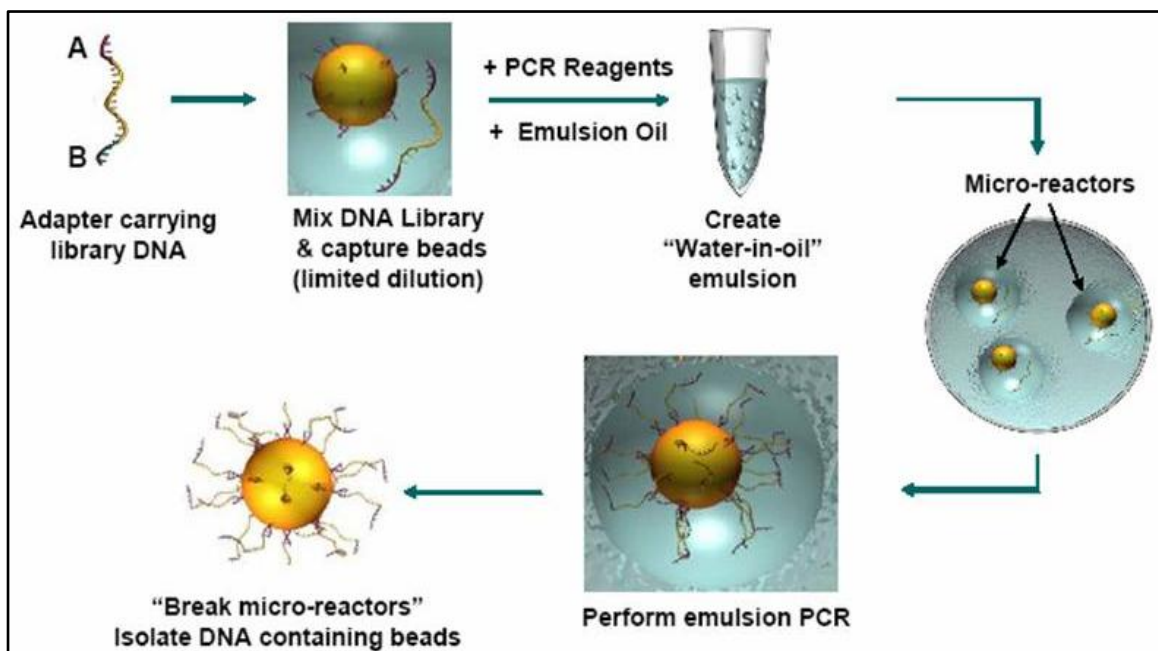


Figura 14. Esquema de la amplificación clonal en la emPCR. Figura modificada a partir de figuras proporcionadas por el fabricante (Roche, Branford, Estados Unidos).²¹⁶

2. Enriquecimiento:

Después de la amplificación por emPCR, se rompe la emulsión y se cuentan las esferas recuperadas, siendo la recuperación óptima entre el 65 y el 85%. Se obtienen dos tipos de esferas, unas con ADN amplificado y otras sin producto amplificado. A continuación, se lleva a cabo el proceso de enriquecimiento, que consiste en capturar esas esferas con ADN y eliminar las que no tienen producto amplificado (Figura 15). Para ello, las esferas recubiertas de ADN se separan de las que no tienen producto amplificado por unión a unas esferas paramagnéticas llamadas *enrichment beads*. La unión de las *enrichment beads* con las esferas recubiertas de ADN tiene lugar mediante un cebador conjugado con biotina (*enrichment primer*). Los amplicones de ADN que recubren las esferas, se desnaturalizan y se hibridan con el *enrichment primer*, que es complementario al adaptador (A o B) del extremo 3' del ADN inmovilizado en las esferas. Además, las *enrichment beads* están recubiertas de estreptavidina que se une a la biotina de los *enrichment primers*. El complejo formado por esferas con ADN amplificado + *enrichment primer* + *enrichment beads* se introduce en un imán donde es retenido, mientras que el sobrenadante con esferas nulas se elimina. Una vez separadas las esferas recubiertas de ADN, se desnaturaliza el ADN que recubre las esferas recuperadas para obtener una librería de ADNss al que se hibrida otro cebador que actuará en la reacción de pirosecuenciación. A continuación, se valora la calidad del procedimiento de enriquecimiento por conteo de las esferas recuperadas al final de este proceso. El enriquecimiento óptimo está entre el 5 y el 20% de los 10 millones de esferas originales añadidas a la emulsión (entre 500.000 y 2.000.000 de esferas). Un enriquecimiento por debajo del 5% daría pocas lecturas de secuencia (*reads*) en la carrera de secuenciación, en cambio, por encima del 20% indicaría que hay gotas de emulsión con más de una molécula de ADNss, lo que supondría una lectura de mala calidad.

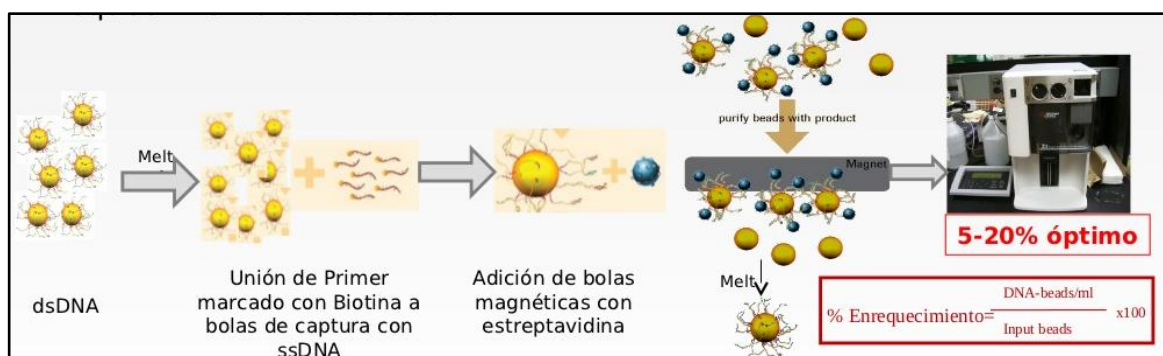


Figura 15. Esquema del enriquecimiento. Figura modificada a partir de figuras proporcionadas por el fabricante (Roche, Branford, Estados Unidos).

3. Pirosecuenciación:

Las esferas enriquecidas se depositan en una placa de fibra óptica PTP (*PicoTiterPlate*), de manera que solo entre una esfera por pocillo.²¹⁶ Junto con estas esferas se añaden otras esferas: unas con las enzimas necesarias para la reacción de pirosecuenciación (*enzyme beads*) y otras para mantener el conjunto de todos los componentes del sistema de pirosecuenciación dentro de los pocillos (*packing beads*). El depósito de las esferas en los pocillos de la PTP se hace por capas a través de varios ciclos de carga y centrifugación (Figura 16). Además, el subsistema de fluídica de la plataforma va liberando los reactivos necesarios para la pirosecuenciación, como son los tampones y los dNTPs, en flujos sobre la PTP.

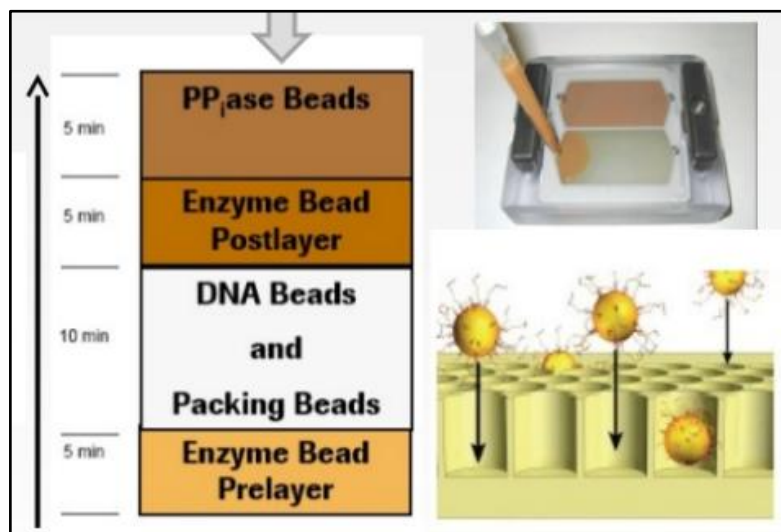


Figura 16. Carga de las esferas en capas dentro de los pocillos de la PTP y disposición de las capas dentro de cada uno de los pocillos. Figura modificada a partir de figuras proporcionadas por el fabricante (Roche, Branford, Estados Unidos).²¹⁶

Una vez cargada, la PTP se deposita dentro de la celda de flujo o *flowcell* del secuenciador Roche 454 GS Junior y empieza el proceso de NGS (Figura 17) que se realiza en los siguientes pasos:²¹⁶

a) El ADNss, unido a las esferas recuperadas después del enriquecimiento e hibridado con el cebador de secuenciación, se incuba con las enzimas ADN polimerasa, ATP sulfurilasa, luciferasa y apirasa, más los sustratos adenosina-5'-fosfosulfato (APS) y luciferina, en los pocillos de la placa PTP.

b) Adición en ciclos, de uno en uno, los cuatro dNTPs, mientras la ADN polimerasa cataliza la incorporación del dNTP en la cadena molde unida a las esferas

con ADN, si éste es complementario. Si hay incorporación, se libera una cantidad de pirofosfato inorgánico (PPi) equivalente a la cantidad de dNTPs incorporados a la nueva cadena.

c) Este PPi iniciará una reacción enzimática acoplada que dará lugar a la producción de quimioluminiscencia: la ATP sulfurilasa convierte cuantitativamente el PPi en adenosina trifosfato (ATP) en presencia de APS. El ATP generado, aporta la energía necesaria que permite la oxidación de la luciferina en oxiluciferina por acción de la luciferasa. Este proceso genera luz visible en cantidades proporcionales a la cantidad de ATP presente, por tanto proporcional al número de nucleótidos incorporados en cada ciclo.

d) La luz emitida se transmite a una cámara de alta resolución que consiste en un CCD, que la registra. En cada ciclo de un dNTP se coge una imagen de los pocillos, se analiza el ruido de fondo y se filtra de acuerdo con los criterios de calidad internos del *software* del instrumento.

e) La aspirasa degrada los dNTPs que no han sido incorporados y el ATP que no se utiliza.

f) Se libera un nuevo dNTP que puede ser añadido para iniciar un nuevo ciclo.

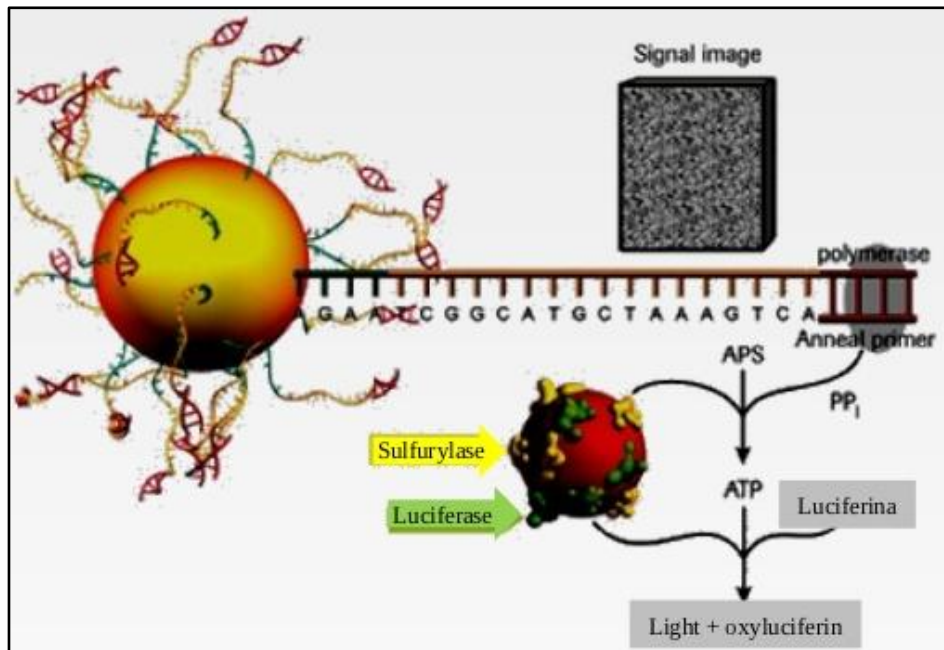


Figura 17. Esquema del proceso de pirosecuenciación. Figura modificada a partir de figuras proporcionadas por el fabricante (Roche, Branford, Estados Unidos).

4. Análisis de la imagen

El primer paso es identificar las imágenes generadas en cada pocillo de la PTP. En este punto, son de gran importancia los cuatro nucleótidos de los cebadores de la PCR MID que forman la *Key sequence* (secuencia TCAG, en color amarillo, en la Tabla 4). Esta secuencia permite reconocer cada pocillo de manera individual y detectar la presencia de una esfera con ADN. Además, esta secuencia también permite calibrar la intensidad de la señal lumínica de cada base, que indica directamente el número de nucleótidos incorporados a cada pocillo a lo largo de cada ciclo. Los parámetros de cada imagen, como la intensidad de la señal, el ruido de fondo, etc. se analizan por el *software* del instrumento (*Amplicon Variant Analyzer, AVA, Roche*). Este *software* aplica una serie de pasos de normalización, corrección y filtraje de la señal obtenida y transforma estas señales en un gráfico de flujos (*flowgram*) (Figura 18). Esto permite convertir estos datos en *reads*, un proceso conocido como asignación de bases o *base calling*.

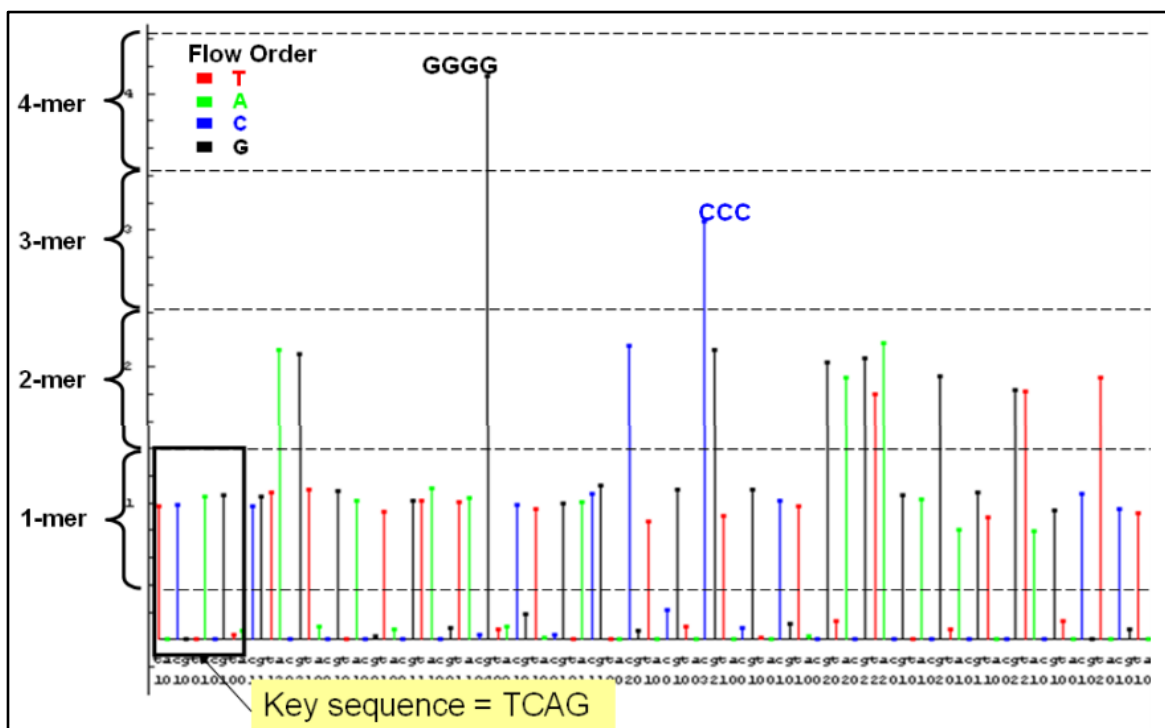


Figura 18. Ejemplo de un gráfico de flujos generado por el *software* de la plataforma. Figura modificada a partir de figuras proporcionadas por el fabricante (Roche, Branford, Estados Unidos).

4.10.2 NGS mediante la plataforma MiSeq:

1. Reparación de extremos

En primer lugar, se deben transformar los extremos cohesivos de los amplicones, mezclados en cantidades equimoleculares en los *pools*, en extremos romos. Este paso se ha realizado con el *Kapa Hyper Prep Kit* (Roche Sequencing, Pleasanton, Estados Unidos) utilizando el *buffer End Repair & A-tailing* y la *End Repair & A-tailing Enzyme Mix*. El *pool* se mezcla con estos reactivos y se incuba a 20°C durante 30 minutos y a 65°C durante 30 minutos en un termociclador. Además, en este paso se fosforila el extremo 5' y se añade un solo nucleótido de Adenina (A) a los extremos 3' (Figura 19).

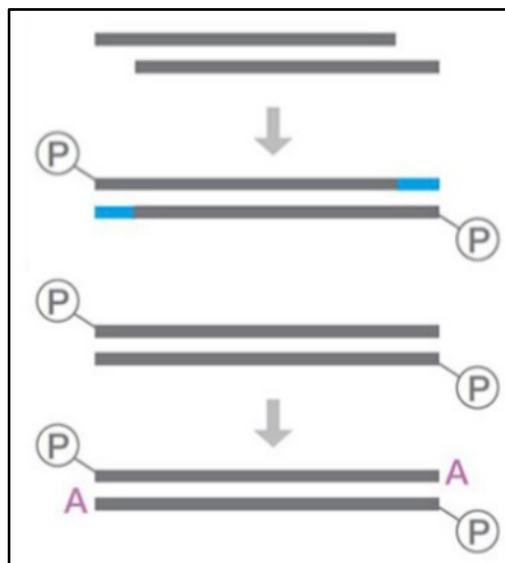


Figura 19. Reparación de extremos: se transforman los extremos cohesivos a extremos romos, se fosforila el extremo 5' y se añade un nucleótido A al extremo 3'. Figura modificada a partir de figuras proporcionadas por el fabricante (Illumina, San Diego, Estados Unidos).

2. Ligación de los adaptadores

En este paso se procede a la ligación de los adaptadores a los extremos de los amplicones del *pool* (Figura 20). Esta ligación se ha realizado con *Nimblegen SeqCap Adapter Kit* (Roche Sequencing, Pleasanton, Estados Unidos) gracias al nucleótido A añadido en el paso anterior, que se unirá a un nucleótido de Timina (T) presente en el adaptador. P5 y P7 permitirán que los amplicones se unan a la celda de flujo, ya que, son complementarios a los oligonucleótidos de su superficie. P5 se unirá al extremo 5' y P7 se unirá al extremo 3'. Los índices permitirán etiquetar cada *pool* de amplicones, ya que, en un mismo experimento se puede secuenciar más de un *pool* a la vez (los amplicones de cada muestra individual dentro de los *pools* se seguirán identificando por las mismas

secuencias MID que con el sistema Roche 454 GS Junior). Rd1SP y Rd2SP son la diana de los cebadores utilizados para iniciar la secuenciación. En la reacción de ligación se necesita el tampón de ligación, la enzima ligasa y el agua, que se unirán a la mezcla creada en el apartado anterior, además del adaptador elegido. Esta mezcla se incubó en el termociclador a 20°C durante 15 minutos.

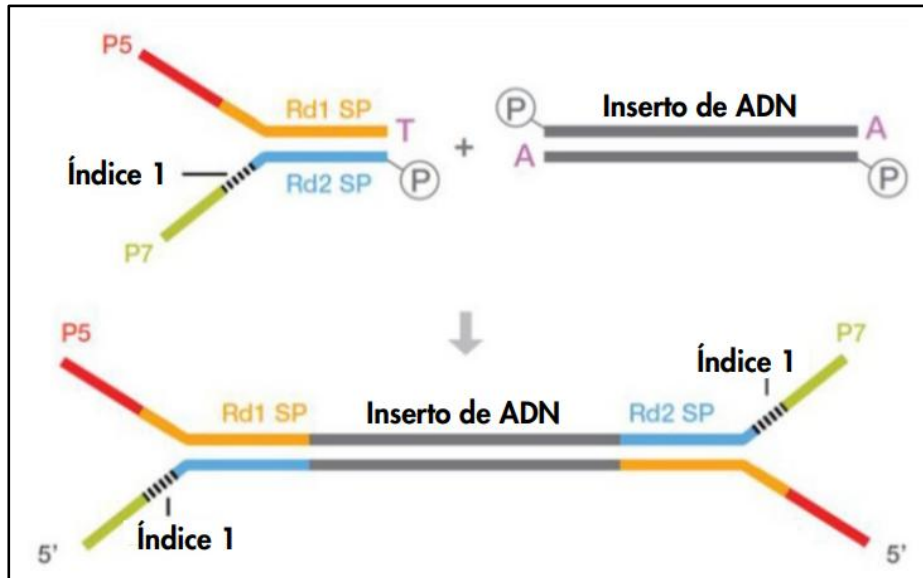


Figura 20. Ligación de los adaptadores. Figura modificada a partir de figuras proporcionadas por el fabricante (Illumina, San Diego, Estados Unidos).

3. Lavado post-ligación

En este paso se utilizan las *KAPA Pure Beads* (Roche Sequencing, Pleasanton, Estados Unidos) que consisten en una suspensión de esferas paramagnéticas en un tampón optimizado. Este tampón incluye PEG/NaCl, un agente de precipitación diseñado para conducir las moléculas de ADN a las esferas para su unión. La relación volumétrica entre las *KAPA Pure Beads* y la muestra es el factor crítico para determinar la distribución del tamaño de los fragmentos de ADN retenidos por las esferas. En este caso se utilizó una *ratio* de 0,8; 1 µL de DNA por 0,8 µL de *KAPA Pure Beads*. Este lavado se ha realizado siguiendo las instrucciones del fabricante y permite eliminar adaptadores no ligados, dímeros de adaptadores, cebadores, dímeros de cebadores, nucleótidos, sales y enzimas.²¹⁷

4. Amplificación intermedia de las librerías de amplicones de los *pools*

Este paso intermedio no es obligatorio, pero sí que es recomendable para *pools* con librerías de amplicones presentes a concentraciones bajas, o si se analiza un virus en que su amplificación tiene ciertas complicaciones como es el caso del VHD.

Para esta amplificación se utilizan 20 µL del producto obtenido en el apartado anterior y el *KAPA HiFi HotStart ReadyMix PCR Kit* (Roche Sequencing, Pleasanton, Estados Unidos), que contiene la *KAPA HiFi HotStart Ready Mix* (2x) y la *KAPA Library Amplification Primer Mix* (10x). La ADN polimerasa *KAPA HiFi HotStart* tiene la capacidad de amplificar secuencias largas, secuencias ricas en GC y AT y actividad de corrección de errores (1 error por cada 3,6E+06 nucleótidos incorporados).²¹⁸

El protocolo que hay que seguir se resume en la Tabla 7. La baja tasa de errores de la polimerasa *KAPA HiFi HotStart* junto con el bajo número de ciclos de la PCR intermedia minimiza la introducción de artefactos de amplificación en las librerías de amplicones de los *pools*.

	Temperatura	Tiempo	Ciclos
Tapa del termociclador a 85°C			
Desnaturalización inicial	98°C	45 segundos	1
Desnaturalización	98°C	15 segundos	5
<i>Annealing</i>	60°C	30 segundos	
Extensión	72°C	30 segundos	
Extensión final	72°C	1 minuto	1
Enfriamiento	4°C	∞	1

Tabla 7. Tiempos, temperatura y ciclos del protocolo de la PCR intermedia.

5. Lavado post-amplificación

Se realiza lo mismo que en el apartado 4. *Lavado post-ligación*.

6. Cuantificación de los *pools* por qPCR

El siguiente paso es cuantificar los *pools* en el instrumento *LightCycler 480* (Roche Diagnostics, Indianapolis, Estados Unidos) mediante una qPCR a tiempo real utilizando el *KAPA Library Quantification Kit* (Roche Sequencing, Pleasanton, Estados Unidos) (Tabla 8). A cada pocillo de la placa de PCR se le añade 6 µL de la *KAPA SYBR FAST qPCR Master Mix* (2x), que incluye el *Primer Premix* (10x) (que se dirigen a los adaptadores P5 y P7 nombrados anteriormente) y la ADN polimerasa *KAPA SYBR FAST*,

2 μ L de agua estéril y 2 μ L de cada muestra a cuantificar.²¹⁹ En la qPCR se cuantifican tres réplicas de los estándares de ADN prediluidos (también incluidos en el *KAPA Library Quantification Kit*), tres réplicas de los *pools* diluidos (1:100 y 1:10.000) con tampón 10 mM Tris-Cl, pH 8,5, tres réplicas de un control negativo (agua estéril) y tres réplicas de un control interno (PhiX). El PhiX, se deriva del genoma pequeño y bien caracterizado del bacteriófago PhiX. Es una librería comercial de Illumina concentrada (10 nM en 10 μ L) de fragmentos de ADN que tienen un tamaño promedio de 500 pb y contienen una composición de bases equilibrada, aproximadamente el 45% de GC y el 55% de AT.²²⁰ Las réplicas se utilizan para corroborar la veracidad de la cuantificación.

En la qPCR se obtienen valores de cuantificación del ciclo (Cq) que es el número del ciclo de PCR en el que la curva de fluorescencia de la muestra se cruza con la línea de umbral. Este valor indica cuántos ciclos tardó en detectar una señal real en las muestras. Los valores de Cq son inversos a la cantidad de ADN presente en la muestra. Para poder cuantificar la concentración de cada muestra incluida en la qPCR, el valor promedio de Cq para cada estándar de ADN se representa gráficamente frente a la concentración en \log_{10} en pM, para generar una curva estándar. El rango de concentración de la curva estándar es de 20 a 0,0002 pM. Las concentraciones de cada muestra se calculan contra la curva estándar, utilizando la cuantificación absoluta.

LightCycler 480				
Programa	Temperatura	Tiempo	Modo de adquisición de fluorescencia	Modo de análisis
Desnaturalización	95°C	10 minutos	Ninguno	Ninguno
Amplificación (35 ciclos)	95°C	30 segundos	Ninguno	Cuantificación
	60°C	45 segundos	Único	
Curva de Melting	65°C	30 segundos	Ninguna	Curvas de Melting
	95°C		Continuo	
Enfriamiento	40°C	30 segundos	Ninguna	Ninguno

Tabla 8. Tiempos y temperaturas del protocolo de cuantificación de los *pools*.

7. Pool de pools

Una vez obtenida la concentración a la que se encuentran las librerías de amplicones en cada *pool*, éstos se diluyen a una concentración de 4 nM con tampón 10 mM Tris-Cl, pH 8,5. Este proceso se lleva a cabo con todos los *pools* que se quieran cargar a la vez en un mismo experimento de secuenciación. Con todas las librerías ya normalizadas a 4 nM se forma un *pool de pools*. En principio se mezclan todos los *pools* representados de igual manera en el *pool de pools*. Sin embargo, según los resultados obtenidos en experimentos anteriores se pueden variar los volúmenes en los que se mezclan los diferentes *pools*.

8. Cuantificación del pool de pools por qPCR

En este paso se cuantifica la concentración de este nuevo *pool de pools* de la misma manera que se ha explicado en el apartado 6. *Cuantificación de los pools por qPCR*. En este caso, se diluirá con tampón 10 mM Tris-Cl, pH 8,5 cada *pool* por separado más el *pool de pools* a 1:400 y 1:4.000. Y, se vuelve a cuantificar cada *pool* por separado para asegurar que la concentración de las librerías de amplicones en el *pool de pools* son las esperadas.

9. Preparación del pool de pools para cargar en la plataforma MiSeq

Una vez cuantificado el *pool de pools* se ajusta su concentración a 4nM (o a 2nM si por problemas en el pipeteo o en la cuantificación esta concentración ha disminuido) con tampón 10 mM Tris-Cl, pH 8,5, dependiendo de la concentración que hayamos obtenido después de la cuantificación. En este caso se ha ajustado a 4 nM. Seguidamente, se desnaturaliza mediante NaOH 0,2 N y se neutraliza el exceso de NaOH añadiendo solución Tris-HCl 200 mM pH 7-8. También se desnaturaliza de la misma manera el PhiX. A continuación, se mezcla el *pool de pools* desnaturalizado, el PhiX desnaturalizado y se añade el tampón HT1 del *MiSeq Reagent Kit v3* (Illumina, San Diego, Estados Unidos).

El PhiX se utiliza para aumentar artificialmente la diversidad de las secuencias en cada experimento de secuenciación. Esto es necesario para no saturar la señal obtenida en los diferentes ciclos de secuenciación debido a la secuenciación de librerías de baja diversidad, es decir, con amplicones de secuencia muy parecida.²²⁰

La mezcla del *pool de pools* con el PhiX se conocerá a partir de ahora como *pool de carga* y para asegurar la desnaturalización se incuba durante 2 minutos a 95°C y se deja en hielo hasta cargarlo en la plataforma MiSeq.

10. Preparación de la plataforma MiSeq

Una vez delante del instrumento, en la interfaz del *software*, se selecciona *Sequence* (Secuenciar) para iniciar los pasos de configuración del experimento. Seguidamente, la celda de flujo se lava con agua Mili-Q, se seca bien y se carga en el instrumento. Después, se carga la botella del tampón de lavado (PR2) del *MiSeq Reagent Kit v3* (Illumina, San Diego, Estados Unidos) y se desechan los residuos de la botella de residuos si los hay.²²¹

11. Carga del *pool* de carga a la plataforma MiSeq

El siguiente paso es cargar el *pool* de carga en la plataforma MiSeq. Hay que tener en cuenta que el cartucho de reactivos debe dejarse descongelar antes de utilizarse. Se añaden los 600 μ L del *pool* de carga al cartucho de reactivos en el lugar que se indica en el propio cartucho. Por último, se introduce el cartucho de reactivos en el instrumento y se siguen los pasos que éste indica hasta iniciar la secuenciación.

12. Generación de *clusters* y secuenciación por síntesis

El funcionamiento de la tecnología de secuenciación por síntesis (SBS) de la plataforma MiSeq se basa en la generación de *clusters* a partir del *pool* de carga. Los *clusters* son agrupaciones de secuencias idénticas, amplificadas clonalmente.

La generación de *clusters* sigue los siguientes pasos (información proporcionada por el fabricante (Illumina, San Diego, Estados Unidos)):²²²

1. Hibridación de los adaptadores ligados al extremo de los amplicones, en forma de ADNss del *pool* de carga, con oligonucleótidos complementarios presentes en la celda de flujo, permitiendo su extensión a través de una polimerasa (Figura 21, 1).
2. Obtención de ADNds, que se desnaturaliza y después se elimina la hebra original.
3. Unión covalente de la hebra recién sintetizada con la superficie de la celda de flujo, esta hebra se voltea y forma un puente al hibridarse con un oligonucleótido adyacente, complementario a su extremo libre.
4. Extensión del cebador hibridado gracias a las polimerasas, obteniendo puentes bicatenarios (Figura 21, 2).

5. Desnaturalización de los puentes obteniendo dos hebras monocatenarias complementarias, unidas covalentemente a los dos respectivos oligonucleótidos (Figura 21, 3).
6. Las hebras vuelven a doblarse e hibridan con cebadores adyacentes, repitiendo este ciclo de amplificación, formando múltiples puentes que amplifican clonalmente la secuencia de la hebra original (Figura 21, 4).
7. Desnaturalización de los puentes y escisión de las cadenas rv, quedando solamente cadenas fw, a las que se les bloquea el extremo 3' libre para evitar el cebado indeseado del ADN.
8. Hibridación del cebador de secuenciación con la secuencia adaptadora y secuenciación de las cadenas fw (Figura 21, 5 y 6) mediante secuenciación por síntesis, tal como se explicará más adelante.
9. Las cadenas nuevas producidas durante la secuenciación por síntesis se eliminan y los extremos 3' de las cadenas fw y de los oligonucleótidos adyacentes se desbloquean. Entonces se procederá a la secuenciación de sus respectivas cadenas complementarias, las cadenas rv, volviendo a empezar las etapas de secuenciación desde el punto 3. Este proceso se conoce como *Paired-end sequencing*.

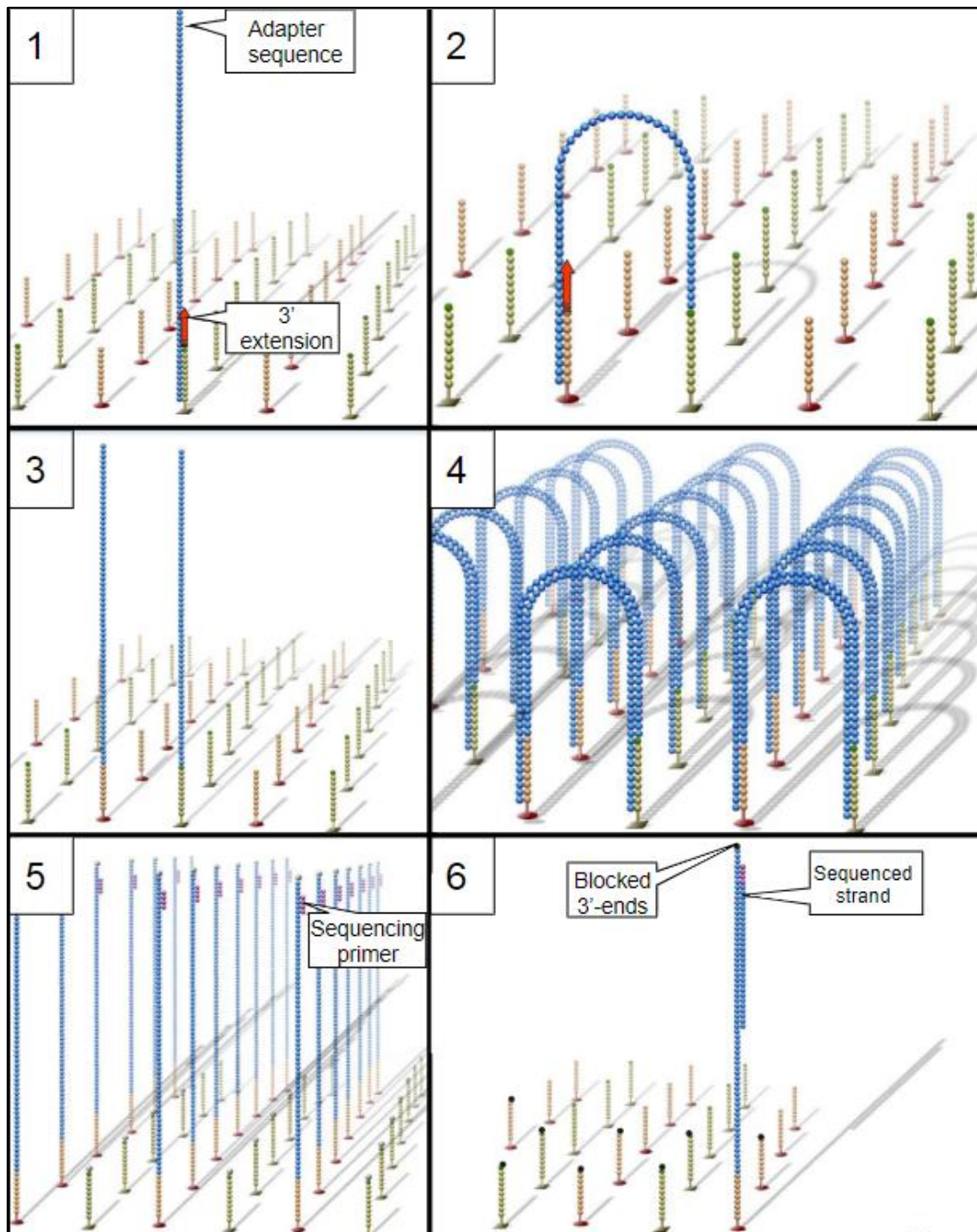


Figura 21. Esquema de la generación de clusters. Figura modificada a partir de figuras proporcionadas por el fabricante (Illumina, San Diego, Estados Unidos).²²²

La secuenciación de los *clusters* se lleva a cabo gracias a un proceso de secuenciación por síntesis (Figura 22). En la reacción, los 4 nucleótidos están unidos a fluorocromos de colores distintos para su identificación. Cuando un nucleótido se incorpora, se detecta el fluorocromo a partir de la longitud de onda de la fluorescencia que emite al ser excitado después de su adición. Esto se repite por cada posición nucleotídica. Los nucleótidos que

no se hayan unido serán lavados para continuar con el ciclo. A continuación, se elimina el fluorocromo de los nucleótidos (liberando su extremo 3' para que se una el siguiente nucleótido) y se comienza el ciclo desde el principio (Información proporcionada por el fabricante (Illumina, San Diego, Estados Unidos)).²²²

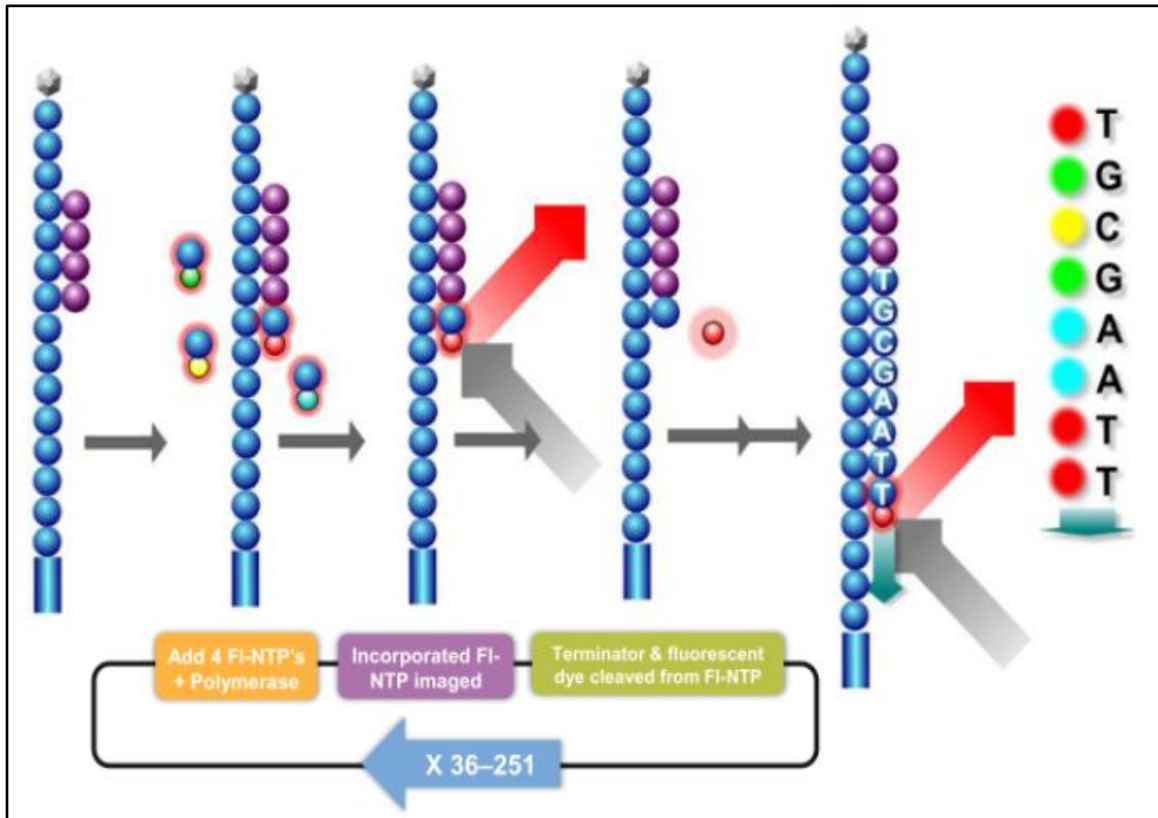


Figura 22. Esquema de la secuenciación por síntesis. Figura modificada a partir de figuras proporcionadas por el fabricante (Illumina, San Diego, Estados Unidos).²²²

4.11 Análisis de los datos de NGS

Todos los pasos de filtrado descritos a continuación, tanto para la plataforma Roche 454 GS Junior como para la plataforma MiSeq, se realizaron con scripts desarrollados internamente en R.^{197,199,223-225}

4.11.1 Análisis de los datos obtenidos mediante el sistema Roche 454 GS Junior

Brevemente, los pasos que se siguieron fueron los siguientes:

1. Obtención del archivo fasta proporcionado por la plataforma Roche 454 GS Junior.
2. Los *reads* del archivo fasta se dividieron en diferentes archivos de acuerdo con la identificación de las secuencias MID, el cebador M13 (fw o rv) y los cebadores específicos de cada amplicón. Se toleró 1 desajuste (*mismatch*) en la identificación MID, 4 *mismatches* en las secuencias M13 fw y rv y 3 *mismatches* en la identificación específica del cebador. Finalmente se eliminaron todas estas secuencias incluidas en los cebadores.
3. Se excluyeron los *reads* que no cubrieron completamente el amplicón y aquellos que mostraron más de 2 indeterminaciones, más de 3 gaps o más de 99 diferencias con respecto a la secuencia de referencia. Las otras indeterminaciones y gaps se repararon según el haplotipo dominante en cada muestra. Finalmente todos los *reads* se colapsaron en haplotipos, secuencias únicas que cubren todo el amplicón y agrupan secuencias iguales entre sí en el conjunto de secuencias.
4. Sólo se aceptaron los haplotipos comunes en la cadena fw y rv presentes en abundancias del 0,1% o superiores. La frecuencia final para cada haplotipo es la suma de los *reads* colapsados observados en cada cadena.
5. Después del primer filtraje, se excluyeron los haplotipos con abundancias inferiores al 0,25%. Este nivel de abundancia se consideró el dintel a partir del que no se podían distinguir con fiabilidad los haplotipos realmente presentes en la quasiespecies de los generados por la artefactuación del procedimiento de creación de librerías o secuenciación masiva.¹⁹⁷
6. Los haplotipos resultantes del paso anterior se tradujeron a secuencias de aminoácidos. Los que tenían la misma secuencia se colapsaron.

4.11.2 Análisis de los datos obtenidos mediante el sistema MiSeq

Brevemente, los pasos que se siguieron fueron los siguientes:

1. Se realiza un control de calidad de los archivos *fastq* obtenidos, analizando los perfiles de calidad por cada posición de las secuencias (eliminando *reads* con indeterminaciones), la longitud de los *reads* y los parámetros generales del instrumento relacionados con la calidad.

2. En la *Paired-end sequencing* se utiliza *FLASh* para solapar los *reads* obtenidos al secuenciar las cadenas en ambos sentidos, R1 y R2, imponiendo un mínimo de 20 pb solapadas con un máximo del 10% de *mismatches* (rendimiento del 60 al 80% para amplicones de 450 a 500 pb).²²⁶
3. Posteriormente se realiza un filtrado que elimina los *reads* con un 5% o más de bases con calidad inferior a Q30, *Phred score* de 30, lo que corresponde a una precisión estimada del 99,9% (rendimiento del 75 al 85%).²²⁷
4. La demultiplexación se realiza mediante la identificación de secuencias de oligonucleótidos, en ambos extremos, dentro de las ventanas de las posiciones esperadas en los *reads* (rendimiento del 70 al 85%). Primero, las secuencias MID se usan para distinguir entre muestras diferentes. Solo se permite 1 *mismatch*. En segundo lugar, se usan los cebadores específicos para distinguir entre diferentes regiones del genoma y entre las dos cadenas. Se permiten hasta 3 *mismatches*. Finalmente, los MID y los cebadores se recortan y se obtiene un archivo *fasta* para cada combinación de MID, cebador y cadena, donde los *reads* se colapsan en haplotipos con las frecuencias correspondientes.
5. En cada archivo *fasta*, los haplotipos se alinean con la secuencia de referencia o la secuencia *master* (haplotipo más abundante en el archivo) y se aplica un filtro de calidad (rendimiento >90%). Este filtro consiste en descartar haplotipos que no cubren el amplicón completo y aquellos que tienen más de 2 indeterminaciones, 3 gaps o más de 99 diferencias con respecto a la secuencia de referencia. Finalmente, las indeterminaciones y los gaps aceptados se reparan según la secuencia de referencia.
6. Se seleccionan los haplotipos con abundancia superior o igual al 0,1% y comunes en ambas cadenas (rendimiento del 50 al 60%), mientras que los únicos para una cadena se descartan. La cobertura de los haplotipos que pasan el filtro se toma como la suma de los *reads* en ambas cadenas.
7. Se descartan todos los haplotipos con abundancias inferiores al 0,25%. Los haplotipos finales son la base para el análisis posterior en este estudio (rendimiento global final del 15 al 25%).

4.12 Genotipaje

Los genotipos del VHD se determinaron mediante NGS y un análisis filogenético de la región del HDAG amplificada [ver apartado 4.5 *Amplificación del ARN del VHD para obtener librerías de amplicones para analizar por secuenciación masiva (NGS)*]. Los haplotipos de nucleótidos alineados al 0,25% obtenidos por la NGS fueron genotipados por análisis discriminante basado en distancias (regla DB) que corrige la distancia entre una secuencia problema y los distintos genotipos atendiendo a la variabilidad propia de cada genotipo.^{228,229} Para este análisis, se utilizaron secuencias de referencia de la región del HDAG analizada, extraídas de genomas completos representativos de los genotipos 1 a 8 del VHD, obtenidos de *GenBank*. Este análisis tiene en cuenta la variabilidad inter e intraclase de cada genotipo. Las distancias genéticas se calcularon según el modelo Kimura-80.²³⁰ Para visualizar las distancias genéticas entre secuencias se usan dendrogramas construidos mediante el método UPGMA (*Unweighted Pair Group Method with Arithmetic mean*). Los dendrogramas son representaciones gráficas en forma de árbol que resumen el proceso de agrupación en un análisis de grupos o *clusters*. En este caso, los *clusters* de secuencias se forman según sus distancias genéticas (Figura 23).

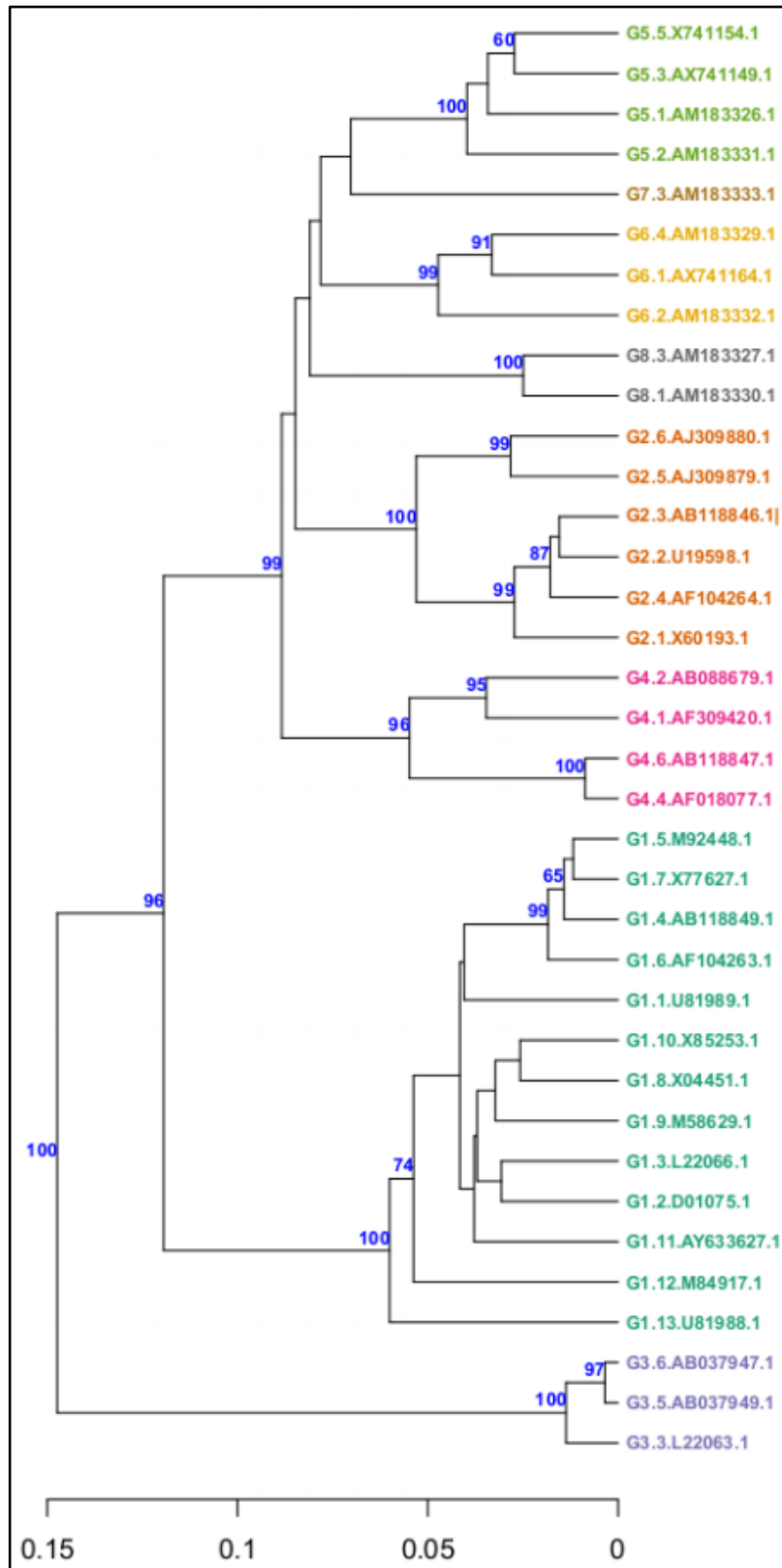


Figura 23. Análisis filogenético de las secuencias de referencia para el genotipado del VHD entre los nucleótidos 910 a 1270 del genoma viral, extraídas de secuencias de longitud completa obtenidas de *GenBank*. Los nodos del árbol UPGMA muestran los valores de *bootstrap* (números azules). Los diferentes colores representan los 8 genotipos del VHD. La escala muestra la distancia genética medida según el modelo de Kimura-80.

4.13 Clonación molecular

En el primer estudio se realizó una clonación usando el *Zero Blunt® TOPO® PCR Cloning Kit for Sequencing* (Thermo Fisher Scientific, Carlsbad, Estados Unidos) a partir del producto de PCR después de la amplificación de la región del HDAg, su comprobación de la amplificación y su purificación, tal como se describe en los apartados *4.5 Amplificación del ARN del VHD para obtener librerías de amplicones para analizar por secuenciación masiva (NGS)*, *4.6 Comprobación de la amplificación del ARN del VHD* y *4.7 Purificación del producto de PCR*. El *Zero Blunt® TOPO® PCR Cloning Kit for Sequencing* proporciona una alta eficiencia en la inserción directa de productos de PCR con extremos romos dentro de un plásmido. Cada kit incluye el vector *pCR™ 4Blunt-TOPO®* (Figura 24), con el que se transforman químicamente células competentes de *E. coli*. Este vector permite seleccionar directamente recombinantes al alterar el gen letal de *E. coli*, *ccdB*.²³¹⁻²³³ Este gen está fusionado con la secuencia codificante del extremo C-terminal del fragmento LacZα. La ligación del producto de PCR de extremos romos interrumpe la expresión de la fusión del gen *lacZα-ccdB* permitiendo el crecimiento de solo recombinantes positivos tras la transformación en *E. coli*. Al incubar las bacterias transformadas en medio Luria-Bertrani sólido (LB) (Tabla 9) durante toda la noche a 37°C, las células que contienen un vector no recombinante no sobreviven y por lo tanto no pueden formar colonias.²³⁴

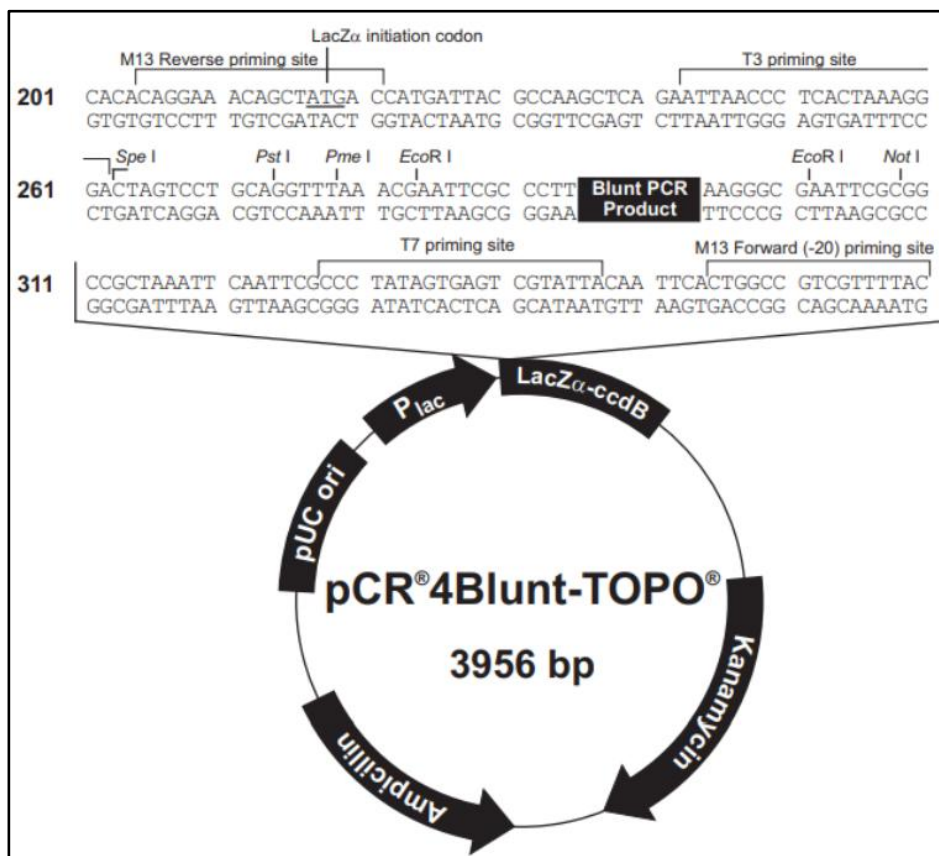


Figura 24. Vector pCR[™] 4Blunt-TOPO[®].²³⁵

Medio LB sólido (500 mL)	
Agar	7,5 gr
Triptona	5 gr
Levadura	2,5 gr
NaCl	5 gr
Agua estéril	500 mL

Tabla 9. Componentes del medio LB sólido.

Los pasos que se han seguido para la clonación molecular son los descritos por el fabricante del kit el Zero Blunt[®] TOPO[®] PCR Cloning Kit for Sequencing,²³⁴ brevemente:

1. Mezclar el producto de PCR y el vector pCR[™] 4Blunt-TOPO[®]. La concentración de vector debe ser tres veces superior a la concentración del ADN amplificado

para garantizar que un solo fragmento de ADN se incorpora a una molécula de vector. Las concentraciones de los productos de PCR se cuantificaron mediante NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, Estados Unidos) como se explica en el apartado 4.9 *Cuantificación del producto de PCR purificado*.

2. A partir de las concentraciones del producto de PCR a clonar, se calcula el volumen (μL) de los componentes de la reacción de clonación (producto de PCR, sal, agua y vector). El uso de sal permite aumentar el número de transformantes porque permite tiempos de incubación más largos. Esto se debe a que evita que la topoisomerasa I vuelva a unirse al plásmido después de ligar el producto de PCR, con lo que podría disociarlo de nuevo.
3. Incubar durante 5 minutos a temperatura ambiente.
4. Transformar la reacción de clonación en bacterias *E. coli* TOP10 competentes, para que en el momento de replicarse, se obtengan clones del fragmento de ADN de interés. Esta inoculación se realiza mediante un choque térmico en la bacteria, que produce poros en la célula que dejan pasar el plásmido al interior. Después del choque térmico, la célula vuelve al estado original sin sufrir ningún daño.
5. Sembrar las bacterias en medio de cultivo LB (Tabla 9) a 37 °C. A este medio se le añade ampicilina (100 mg/mL) diluida 1:100, para seleccionar las bacterias que hayan incorporado el plásmido. Esto es posible gracias a que este plásmido contiene un gen de resistencia a la ampicilina, lo que permite a estas células crecer en este medio.
6. A las 24 horas, se seleccionan 20 colonias de cada muestra obteniendo así 20 clones de cada paciente.
7. Cultivar estas colonias por separado en medio de cultivo LB líquido (se prepara igual que el medio LB sólido (Tabla 9) pero no se añade Agar), al que también se añade ampicilina (100 mg/mL) diluida 1:100, durante 24 horas a 37°C.
8. Una vez crecidas en este medio, centrifugar cada muestra para precipitar las células y eliminar el sobrenadante.

4.14 Purificación del producto de clonación

A partir del precipitado de células obtenido, se purifica el plásmido siguiendo el protocolo del *QIAprep® Spin Miniprep Kit* (QIAGEN®, Hilden, Alemania). El procedimiento se basa en la lisis alcalina de células bacterianas, seguido de la adsorción del ADN sobre la membrana de sílice en presencia de alto contenido de sal. En dicha membrana el ADN se someterá a distintos pasos de lavado y se eluye en tampón 10 mM Tris-Cl, pH 8,5. Este kit está diseñado para el aislamiento de hasta 20 µg de ADN plasmídico de alta pureza para su uso en aplicaciones rutinarias de biología molecular, incluida la secuenciación.²³⁶

Para comprobar la recuperación del ADN plasmídico y su tamaño, el producto purificado obtenido se verifica mediante una electroforesis en gel de agarosa al 1,5% con *Sybr Safe® DNA Gel Stain* como se explica en el apartado 4.6 *Comprobación de la amplificación del ARN del VHD*. Además este producto se cuantifica mediante NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, Estados Unidos) como se explica en el apartado 4.9 *Cuantificación del producto de PCR purificado*.

4.15 Reacción de secuenciación Sanger

Los plásmidos purificados se secuenciaron mediante el método Sanger. En primer lugar se realiza la reacción de secuenciación utilizando el *BigDye® Terminator v3.1 Cycle Sequencing Kit* (Thermo Fisher Scientific, Foster City, Estados Unidos) que contiene la *BigDye™ Terminator v3.1 Ready Reaction Mix*, que consta de dNTPs, didesoxiribonucleótidos trifosfato (ddNTPs), una ADN polimerasa, y el 5x *Sequencing Buffer*. El método de Sanger consiste en la utilización de diferentes ddNTPs (ddATP, ddTTP, ddCTP y ddGTP), cada uno marcado con un fluorocromo diferente, como terminadores de cadena que se incorporan en el extremo 3' de la cadena en crecimiento y el alargamiento de la cadena se termina selectivamente en A, C, G o T porque el ddNTP carece de un grupo 3'-hidroxilo. Esto da lugar a una mezcla de fragmentos de ADN de diferentes longitudes. En esta reacción se utilizan los cebadores descritos en la Tabla 10, los componentes descritos en la Tabla 11 y el protocolo descrito en la Tabla 12.

Cebador	Secuencia 5' - 3'
T7 FW	TAATACGACTCACTATAGGG
T3 RV	ATTAACCCTCACTAAAGGGA

Tabla 10. Cebadores utilizados en la reacción de secuenciación. Cada cebador se utiliza en una reacción de secuenciación diferente para diferenciar entre cadena fw y cadena rv en el posterior análisis.

Componentes	Concentración	Volumen (µL)
Agua estéril	-	5,75
Tampón de secuenciación	5 x	1,75
Cebador T7 FW / T3 RV	10 µM	1
BigDye™ Terminator v3.1 Ready Reaction Mix	-	0,5
Producto plasmídico purificado	-	1

Tabla 11. Componentes, concentración y volumen utilizados en la reacción de secuenciación por el método de Sanger.

	Temperatura	Tiempo
Desnaturalización	96°C	1 minuto
Amplificación (30 ciclos)	96°C	10 segundos
	50°C	5 segundos
	60°C	4 minutos
Enfriamiento	20°C	∞

Tabla12. Tiempos y temperaturas de la reacción de secuenciación por el método de Sanger.

En segundo lugar, se realiza la purificación del producto de la reacción de secuenciación mediante el *BigDye® XTerminator™ Purification Kit* (Thermo Fisher Scientific, Foster City, Estados Unidos). Éste método de purificación requiere la adición de sólo dos reactivos que pueden agregarse secuencialmente o premezclados, la Solución XTerminator™ (que consiste en una resina sintética que captura los ddNTPs fluorescentes no incorporados y las sales libres) y la Solución SAM™ (que mejora el rendimiento de la solución XTerminator™ y estabiliza las reacciones posteriores a la

purificación). Una vez capturados los ddNTPs y las sales, la resina XTerminator es precipitada por centrifugación.²³⁷

Por último, los fragmentos de ADN de diferentes longitudes producidos durante la reacción de secuenciación son separados según su longitud mediante electroforesis capilar, utilizando el secuenciador automático, ABI PRISM 3130XL (Thermo Fisher Scientific, Foster City, Estados Unidos). La fluorescencia emitida por los fragmentos de la misma longitud se registra por una cámara CCD durante su migración a través del capilar.²³⁸⁻²⁴⁰ De la integración de estos datos de fluorescencia se obtienen electroferogramas (Figura 25), a partir de los que se generan archivos *fasta* con la secuencia obtenida, que se han revisado y editado utilizando el programa BioEdit Sequence Alignment Editor.²⁴¹

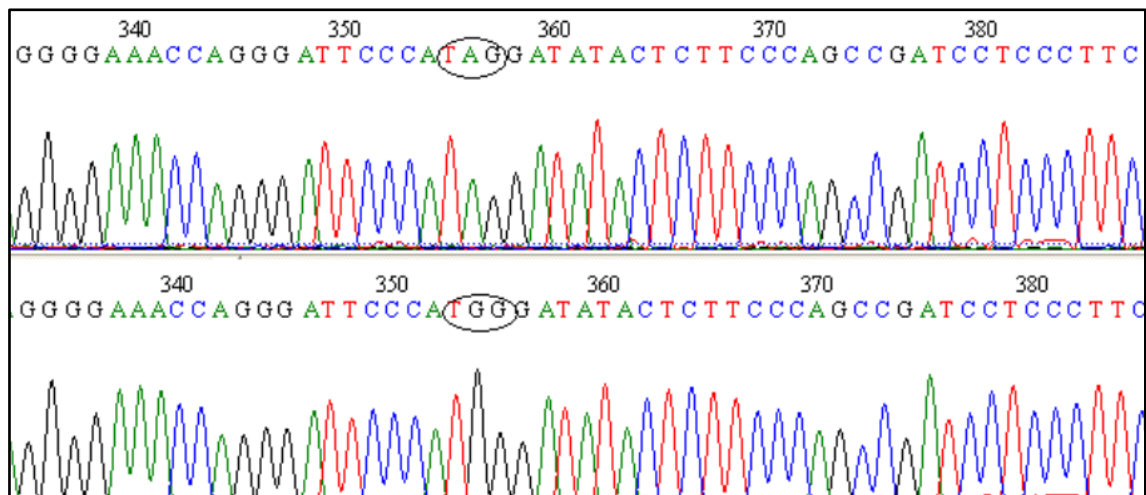


Figura 25. Ejemplo de dos electroferogramas, el de arriba de un genoma no editado y el de abajo de un genoma editado. Cada pico del electroferograma corresponde a una base y a cada una de las bases le corresponde un color. Los codones de edición están marcados con un círculo.

Esta metodología permite determinar la secuencia de fragmentos de ADN de entre 300 y 900 nucleótidos pero no permite determinar secuencias de moléculas de ADN individuales como las técnicas de secuenciación masiva, sino una única secuencia donde los nucleótidos de cada posición sean el promedio de los representados en una proporción $\geq 20\%$ de las moléculas secuenciadas. Así pues, al analizar una quasiespecies directamente por esta metodología no se obtendrá la secuencia de haplotipos individuales, si no que la secuencia obtenida es la media de los haplotipos presentes en la quasiespecies en una proporción $\geq 20\%$.²⁴² Para secuenciar los haplotipos individuales mediante el método de Sanger es necesario un paso previo de clonaje molecular,

descrito en los dos apartados anteriores, que hace posible amplificar estos haplotipos en los diferentes clones lo cual permite su secuenciación.

4.16 Análisis del porcentaje de genomas no editados y editados

En el primer estudio se cuantificaron los porcentajes de genomas no editados y editados en los archivos *fasta* obtenidos mediante la NGS en la plataforma Roche 454 GS Junior y la secuenciación Sanger de los productos de clonación en muestras de pacientes infectados por VHD. El porcentaje de haplotipos por muestra con codones CUA o CCA en las posiciones nucleotídicas 1011 a 1013 del genoma del VHD (traducido a TAG o TGG en el codón 196, respectivamente) determinó estos porcentajes de genomas no editados (TAG, codón stop, que codifica el S-HDAg) y editados (TGG, codón triptófano [W], que codifica el L-HDAg). Dichos porcentajes se calcularon en los archivos *fasta* de cada muestra, utilizando el programa BioEdit Sequence Alignment Editor.²⁴¹ En los *fasta* obtenidos por NGS se sumaron las frecuencias de los haplotipos con codones TAG y con codones TGG. En los obtenidos por secuenciación Sanger de los productos de clonación, se contó y sumó el número de haplotipos de cada tipo (Figura 26).

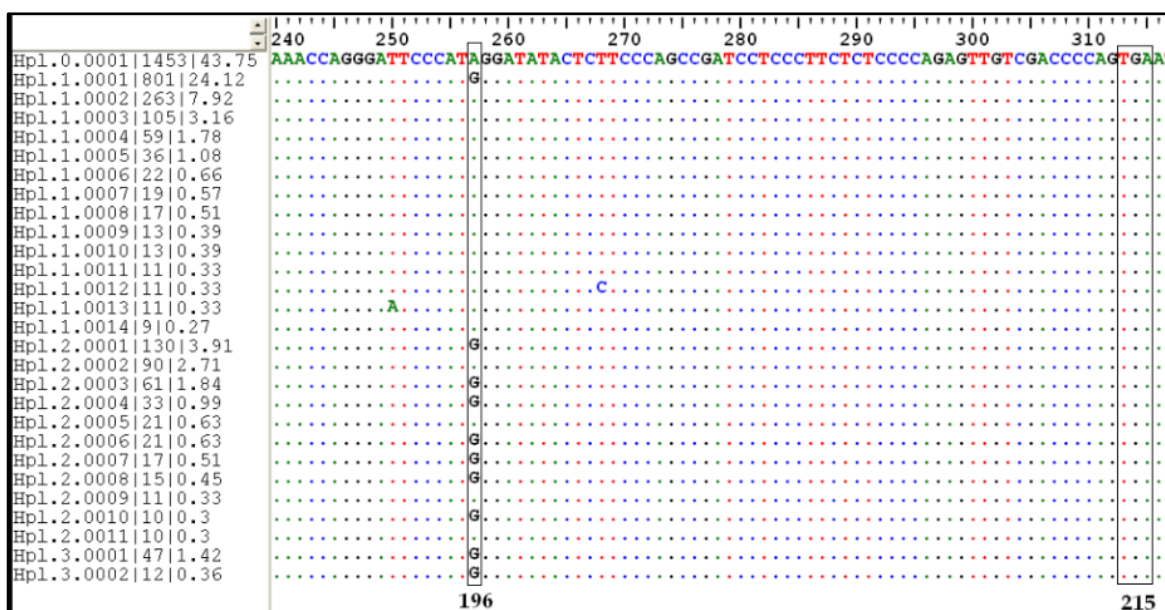


Figura 26. Alineamiento de haplotipos obtenidos de una de las muestras incluidas utilizando el programa BioEdit Sequence Alignment Editor.²⁴¹ El codón de edición 196 (TAG para haplotipos no editados y TGG para haplotipos editados) y el codón 215 (codón stop TGA para L-HDAg) están enmarcados.

Después de la cuantificación de haplotipos de formas editadas y no editadas del genoma viral por ambos métodos, se realizó la comparación entre la cuantificación de estas formas obtenida por NGS y por secuenciación Sanger de los productos de clonación. Teniendo en cuenta que la edición en el codón 196 es un evento binario (CUA o CCA), que tiene lugar de forma independiente en los diferentes genomas del VHD secuenciados, se ha considerado que la frecuencia de este evento sigue una distribución binomial. Por tanto, las frecuencias de los codones no editados y editados son complementarias (la suma de sus proporciones es 1). Entonces, para simplificar la comparación de ambas metodologías, se han analizado gráficamente los intervalos de confianza al 95% de los porcentajes de codones no editados obtenidos por ambas metodologías.

El análisis de este apartado, del apartado *4.17 Cuantificación por NGS de las mezclas de clones con diferentes porcentajes de genomas no editados y editados* y del apartado *4.18 Análisis del tipo de cambios nucleotídicos* se han realizado con scripts desarrollados internamente en R.²⁴³

4.17 Cuantificación por NGS de las mezclas de clones con diferentes porcentajes de genomas no editados y editados

En el primer estudio, para evaluar la fiabilidad de los porcentajes de haplotipos editados y no editados reportados por NGS en cada muestra, se crearon una serie de mezclas artificiales de clones. De las muestra de los pacientes incluidos en este primer estudio, se escogió un clon de un genoma editado y otro sin editar. El ADN plasmídico purificado de estas muestras se cuantificó por fluorescencia, tal como se ha explicado en el apartado *4.9 Cuantificación del producto de PCR purificado*. Se debe verificar que la cantidad de ADN está entre 5-30 µg, el rango óptimo de concentraciones a las que la ADN polimerasa Pfu Ultra II (Agilent Technologies, Santa Clara, Estados Unidos) puede amplificar un vector. El fragmento de ADN viral insertado en cada plásmido fue luego amplificado por una PCR MID, utilizando los cebadores, los componentes y el protocolo para la PCR MID del primer estudio, descritos en la Tabla 4, en la Tabla 5 y en la Tabla 6, respectivamente. Los clones de una misma mezcla (no editado y editado), pertenecían a un mismo paciente y, por tanto, tenían el mismo MID. A continuación, para verificar la presencia del ADN que contiene el producto de la PCR MID, se realiza una electroforesis en gel de agarosa al 1,5% como se ha explicado en el apartado *4.6 Comprobación de la amplificación del ARN del VHD* y posteriormente estos productos se purifican como se ha

explicado en el apartado 4.7 *Purificación del producto de PCR*. Para verificar la purificación del ADN se utiliza la plataforma de electroforesis capilar Agilent 2200 TapeStation como se ha explicado en el apartado 4.8 *Comprobación de la calidad del producto de PCR purificado*. Seguidamente, se cuantifica de nuevo el ADN purificado mediante fluorimetría como se ha explicado en el apartado 4.9 *Cuantificación del producto de PCR purificado*. A través de diluciones con Tris-EDTA, pH 8, se llevan todas las muestras a igual concentración. A continuación, se mezclan los ADN de secuencias editadas y no editadas de cada paciente (ambas con el mismo MID) en las proporciones deseadas (90 no editado:10 editado, 80:20, 70:30, 60:40, 30:70, 10:90 y 5:95). Por tanto, se obtienen 7 mezclas de clones. Seguidamente, estas mezclas se vuelven a cuantificar del mismo modo que antes, para verificar que la mezcla no ha variado la concentración.

Una vez verificada la concentración, se crea un *pool* que contiene todas las mezclas de clones con el genoma editado y no editado a igual concentración. Esto sirve para asegurar que los amplicones de cada mezcla estén representados en cantidades equimoleculares en el posterior análisis de los datos de NGS, tal como se describe en el apartado 4.9 *Cuantificación del producto de PCR purificado*.

Para realizar la NGS se utiliza el sistema Roche 454 GS Junior siguiendo su protocolo descrito en el apartado 4.10.1 *NGS mediante la plataforma Roche 454 GS Junior*. De las secuencias obtenidas, se cuantifica el porcentaje de haplotipos de genomas no editados y editados (4.16 *Análisis del porcentaje de genomas no editados y editados*) y se analiza si cada mezcla contiene el porcentaje de genomas no editados y editados esperado, correspondiente a sus proporciones en la mezcla artificial creada. Para ello se asume una distribución binomial de porcentajes de codones editados y no editados, tal como se ha explicado en el apartado 4.16 *Análisis del porcentaje de genomas no editados y editados*. Por lo tanto, solo se utilizan los codones no editados para comparar las proporciones esperadas con las observadas en cada mezcla, utilizando el análisis de la regresión entre estas proporciones.

4.18 Análisis del tipo de cambios nucleotídicos

En el primer estudio, los cambios nucleotídicos que ocurren a lo largo de la región amplificada del HDAG se evaluaron en todas las muestras, con el fin de confirmar si la edición del genoma viral es un fenómeno general que tiene lugar en toda esta región o solo en el codón de edición (196). Los tipos de cambios se determinaron agrupando y

alineando todos los haplotipos obtenidos en un único archivo *fasta*. Así se registraron todos los cambios de nucleótidos observados, en comparación con los nucleótidos más frecuentes observados en cada posición del alineamiento. Cada cambio de nucleótido se contó sin tener en cuenta la frecuencia de los haplotipos en los que se encontraba en la quasiespecies del VHD, para evitar un factor de confusión debido a la aptitud replicativa (*fitness*) del haplotipo. Además, para evitar sesgos al determinar las frecuencias de transiciones y transversiones, se excluyó el sitio de edición del HDAg (nucleótido 1012 dentro del codón 196), ya que esta posición es un punto caliente para cambios de A a G [A:G] en el ARN antígeno.

4.19 Análisis de la conservación y variabilidad de la región del HDAg y de la región de la ribozima

En el segundo estudio, se ha analizado la conservación/variabilidad de la quasiespecies del VHD en las regiones del HDAg y de la ribozima, definidas en el apartado 4.5 *Amplificación del ARN del VHD para obtener librerías de amplicones para analizar por secuenciación masiva (NGS)* mediante el contenido de información (CI). El CI se define como el número de decisiones binarias (número de preguntas cuya respuesta es sí o no) requerido para encontrar el elemento correcto en un conjunto de N elementos. Como la probabilidad de encontrar un nucleótido o aminoácido en una posición determinada de un alineamiento múltiple es incierta, podemos definir el CI basado en la entropía de Shannon como se puede observar en la Figura 27. El CI se mide en bits.²⁴⁴

$$CI_j = \log_2(N) - \sum_{i=1}^j p_{ij} \log_2(p_{ij}) *$$

Figura 27. Fórmula para calcular el CI. j: posición j en el alineamiento y, p_j: frecuencia del haplotipo j en la quasiespecies viral, en la posición j, N: el número de posibilidades (4 nucleótidos y 20 aminoácidos). El CI varía desde 0, lo que significa máxima incertidumbre o variabilidad; hasta 2 bits en nucleótidos y 4,322 bits en aminoácidos, lo que significa máxima información o conservación.

El cálculo del CI se puede realizar, por un lado, a partir de un alineamiento múltiple de los haplotipos presentes en la quasiespecies y sin tener en cuenta la frecuencia relativa de cada uno de estos haplotipos (análisis por haplotipos). Es decir, desde este enfoque cada secuencia tiene el mismo peso que las demás y los resultados de conservación no se

ven afectados por el *fitness* de cada haplotipo. De esta forma también se evita que las variantes minoritarias, que podrían escapar a una posible terapia génica, queden enmascaradas en el análisis. Por otro lado, el análisis por frecuencia de haplotipos, donde el alineamiento múltiple sobre el cual se calcula el CI se hace también a partir de los haplotipos presentes en la quasispecies pero teniendo en cuenta la frecuencia relativa de cada uno de estos haplotipos (análisis por frecuencia de haplotipos).

Para el estudio de la conservación/variabilidad se ha utilizado el análisis mediante *Sliding window*, en el cual se ha hecho el análisis por haplotipos y por frecuencia de haplotipos. Esta herramienta sirve para analizar fragmentos de la secuencia de los haplotipos (ventanas) de N nucleótidos o aminoácidos, comenzando desde la primera posición en el alineamiento múltiple y avanzando en pasos de 1 posición. Es decir, cada punto del *sliding window* nos indica la media del CI de cada ventana de N nucleótidos o aminoácidos, el siguiente punto representaría otra ventana de N nucleótidos o aminoácidos en los que se incluirían todos los de la ventana anterior excluyendo el primero y añadiendo el siguiente del alineamiento y así sucesivamente. El objetivo es representar gráficamente las medias obtenidas en cada ventana para localizar el fragmento o fragmentos de N nucleótidos o aminoácidos más conservado o los más variables, dentro del alineamiento múltiple. En el caso de nucleótidos hablaremos de ventanas de 25 y en el caso de aminoácidos de ventanas de 10.

Otra manera de obtener una representación gráfica del CI es mediante logos de secuencias, que se han hecho mediante un análisis por haplotipos. Estos logos muestran la conservación/variabilidad posición a posición de un fragmento de interés dentro de un haplotipo, ya sea representando nucleótidos o aminoácidos. En el caso del HDAg se han representado los logos más conservados y el más variable y en el caso de la ribozima, aparte del logo más conservado y del más variable, también se ha realizado un logo de toda su secuencia de nucleótidos, en el cual se han definido las diferentes regiones. En estas representaciones aparecen apiladas en cada posición las letras que representan los nucleótidos o aminoácidos identificados en el alineamiento múltiple. En cada pila, los tamaños relativos de las letras indican sus frecuencias relativas en cada posición dentro del alineamiento múltiple de los haplotipos. Además, la altura total de cada letra o pila de letras representa el CI de cada posición, medido en bits desde la mínima (0) hasta la máxima conservación (4.322 bits para la secuencia de aminoácidos y 2 bits para la secuencia de nucleótidos).²⁴⁵

4.20 Estudio de los patrones de cambios de aminoácidos de la región del HDAg

En la región del HDAg analizada, también se realizó un estudio de posibles patrones de cambios de aminoácidos (entre las posiciones 111 y 214 del L-HDAg). En primer lugar, se determinaron las mutaciones aminoacídicas presentes en cada haplotipo obtenido en cada muestra respecto a una secuencia de referencia del VHD de genotipo 1 (Figura 28).²⁴⁶



Figura 8. Secuencia de referencia de aminoácidos del VHD de genotipo 1 utilizada para el análisis (*GenBank Data Library*, número de acceso: AF008309 – AF008375).²⁴⁶ La numeración indica la posición de los diferentes aminoácidos en la secuencia del HDAg.

Las frecuencias relativas de las mutaciones aminoacídicas identificadas se correlacionaron entre sí mediante una prueba de Spearman y se eligieron las mutaciones cuyas frecuencias tenían un coeficiente de correlación (ρ) $>0,8$ (nivel de correlación elevado) y un p -valor $<0,05$ (dintel seleccionado para establecer la significación estadística). Con estas mutaciones seleccionadas, se realizó un *Cluster analysis* que es un método estadístico que tiene como objetivo clasificar varios objetos en grupos (*clusters*) según la distancia o similitud entre ellos (en este caso, entre las frecuencias relativas de las mutaciones elegidas), de manera similar a la clasificación de los haplotipos de nucleótidos según su distancia genética (como se explica en el apartado 4.12 *Genotipaje*). Este análisis se llevó a cabo mediante el software de análisis R (versión 3.3.3)²⁴³ y se utilizó *Pvclust*^{247,248} que es una implementación del *bootstrap analysis* para evaluar la incertidumbre en el *hierarchical cluster analysis*, una forma de agrupamiento jerárquico en la que se representan los *clusters* en forma de dendrograma. Los objetos similares se conectan mediante enlaces cuya posición en el dendrograma está determinada por el nivel de similitud/disimilitud entre los objetos. El valor de la distancia será cero cuando la variable esté perfectamente correlacionada. El método del promedio se utilizó para definir el vínculo entre las mutaciones. Este método calcula el promedio de

proximidad de una pareja de observaciones entre todas las observaciones que forman los *clusters*. Para cada *cluster* se calculan dos tipos de parámetros: el *Bootstrap Probability value* (BP *value*) que indica la frecuencia con la que aparece un *cluster* en las diferentes réplicas de *bootstrap* y el *Approximately Unbiased P-value* (AU *p-value*) que es el p-valor en multiescala (100 - p-valor) de cada *cluster*, por lo tanto un AU *p-value* igual o por encima del 95% da una fiabilidad muy alta a los *clusters* obtenidos.^{249,250}

Además, la evolución a través del seguimiento longitudinal de la frecuencia relativa de los patrones de mutaciones identificados se representó gráficamente para estudiar su variación en cada una de las muestras que contenían el patrón de mutaciones.

Por último, se hizo una búsqueda bibliográfica para determinar la funcionalidad de las regiones del HDAg en las que se encontraban dichas mutaciones.

RESULTADOS

5. RESULTADOS

5.1 Primer estudio: Caracterización cuantitativa de la edición del genoma del VHD por secuenciación masiva (NGS)

En este primer estudio se validó el método de NGS para cuantificar porcentajes de poblaciones de genomas editados y no editados en el VHD, comparándolo con el método de secuenciación Sanger de los productos de clonación. Además, se pretendía confirmar si la edición es un fenómeno general que tiene lugar en toda la región codificante del HDAg analizada.

La cuantificación de genomas editados y no editados se realizó por secuenciación Sanger de los productos de la clonación molecular y por NGS en las muestras de los pacientes incluidos. Además, para evaluar la fiabilidad de los valores obtenidos por NGS, se crearon mezclas artificiales con diferentes proporciones conocidas de clones editados y no editados y se cuantificaron por NGS comparando las proporciones obtenidas con las esperadas (es decir, las mismas proporciones en las que se habían creado las mezclas). Por último, se determinaron los cambios de nucleótidos que ocurren en la región amplificada después de excluir el codón de edición (codón 196), para evaluar la posible edición del genoma a lo largo de esta región.

Por estos motivos, el diseño de este estudio fue transversal retrospectivo, incluyendo una muestra de plasma de 4 pacientes con hepatitis crónica delta atendidos en el Hospital Universitario Vall d'Hebron (Barcelona, España) (Figura 29). Las características demográficas y clínicas más relevantes de los pacientes se muestran en la Tabla 13. Las muestras de estos pacientes se seleccionaron de acuerdo con los siguientes criterios de inclusión: test anti-VHD positivo en los 6 meses anteriores (para confirmar infección crónica), VHD genotipo 1, ausencia de otras coinfecciones por el VHC o el VIH, ausencia de enfermedad hepática autoinmune y ausencia de tratamiento antiviral en el momento de recolección de la muestra.

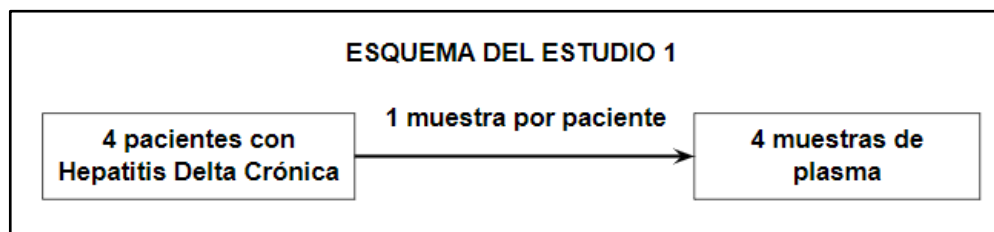


Figura 29. Esquema de los pacientes y las muestras incluidos en el estudio 1: Caracterización cuantitativa de la edición del genoma del VHD por NGS.

Paciente	Sexo	Edad (años)	ALT (UI/L)	HBeAg	ARN-VHD (logUI/mL)
1	Masculino	35	42	Negativo	7,02
2	Femenino	39	83,4	Negativo	5,56
3	Masculino	47	295,8	Negativo	6,88
4	Masculino	49	104,4	Negativo	6,54

Tabla 13. Características demográficas y clínicas de los pacientes incluidos en el estudio.

5.1.1 Comparación entre las proporciones de genomas no editados y editados obtenidas por NGS y por secuenciación Sanger de los productos de la clonación molecular

Mediante las secuencias obtenidas por NGS en la plataforma Roche 454 GS Junior y los electroferogramas obtenidos en la secuenciación Sanger de cada clon, entre las posiciones nucleotídicas 910 a 1270 del genoma viral, se calculan las proporciones de poblaciones con genomas no editados y editados de cada paciente (Figura 30 y Tabla 14), tal y como se explica en el apartado 4.16 *Análisis del porcentaje de genomas no editados y editados* de Material y Métodos.

Por un lado, se secuenciaron un total de 70 clones (entre 15 y 19 por paciente, Tabla 14). Y, por otro lado, se obtuvieron 11.208 *reads* en el análisis por NGS de los 4 pacientes (media= 2.802 *reads*/muestra, SD= 612,71).

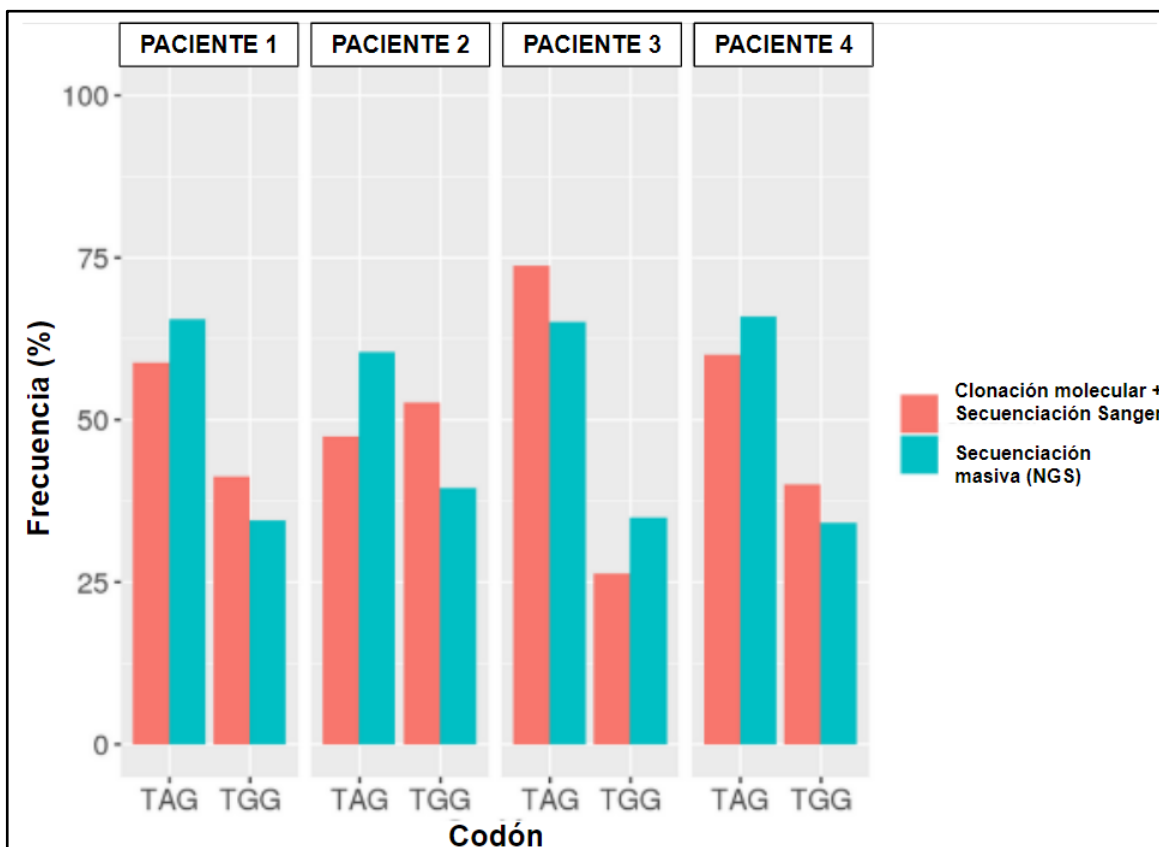


Figura 30. Gráfico de las proporciones de genomas no editados (TAG) y editados (TGG) en el codón 196 del HDAG de los 4 pacientes analizados, observadas mediante clonación molecular más secuenciación Sanger (en rosa) y secuenciación masiva (NGS) (en azul).

	Clonación + Secuenciación Sanger				Secuenciación masiva (NGS)			
	Codón TAG		Codón TGG		Codón TAG		Codón TGG	
	Clones (n/total)	%	Clones (n/total)	%	Reads (n/total)	%	Reads (n/total)	%
Paciente 1	10/17	58,8	7/17	41,2	2175/3321	65,5	1146/3321	34,5
Paciente 2	9/19	47,4	10/19	52,6	1512/2499	60,5	987/2499	39,5
Paciente 3	14/19	73,7	5/19	26,3	2150/3303	65,1	1153/3303	34,9
Paciente 4	9/15	60	6/15	40	1374/2085	65,9	711/2085	34,1

Tabla 14. Proporciones de genomas no editados (TAG) y editados (TGG) en el codón 196 del HDAG de los 4 pacientes, observadas mediante clonación más secuenciación Sanger y secuenciación masiva (NGS).

Los genomas no editados predominaron sobre los editados en las 4 muestras analizadas por NGS y en 3 de las 4 muestras analizadas por secuenciación Sanger de los clones (Figura 30 y Tabla 14). Para simplificar la comparación de las proporciones de edición obtenidas por NGS con las obtenidas de los clones secuenciados usando el método de Sanger se utilizaron los genomas no editados, como se explica en el apartado 4.16 *Análisis del porcentaje de genomas no editados y editados* de Material y Métodos. Las diferencias entre las proporciones de edición obtenidas a partir de estos dos métodos se evaluaron mediante la comparación de los intervalos de confianza del 95% (Figura 31). Los intervalos de confianza para ambos métodos se superponen en todas las muestras, lo que indica que las proporciones obtenidas por la secuenciación Sanger de los clones y por la NGS fueron comparables. Además, los intervalos de confianza de las frecuencias obtenidas por la clonación más la secuenciación Sanger fueron más amplios que los obtenidos por NGS, lo que indica que esta metodología fue más precisa que la clonación.

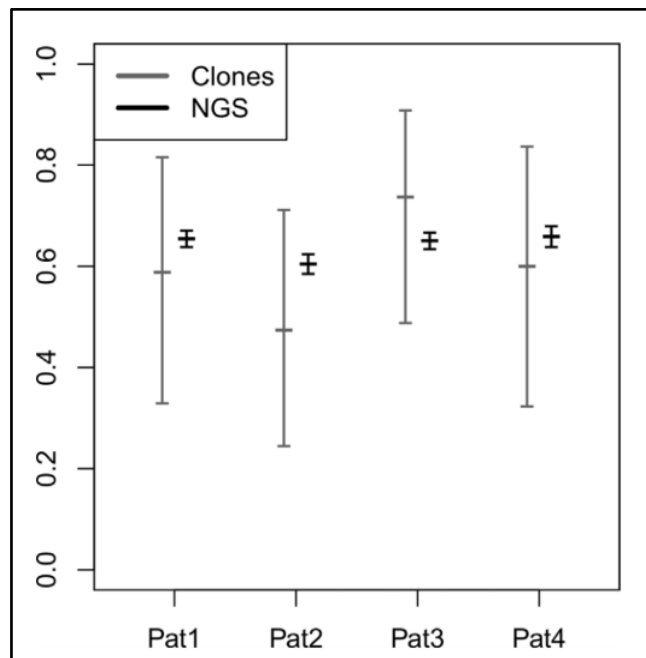


Figura 31. Intervalos de confianza del 95% para las proporciones de codones no editados obtenidos por NGS y por secuenciación Sanger de los productos de clonación, en las muestras de los 4 pacientes estudiados.

5.1.2 Cuantificación por NGS de las mezclas de clones con diferentes proporciones de genomas no editados y editados

Las mezclas de clones creadas artificialmente se cuantificaron mediante NGS, tal y como se explica en el apartado 4.17 *Cuantificación por NGS de las mezclas de clones con*

diferentes porcentajes de genomas no editados y editados de Material y Métodos. En total, se obtuvieron 83.276 reads (media= 13.879 reads/muestra; SD= 4.154,58), después de pasar los filtros de calidad. En la Tabla 15, se muestran las proporciones de genomas no editados y editados de las mezclas creadas y que por lo tanto se espera obtener al cuantificar estos genomas, en comparación con las proporciones obtenidas al cuantificar estas mismas mezclas mediante NGS. Las proporciones de genomas no editados y editados son muy similares tanto en las mezclas de clones creadas artificialmente como en las mezclas de clones cuantificadas por NGS. En la Figura 32 se muestran los reads obtenidos mediante NGS de alguna de las diferentes mezclas de genomas no editados y editados cuantificadas.

Mezclas de clones creadas artificialmente (Proporciones esperadas)		Mezclas de clones cuantificadas por NGS (Proporciones obtenidas)	
% TAG Genoma no editado	% TGG Genoma editado	% TAG Genoma no editado	% TGG Genoma editado
90	10	87,68	12,32
80	20	81,64	18,36
70	30	67,68	32,32
60	40	60,56	39,44
30	70	31,69	68,31
10	90	8	92
5	95	3,81	96,19

Tabla 15. Comparación de las proporciones de genomas no editados y editados de las mezclas creadas artificialmente y de las mezclas cuantificadas por NGS.

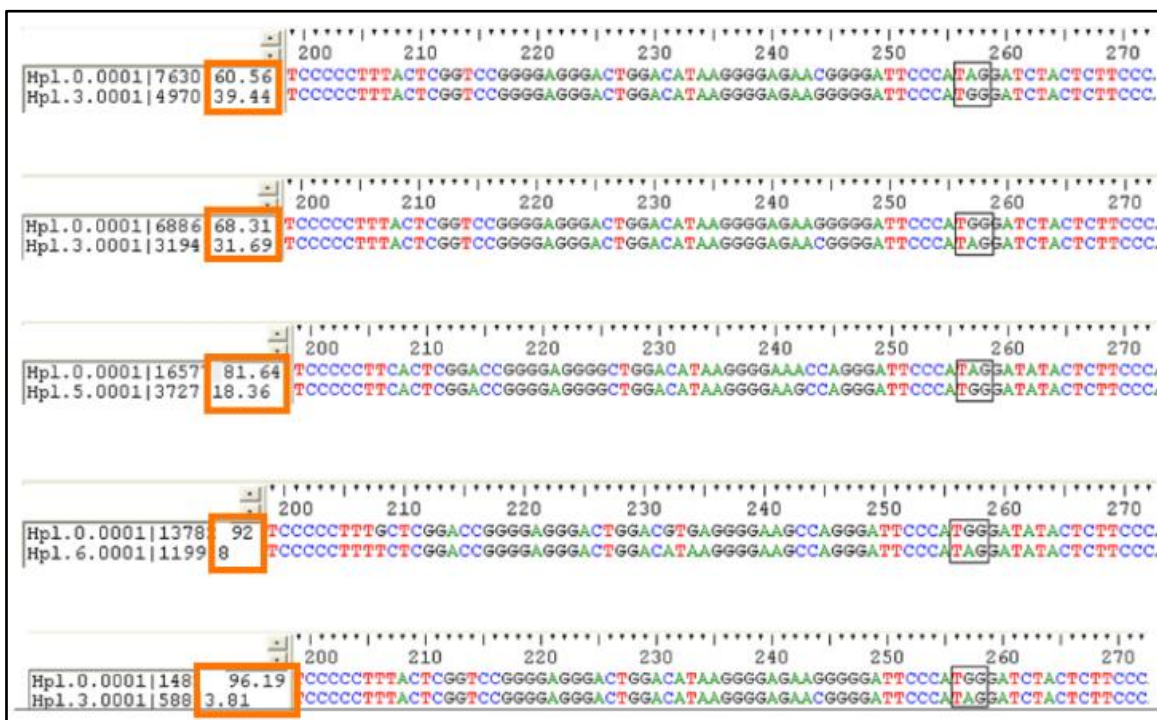


Figura 32. Alineamiento de los haplotipos obtenidos de algunas de las mezclas de clones cuantificadas por NGS utilizando el programa BioEdit Sequence Alignment Editor.²⁴¹ El codón de edición 196 (TAG para haplotipos no editados y TGG para haplotipos editados) está enmarcado en negro y la proporción de cada codón está enmarcada en naranja.

Para una comparación más exhaustiva de los datos, se calculó la regresión lineal entre las proporciones de los genomas no editados obtenidas en cada mezcla de clones con los porcentajes esperados en la cuantificación por NGS en dicha mezcla. Dicha regresión se superpuso casi al ajuste perfecto, indicando una buena correlación entre estos porcentajes (Figura 33). Por otro lado, la resta de las proporciones esperadas de las observadas reveló que NGS tendía a subestimar los porcentajes de codones no editados, ya que la línea de regresión estaba ligeramente por debajo del ajuste perfecto (Figura 34).

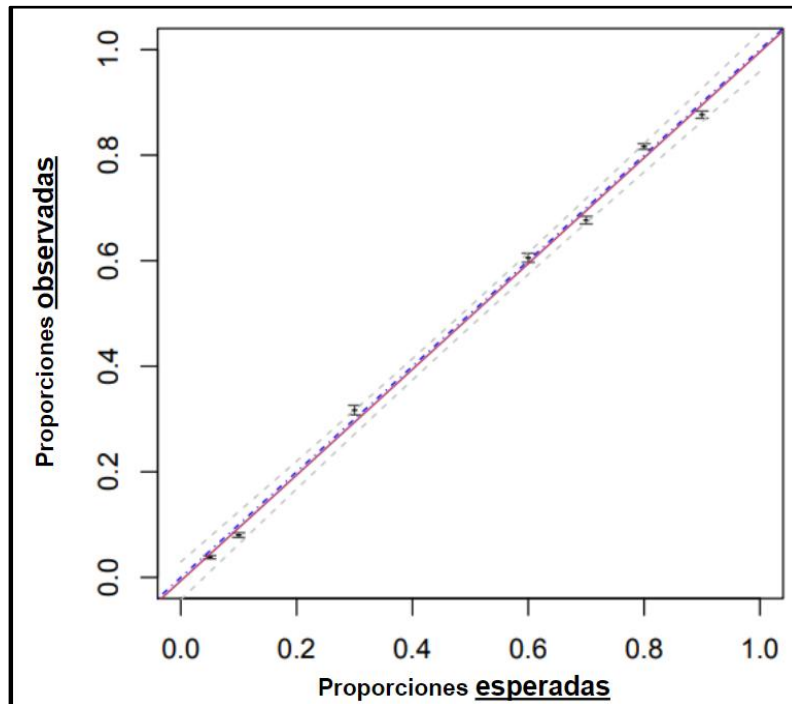


Figura 33. Diagrama de dispersión de las proporciones de genomas no editados observadas versus las esperadas con las líneas de regresión e intervalo de confianza del 95%. La línea de puntos y guiones de color azul representa el ajuste perfecto, la línea continua de color rosa es la línea de regresión y la línea discontinua de color gris representa los límites del intervalo de confianza del 95% de la regresión.

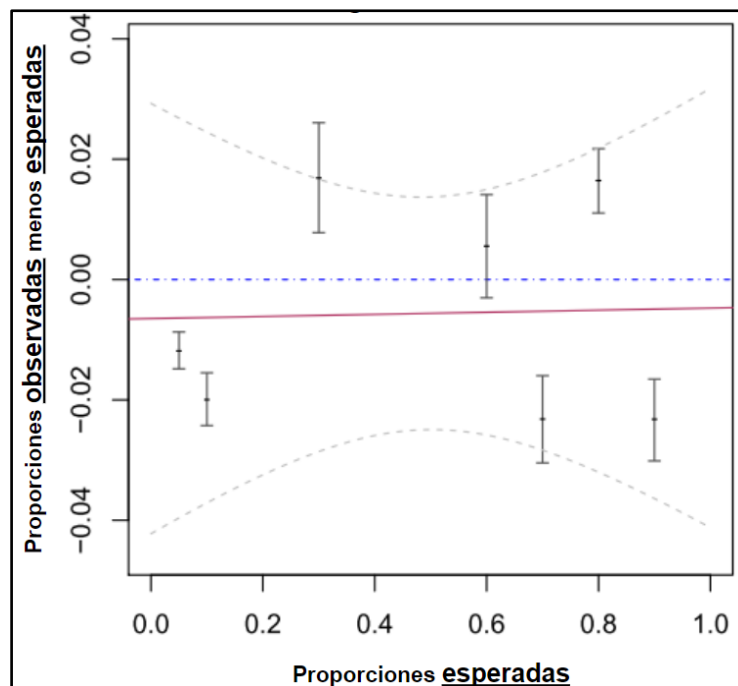


Figura 34. Proporciones de genomas no editados observadas menos esperadas, con línea de regresión, intervalos de confianza binomial del 95% en cada punto e intervalo de confianza de regresión del 95%. La línea de puntos y guiones de color azul representa el ajuste perfecto, la línea continua de color rosa es la línea de regresión y la línea discontinua de color gris representa los límites del intervalo de confianza de regresión del 95%.

5.1.3 Tipos de cambios nucleotídicos en la región del HDAg analizada

Los diversos cambios de nucleótidos presentes en las muestras de los 4 pacientes se contaron a medida que se registraron en los amplicones secuenciados (Figura 35) tal y como se explica en el apartado *4.18 Análisis del tipo de cambios nucleotídicos* de Material y Métodos. Después de excluir el sitio de edición, las transiciones, sustitución de una base púrica por otra base púrica, o una base pirimidínica por otra base pirimidínica (A:G, G:A, C:T y T:C), representaron el 63,09% de los cambios en la región analizada. En cambio, las transversiones, sustitución de una base púrica por una base pirimidínica o viceversa (C:G, G:C, C:A, A:C, T:A, A:T, T:G y G:T), representaron el 36,91% de los cambios en la región analizada. Curiosamente, entre los 4 posibles cambios de tipo transición, G:A y A:G (C:U y U:C en el ARN genómico del VHD circulante, que tiene una secuencia complementaria inversa en relación con nuestros amplicones) representaron el 73,86%, mientras que C:T y T:C (G:A y A:G en el ARN genómico del VHD circulante) estaban presentes en sólo el 26,17%.

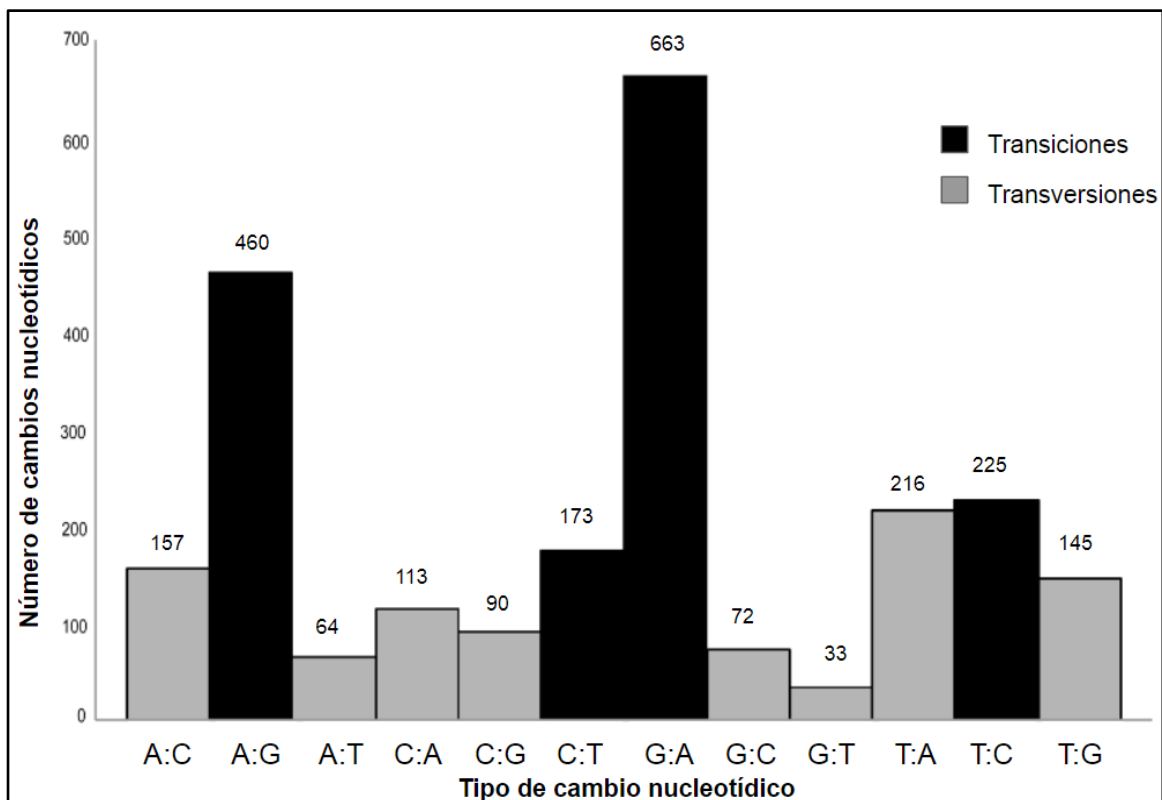


Figura 35. Cambios de nucleótidos de tipo transición (en negro) y de tipo transversión (gris) detectados en las muestras de los 4 pacientes, excluyendo el sitio de edición del HDAg (nucleótido 1012 dentro del codón 196).

5.2 Segundo estudio: Estudio longitudinal en pacientes con hepatitis crónica delta sobre la variabilidad y conservación de las regiones del HDAg y de la ribozima

En este estudio se analizó el grado de conservación o variabilidad de la quasiespecies del VHD en las regiones analizadas del HDAg y de la ribozima descritas en el apartado 4.5 *Amplificación del ARN del VHD para obtener librerías de amplicones para analizar por secuenciación masiva (NGS)*. Además se determinó la existencia de patrones de mutaciones a nivel aminoacídico en la región del HDAg.

Se trata de un estudio longitudinal retrospectivo de 12 pacientes con hepatitis crónica delta atendidos en el Hospital Universitario Vall d'Hebron (Barcelona, España), de los que se incluyeron 36 muestras de plasma (3 muestras por paciente, separadas por 3 años entre ellas) (Figura 36). Las características demográficas y clínicas más relevantes de los pacientes se muestran en la Tabla 16. Las muestras de estos pacientes se seleccionaron de acuerdo con los siguientes criterios de inclusión: mínimo de 6 años de seguimiento del paciente, ARN-VHD superior a 3 logUI/mL, VHD genotipo 1, ausencia de otras coinfecciones por el VHC o el VIH, ausencia de enfermedad hepática autoinmune y ausencia de tratamiento antiviral con PEG-IFN en el momento de recolección de la muestra y al menos cinco años antes. Nueve de los pacientes incluidos recibieron tratamiento contra el VHB basado en análogos de nucleós(t)idos a lo largo del seguimiento analizado.

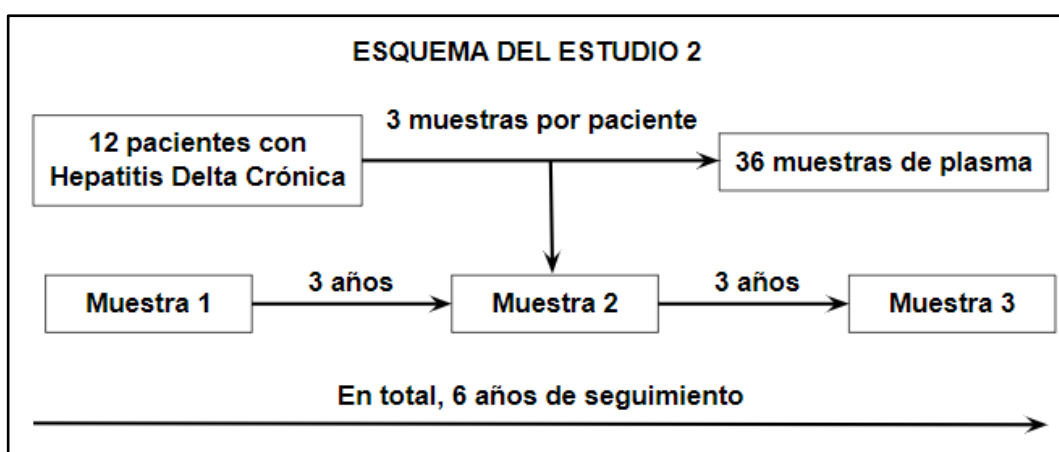


Figura 36. Esquema de los pacientes y las muestras incluidos en el estudio 2: Estudio longitudinal en pacientes con hepatitis crónica delta sobre la variabilidad y conservación de las regiones del HDAg y de la ribozima.

N Varones (%)	7 (58,33%)
Mediana de edad (años, IQR)	45,50 (38,75 - 47,25)
Etnia N caucásicos (%)	11 (91,66%)
Mediana de ALT (UI/L, IQR) Muestra 1 Muestra 2 Muestra 3	84,45 (51,75 - 112,50) 76,50 (55,50 - 107,75) 84,50 (52,50 - 155,75)
Mediana de ADN-VHB, logUI/mL (IQR)* Muestra 1 Muestra 2 Muestra 3	1,56 (1,30 - 2,12) 1,39 (1,30 - 2,04) 1,30 (1,30 - 1,45)
Mediana de ARN-VHD, logUI/mL (IQR) Muestra 1 Muestra 2 Muestra 3	5,64 (4,26 - 6,20) 4,91 (4,00 - 5,97) 5,01 (4,53 - 5,31)

Tabla 16. Características demográficas y clínicas de los pacientes incluidos en el estudio. IQR, rango intercuartílico. *Todas las muestras del ADN-VHB inferiores a 1,30 log UI/mL se han contabilizado como 1,30 logUI/mL y todas las muestras indetectables se han contabilizado como 0 logUI/mL.

5.2.1 Análisis de la conservación y variabilidad de la región del HDAG estudiada

Una vez aplicados los filtros bioinformáticos de calidad a las secuencias inicialmente obtenidas mediante NGS por la plataforma MiSeq, se han obtenido un total de 402.818 *reads* (mediana (IQR)= 9.091 (5.618,5 - 13.970) *reads*/muestra).

La conservación y variabilidad de la región del HDAG se estudió entre las posiciones nucleotídicas 910 y 1270 (que codifican los aminoácidos 111 a 214 del L-HDAG) a través de los haplotipos obtenidos en cada muestra, a nivel de nucleótidos y aminoácidos, aplicando un análisis *sliding window*, tal y como se explica en el apartado 4.19 *Análisis de la conservación y variabilidad de la región del HDAG y de la región de la ribozima de Material y Métodos*. En esta región analizada no se observaron ni inserciones ni deleciones.

Por un lado, en la Figura 37 se muestra el *sliding window* a nivel de nucleótidos de la región del HDAg analizada. Mediante este análisis se identificaron 2 regiones conservadas que corresponden a los nucleótidos 1122-1091 (Figura 38), que codifican los aminoácidos 160-169, y a los nucleótidos 994-970 (Figura 39), que codifican los aminoácidos 203-209. En estas regiones, la mayor parte de posiciones nucleotídicas tienen valores de contenido de información (CI) en torno a 2 bits (100% de conservación en secuencias de nucleótidos). Además, también se identificó la región más variable que corresponde a los nucleótidos de la posición 1048-1024 (Figura 40), que codifica los aminoácidos 185-191. El primer grupo de nucleótidos más conservado (1122-1091) y el grupo de nucleótidos más variable (1048-1024) corresponden a la región de unión a la ARN polimerasa II (ARN pol II) del HDAg (aminoácidos 150-195) y el segundo grupo de nucleótidos más conservado (994-970) pertenece a un dominio funcional del L-HDAg, la señal de ensamblaje viral (aminoácidos 195-214) que además contiene la señal de exportación nuclear (aminoácidos 198-210).^{41,80}

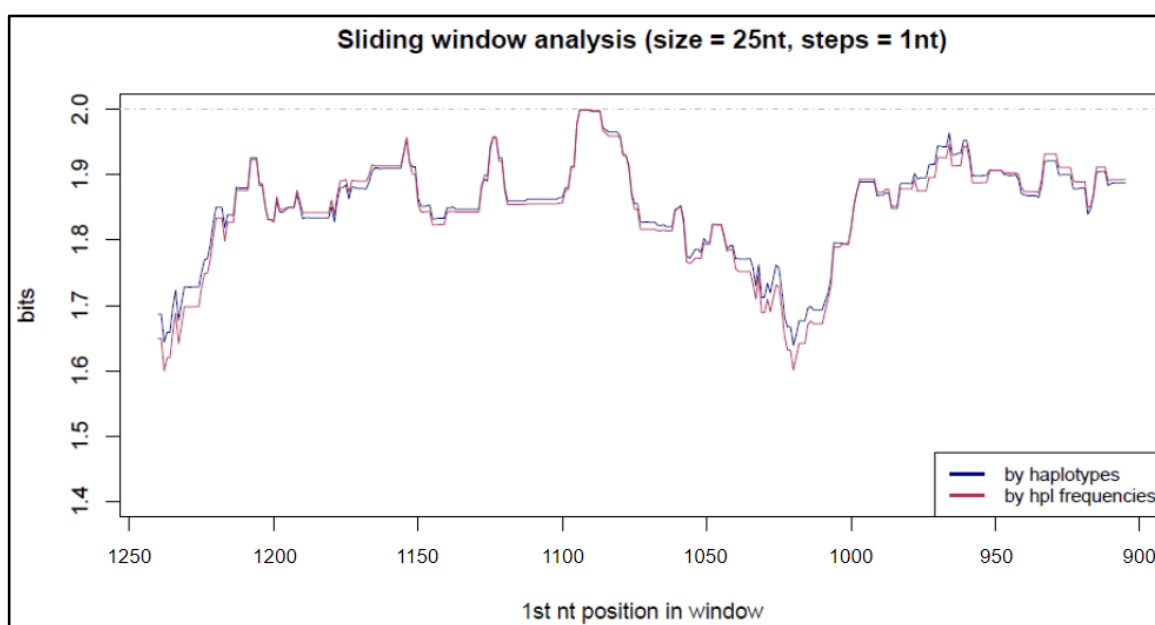


Figura 37. Análisis *sliding window* de nucleótidos entre las posiciones 910 y 1270 del genoma del VHD. La orientación de la secuencia se ha puesto en sentido antigenómico (la reversa complementaria de la genómica) porque es la secuencia codificante. Cada punto en el gráfico representa el contenido de información (CI) medio (en bits) de las ventanas de 25 nucleótidos, con desplazamiento de 1 nucleótido hacia adelante entre cada ventana. La línea azul muestra el análisis por haplotipos (by haplotypes) y la línea en rojo muestra el análisis por frecuencias de haplotipos (by hpl frequencies). La línea discontinua horizontal corresponde al máximo CI.

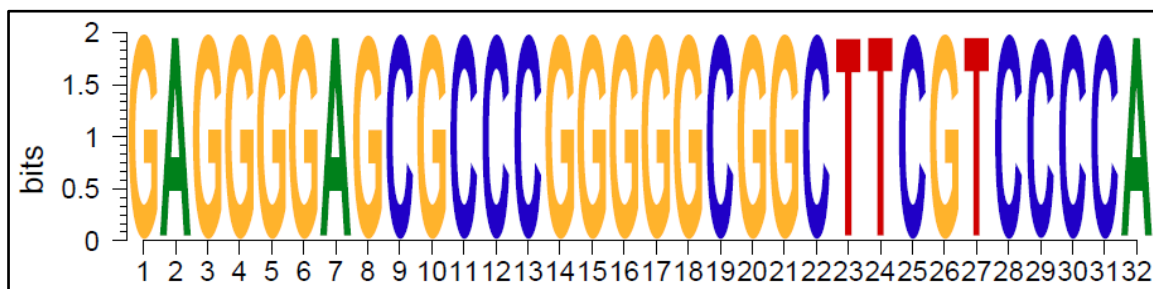


Figura 38. Representación en forma de logos de secuencia de nucleótidos que corresponde a una de las regiones conservadas que comprende de los nucleótidos 1122-1091 (aminoácidos del 160-169). La orientación de la secuencia se ha puesto en sentido antígenómico (la reversa complementaria de la genómica) porque es la secuencia codificante.

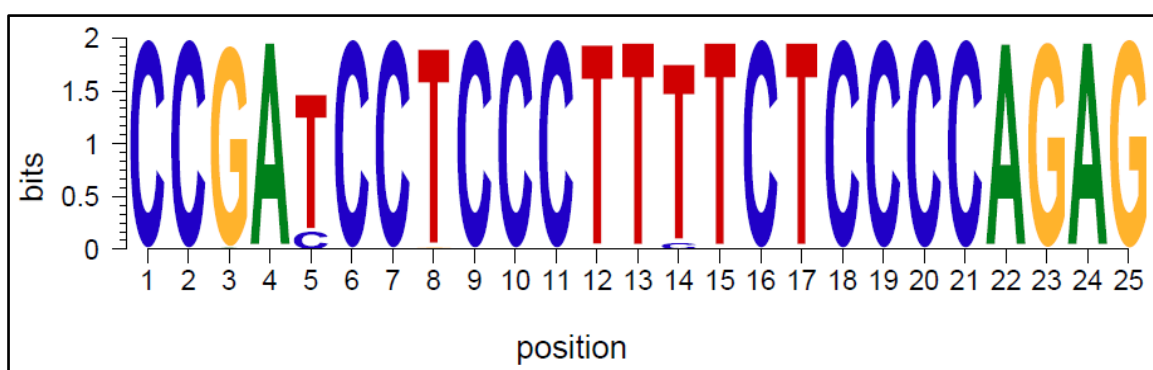


Figura 39. Representación en forma de logo de secuencia de nucleótidos que corresponde a una de las regiones más conservadas que comprende de los nucleótidos 994-970 (aminoácidos del 203-209). La orientación de la secuencia se ha puesto en sentido antígenómico (la reversa complementaria de la genómica) porque es la secuencia codificante.

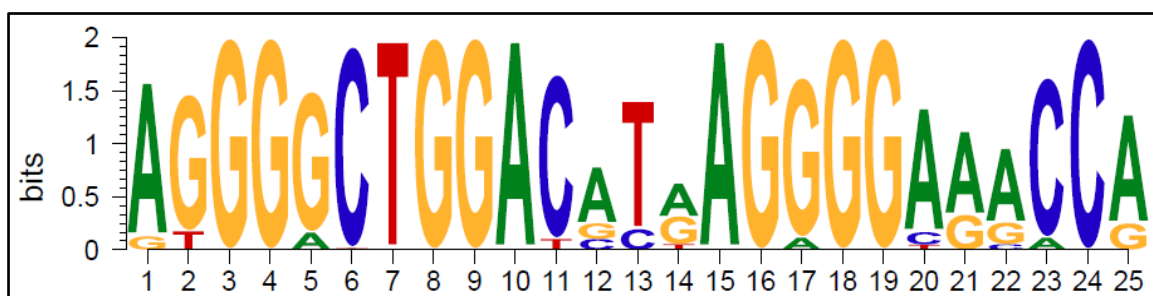


Figura 40. Representación en forma de logo de secuencia de nucleótidos que corresponde a la región más variable que comprende de los nucleótidos 1048-1024 (aminoácidos del 185-191). La orientación de la secuencia se ha puesto en sentido antígenómico (la reversa complementaria de la genómica) porque es la secuencia codificante.

Por otro lado, en la Figura 41 se muestra el *sliding window* a nivel de aminoácidos de la región del HDAg analizada. Mediante este análisis se identificaron 2 regiones conservadas que corresponden a los aminoácidos 160-169 (Figura 42), codificada por los nucleótidos 1121-1092, y a los aminoácidos 199-212 (Figura 43), codificada por los nucleótidos 1004-963, en la que la mayor parte de posiciones aminoacídicas tienen

valores de CI en torno a 4,32 bits (100% de conservación de las secuencias de aminoácidos). Ambas regiones de aminoácidos, por lo tanto, coinciden con las codificadas por las regiones más conservadas a nivel de nucleótidos. Además, también se identificó la región más variable que corresponde a los aminoácidos 112-121 (Figura 44), codificados por los nucleótidos 1265-1236, que no coinciden con la región más variable a nivel de nucleótidos (1048-1024). Como se ha comentado antes, los aminoácidos 160-169 pertenecen a la región de unión de la ARN polimerasa II (aminoácidos 150-195), mientras que los aminoácidos 199-212 pertenecen a la señal de ensamblaje viral del L-HDAg (aminoácidos 195-214) que además contiene la señal de exportación nuclear (aminoácidos 198-210). Por último, el grupo de aminoácidos más variable (112-121) pertenece a la región *helix-loop-helix* del HDAg (aminoácidos 108-137).^{41,80}

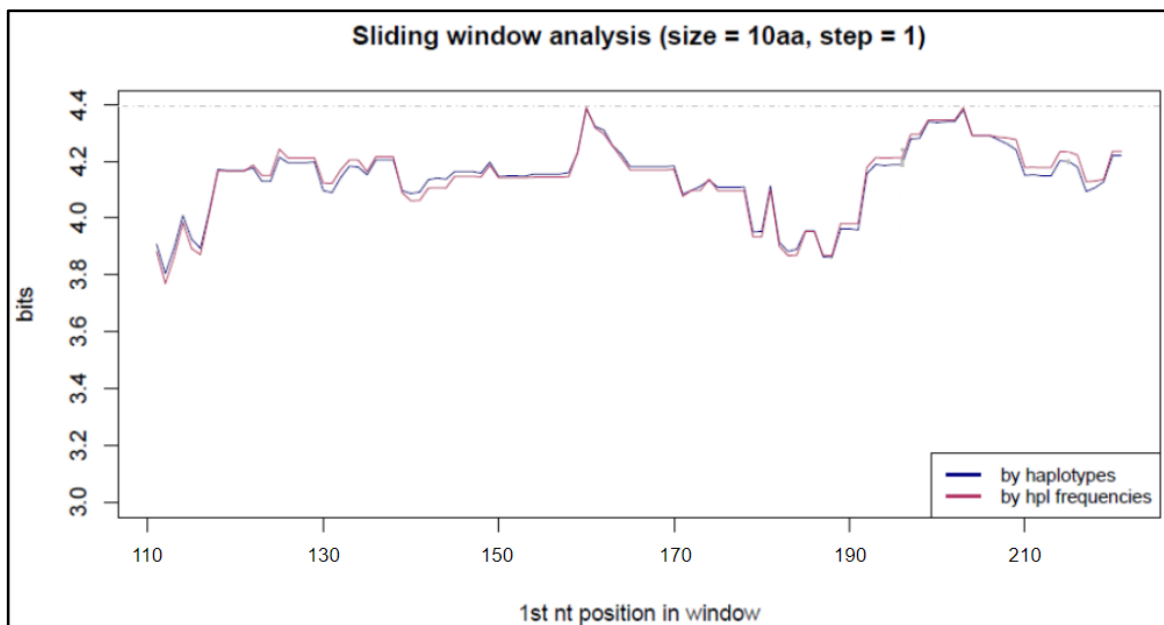


Figura 41. Análisis *sliding window* de aminoácidos de la región del HDAg analizada (posiciones 111 a 214). Cada punto en el gráfico representa el contenido de información (CI) medio (en bits) de las ventanas de 10 aminoácidos, con desplazamiento de 1 aminoácido hacia adelante entre cada ventana. La línea azul muestra el análisis por haplotipos (by haplotypes) y la línea en rojo muestra el análisis por frecuencias de haplotipos (by hpl frequencies). La línea discontinua horizontal corresponde al máximo CI.

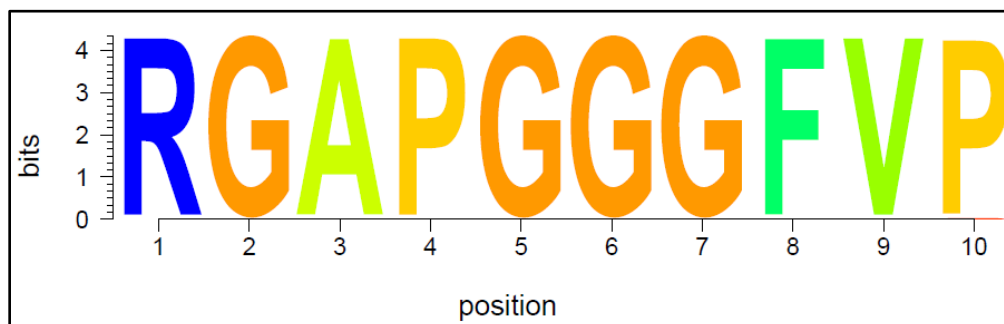


Figura 42. Representación en forma de logos de secuencia de aminoácidos que corresponde a una de las regiones conservadas que comprende de los aminoácidos 160-169 (nucleótidos de la posición 1121-1092).

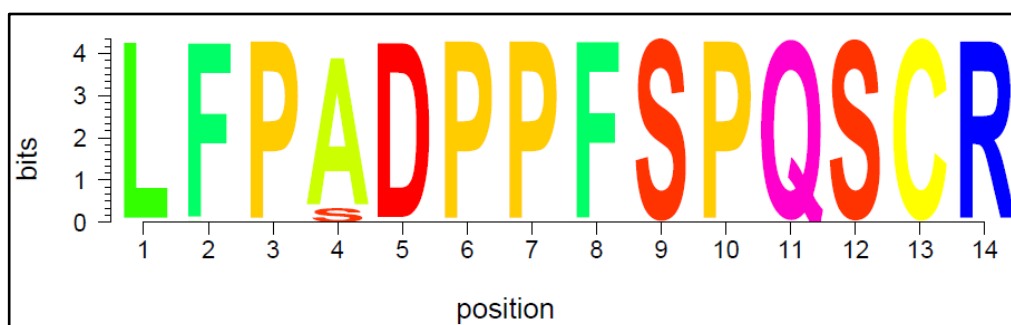


Figura 43. Representación en forma de logos de secuencia de aminoácidos que corresponde a una de las regiones conservadas que comprende de los aminoácidos 199-212 (nucleótidos de la posición 1004-963).

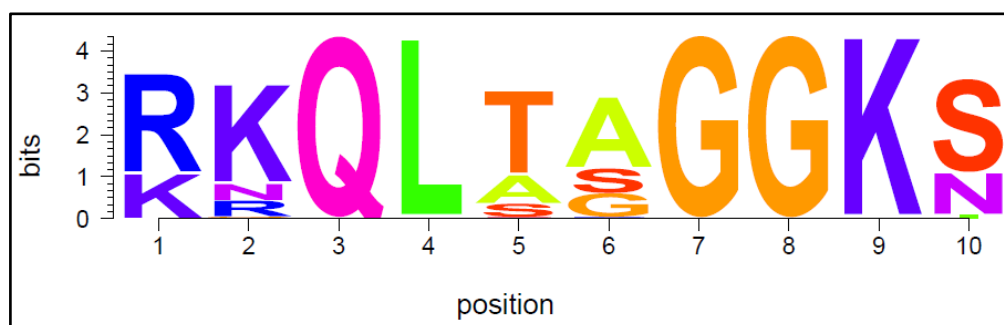


Figura 44. Representación en forma de logo de secuencia de aminoácidos que corresponde a la región más variable que comprende de los aminoácidos 112-121 (nucleótidos de la posición 1265-1236).

5.2.2 Estudio de los patrones de mutaciones de la región del HDAg analizada

Mediante el análisis descrito en el apartado 4.20 *Estudio de los patrones de cambios de aminoácidos de la región del HDAg* se identificaron 98 mutaciones aminoacídicas diferentes en proporciones superiores al 0,25% de la quasiespecies, que afectan a 54 posiciones en la región del HDAg analizada (entre las posiciones aminoacídicas 111 y 214) (Tabla 17).

Posición	Mutaciones	Posición	Mutaciones	Posición	Mutaciones
K111	K111R	V144	V144I	T180	T180V; T180A; T180S
R112	R112K	P148	P148Q; P148R; P148S	T182	T182H
K113	K113R;K113E; K113N; K113G	V149	V149I; V149T; V149A; V149P;	G183	G183A
Q114	Q114R	G150	G150R	E184	E184D; E184G
S116	S116A;S116G; S116T; S116E	N153	N153I	V188	V188I; V188T; V188L; V188P
A117	A117G; A117V; A117S; A117R	G157	G157C	G190	G190R
N121	N121S; N121I	G158	G158A	D191	D191G; D191A; D191T; D191N; D191S; D191E
K124	K124R	S159	S159A; S159Q; S159T	Q192	Q192R
E125	E125D ; E125G	A162	A162T	F194	F194I; F194L
E128	E128G	F167	F167L	D197	D197E
K131	K131R; K131G	P169	P169S	I198	I198L; I198M
R132	R132K	S170	S170N	F200	F200L
E135	E135K; E135G; E135R	M171	M171L	A202	A202S
R139	R139K; R139E;	Q172	Q172L; Q172H	D203	D203N
R140	R140G	G173	G173R; G173D; G173E	F206	F206V
E141	E141G; E141K	V174	V174I; V174A	Q209	Q209R
R142	R142G	S177	S177L	S210	S210G
R143	R143G	F179	F179L; F179Y	P213	P213T; P213N; P213A; P213I

Tabla 17. Posiciones afectadas por mutaciones en la región amplificada del HDAg (aminoácidos 111-214) y mutaciones detectadas en ellas en proporciones superiores al 0,25% de la quasiespecies en alguna de las muestras analizadas. La posición 196 (codón *editing*) solo mostró la variante wild-type (stop) y la editada (aminoácido W).

De estas 98 mutaciones se seleccionaron 19 según la prueba de Spearman, ya que tenían una rho $>0,8$ y un p-valor $<0,05$. Seguidamente, se identificaron diferentes patrones de mutaciones aminoacídicas. En la Figura 45 se observa un dendograma con los patrones con más de 2 mutaciones aminoacídicas:

- Patrón 1: E141K, R139E, V149P, S159Q, T182H, E184D, A202S. En algunos casos se le puede sumar la mutación E135G y después la mutación S170N. Este patrón de mutaciones está presente en 1 de los 12 (8,3%) pacientes incluidos en el estudio.
- Patrón 2: P213T y S159A. En una elevada proporción de casos se le suma la mutación G158A. Este patrón de mutaciones está presente en 7 de los 12 (58,3%) pacientes incluidos en el estudio.

En la Figura 46 se observa un dendograma con los patrones de mutaciones que comprenden parejas de mutaciones aminoacídicas. Se identificaron 3 patrones de mutaciones:

- Patrón 3: A117G y Q192R, presente en 3 pacientes de los 12 (25%) incluidos en el estudio.
- Patrón 4: R139K y Q172L, presente en 2 paciente de los 12 (16,7%) incluidos en el estudio.
- Patrón 5: R132K y I198L, presente en 2 pacientes de los 12 (16,7%) incluidos en el estudio.

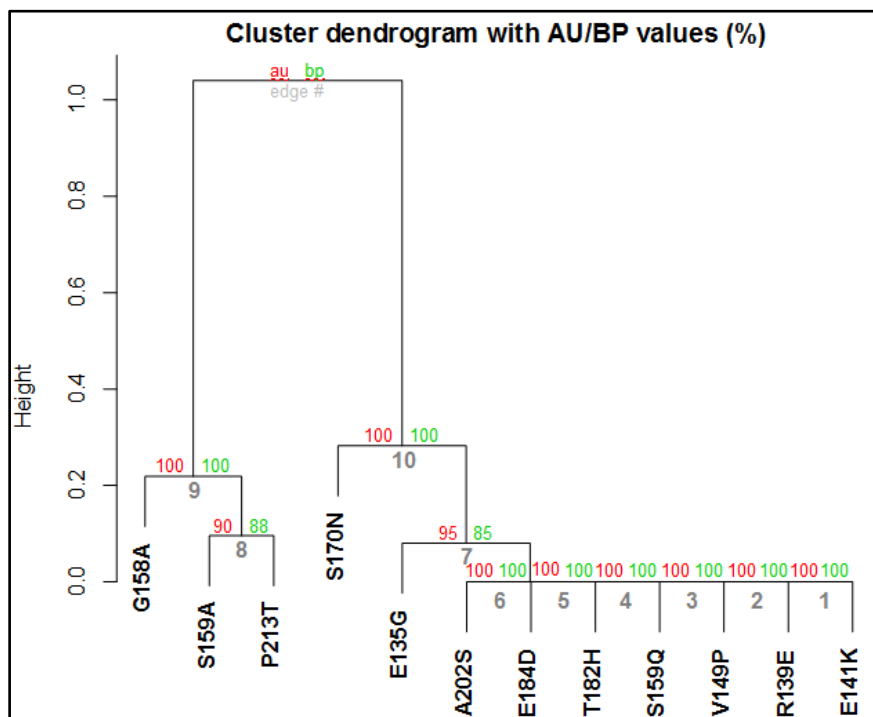


Figura 45. Dendrograma de los patrones de mutaciones con más de 2 mutaciones aminoacídicas. El orden en el que se han ido agrupando entre sí las parejas de mutaciones en los distintos *clusters* se indica con los números en color gris. El valor del *Approximately Unbiased P-value* (AU, 100 - p-valor del *cluster*) se muestra con el número en color rojo y el valor de *Bootstrap Probability value* (BP, frecuencia de aparición del *cluster* en diferentes réplicas del *bootstrap*) se muestra con el número en color verde.

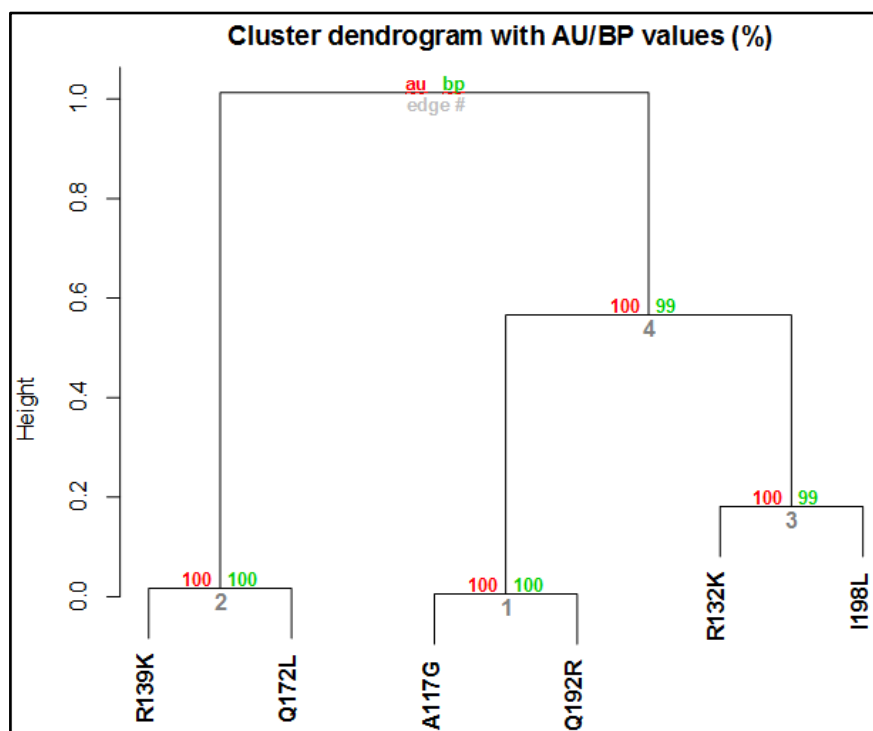


Figura 46. Dendrograma de los patrones de mutaciones que comprenden 2 mutaciones aminoacídicas. El orden en el que se han ido agrupando entre sí las parejas de mutaciones en los distintos *clusters* se indica con los números en color gris. El valor del *Approximately Unbiased P-value* (AU, 100 - p-valor del *cluster*) se muestra con el número en color rojo y el valor de *Bootstrap Probability value* (BP, frecuencia aparición del *cluster* en diferentes réplicas del *bootstrap*) se muestra con el número en color verde.

En la Figura 47, se muestran los gráficos de la evolución de las frecuencias relativa de cada patrón a lo largo del seguimiento de los pacientes en los que se han detectado.

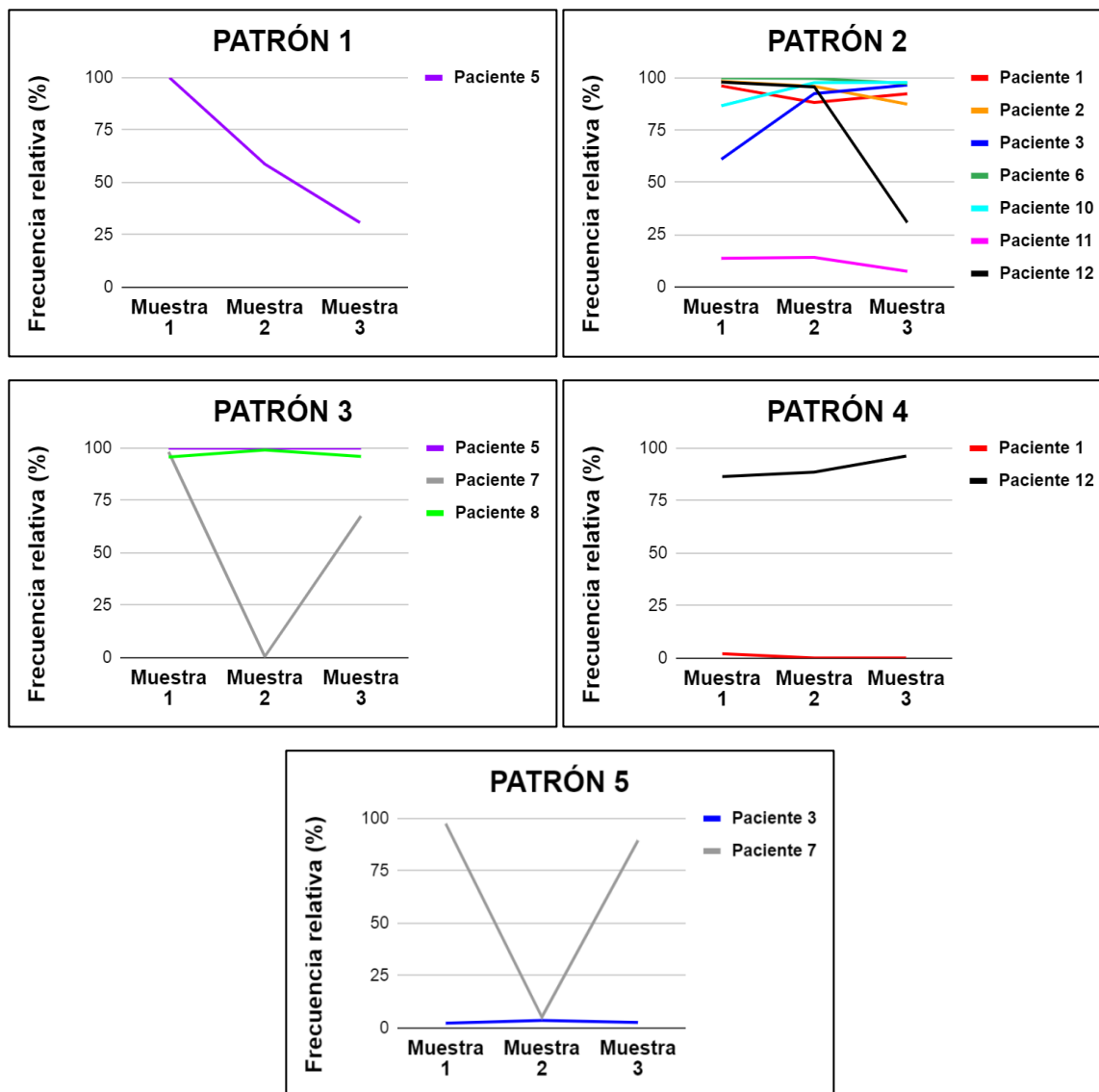


Figura 47. Gráficos de cada patrón de mutaciones. Cada paciente se representa con un color diferente. En el eje x se representan las muestras de cada paciente que contiene el patrón de mutaciones y en el eje y se muestra la frecuencia relativa (%) en que está presente el patrón de mutaciones en cada muestra de cada paciente.

En la mayoría de pacientes las frecuencias de los patrones de mutaciones se mantienen estables a lo largo del período de seguimiento de 6 años analizado. Sin embargo, en el patrón de mutaciones 1, presente solamente en el paciente 5, se observa que la frecuencia relativa disminuye a lo largo del tiempo. Además, en el patrón de mutaciones 2, el paciente 12, la frecuencia relativa se mantiene elevada en la primera y segunda muestra pero disminuye considerablemente de la segunda a la tercera muestra. En

cambio, en este mismo patrón, en el paciente 3 la frecuencia relativa aumenta entre la primera muestra y la segunda. Por último, en el paciente 7, el cual muestra el patrón de mutaciones 3 y el 5, la frecuencia relativa fluctúa a lo largo del tiempo: en la primera muestra estos patrones muestran una frecuencia muy elevada, en la segunda muestra el patrón prácticamente desaparece mientras que en la tercera muestra vuelve a seleccionarse intensamente y sus proporciones aumentan considerablemente. A destacar también los pacientes 1 y 12 que contienen los patrones de mutaciones 2 y 4. En el paciente 12, en cuanto disminuye la frecuencia relativa del patrón 2, aumenta la frecuencia relativa del patrón 4. Y, en el paciente 1, a lo largo de todo el seguimiento, el patrón 2 se muestra en frecuencias muy altas, en cambio, el patrón 4 se muestra en frecuencias muy bajas. Por último, mencionar que la mayoría de pacientes contienen el patrón 2, y en la mayoría de los pacientes está presente en proporciones superiores al 75%.

Una vez identificados los patrones de mutaciones se analiza dónde se localizaba cada mutación de cada patrón. En el patrón 1, la mutación E135G pertenece a la región *helix-loop-helix* (108-137);²⁵¹⁻²⁵³ las mutaciones R139E, E141K y V149P pertenecen a un dominio de unión a ARN (97-146);²⁵⁴ las mutaciones S159Q, S170N, T182H y E184D pertenecen a la región de unión a la ARN pol II (150-195)^{80,81} y la mutación A202S pertenece a la señal de empaquetamiento viral (195-214).²⁵¹ Además, algunas mutaciones se encuentran dentro de epítomos: E135G, R139E, E141K, V149P, S159Q, S170N, T182H y E184D se encuentran en un epítipo inmunogénico de células T CD8+²⁵¹ y S159Q, S170N, T182H y E184D se encuentran en un epítipo de células B.¹⁰⁹ En el patrón 2, las mutaciones G158A y S159A pertenecen a la región de unión a la ARN pol II^{80,81} y la mutación P213T a la señal de empaquetamiento viral y a la señal de isoprenilación (211-214).²⁵¹ Además, las mutaciones G158A y S159A pertenecen a un epítipo inmunogénico de células T CD8+ y de un epítipo de células B.^{109,251} En el patrón 3, la mutación A117G pertenece a la región *helix-loop-helix*²⁵¹⁻²⁵³ y la mutación Q192R a la región de unión a la ARN pol II.^{80,81} Además, la mutación A117G está dentro del epítipo de células restringido al complejo mayor de la histocompatibilidad II y dentro de un *cytotoxic T ligand epitope*¹⁰⁹ y la mutación Q192R está dentro de un epítipo de células B y dentro de un epítipo inmunogénico de células T CD8+.^{109,251} En el patrón 4, la mutación R139K pertenecen a un dominio de unión a ARN²⁵⁴ y la mutación Q172L a la región de unión a la ARN pol II.^{80,81} Además, estas dos mutaciones están dentro de un epítipo inmunogénico de células T CD8+²⁵¹ y la mutación Q172L también está dentro de un epítipo de células B.^{109,251} En el patrón 5, la mutación R132K pertenece a la región

*helix-loop-helix*²⁵¹⁻²⁵³ y la mutación I198L a la señal de empaquetamiento viral y a la señal de exportación nuclear (198-210). Además, la mutación R132K pertenece a un epítipo inmunogénico de células T CD8+.²⁵¹

5.2.3 Análisis de la conservación y variabilidad de la región de la ribozima estudiada

Una vez aplicados los filtros bioinformáticos de calidad a las secuencias inicialmente obtenidas mediante NGS por la plataforma MiSeq se han obtenido un total de 450.186 *reads* (mediana (IQR)= 10.818 (5.884 - 16.510,5) *reads*/muestra).

La conservación de la región de la ribozima se estudió entre las posiciones nucleotídicas 688 y 771 a través de los haplotipos obtenidos en cada muestra solamente a nivel de nucleótidos (ya que se trata de una región no codificante), aplicando un análisis *sliding window* (Figura 48), tal y como se explica en el apartado 4.19 *Análisis de la conservación y variabilidad de la región del HDAg y de la región de la ribozima* de Material y Métodos. Mediante este análisis se observó que esta región es muy conservada, ya que, la mayoría de las posiciones nucleotídicas están en torno a 2 bits (100% de conservación en secuencias de nucleótidos), tal como muestra el logo de toda la secuencia de la ribozima (Figura 49). En este logo, se puede observar que las partes más variables, que corresponden a las posiciones nucleotídicas 746-757 (Figura 50), pertenecen a un nucleótido del bucle L4 y a una parte de la región de doble cadena P4. Por otro lado la región más conservada se identificó entre las posiciones nucleotídicas 715-745 (Figura 51), que pertenecen a algunos nucleótidos de las regiones de doble cadena P3, P1 y P1.1, a la región monocatenaria J1/4, a la región de doble cadena P4 y al bucle L4.⁹⁷

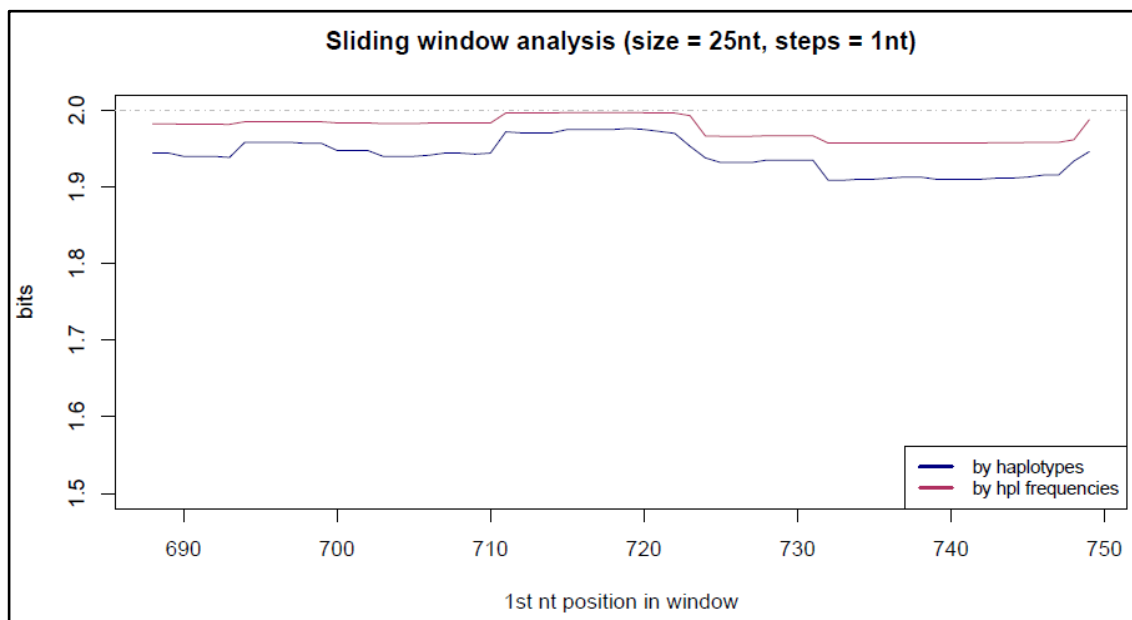


Figura 48. Análisis *sliding window* de la región de la ribozima (entre las posiciones 688 y 771). La orientación de la secuencia se ha puesto en sentido genómico. Cada punto en el gráfico representa el CI medio (en bits) de las ventanas de 25 nucleótidos, con desplazamiento de 1 nucleótido hacia adelante entre cada ventana. La línea azul muestra el análisis por haplotipos (by haplotypes) y la línea en rojo muestra el análisis por frecuencias de haplotipos (by hpl frequencies). La línea discontinua horizontal corresponde al máximo CI

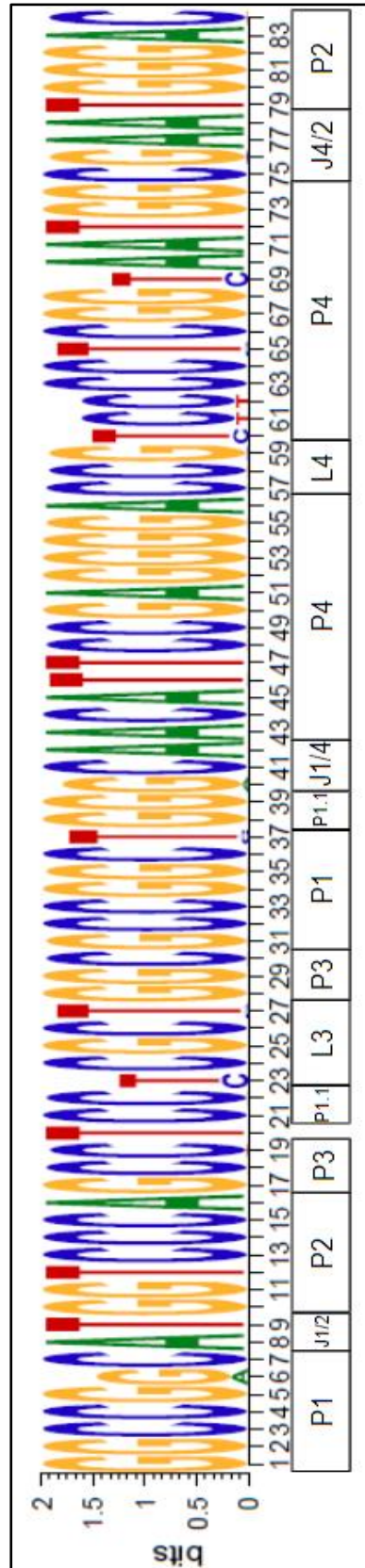


Figura 49. Representación en forma de logo de la secuencia de nucleótidos que corresponde a toda la región de la ribozima. En la parte inferior se indica a qué región de la ribozima pertenece cada grupo de nucleótidos. La orientación de la secuencia se ha puesto en sentido genómico.⁹⁷

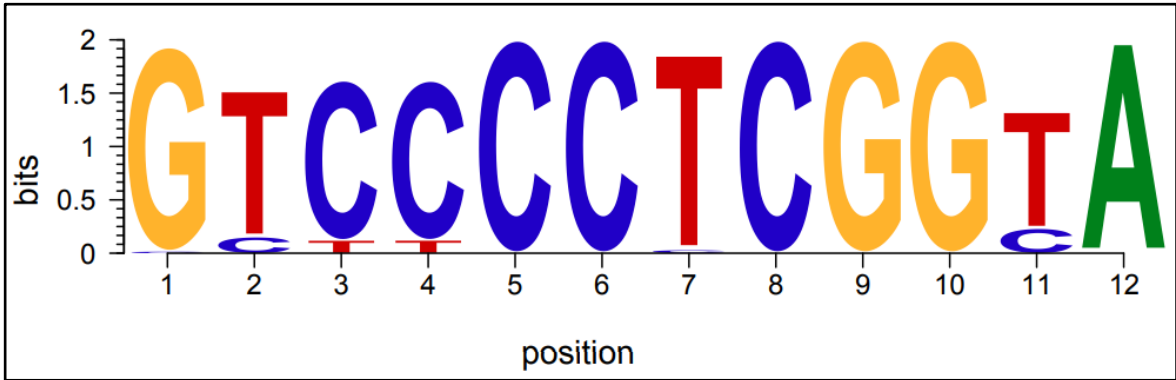


Figura 50. Representación en forma de logos de secuencia de nucleótidos que corresponde a la región más variable que comprende de las posiciones nucleotídicas 746-757. La orientación de la secuencia se ha puesto en sentido genómico, ya que es la secuencia circulante que se ha secuenciado.

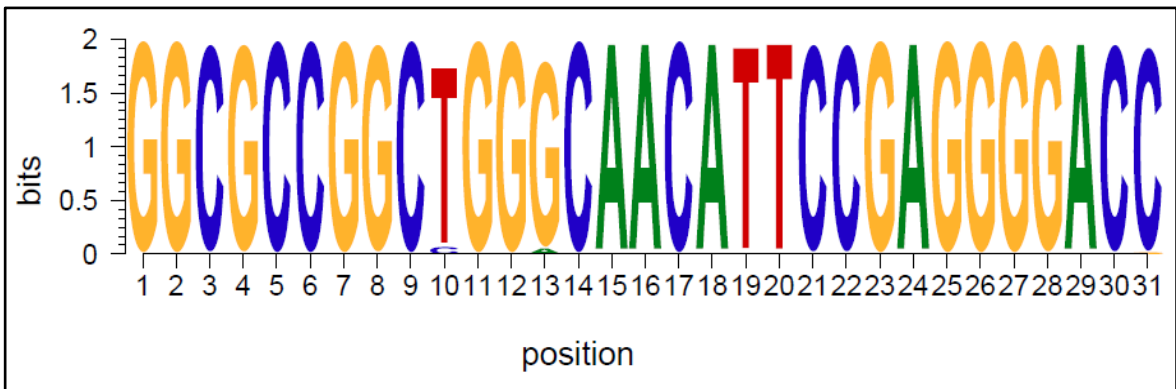


Figura 51. Representación en forma de logos de secuencia de nucleótidos que corresponde a la región más conservada que comprende de las posiciones nucleotídicas 715-745. La orientación de la secuencia se ha puesto en sentido genómico, ya que es la secuencia circulante que se ha secuenciado.

DISCUSIÓN

6. DISCUSIÓN

El virus de la hepatitis delta (VHD) es un virus de ARN de una sola hebra, circular, de alrededor de 1700 nucleótidos y de polaridad negativa. Contiene una región con actividad ribozima, esencial en su replicación, y un único marco de lectura abierto que codifica el antígeno delta (HDAg). El HDAg por la acción editora del enzima del huésped ADAR-1 da lugar a dos formas del mismo, una isoforma de menor longitud (S-HDAg), con un codón stop (TAG en los nucleótidos 1011-1013, que codifican el aminoácido 196), que es la no editada, y una más larga (L-HDAg) que es el resultado de la edición en el nucleótido 1012 (cambio de A:G) manteniéndose la traducción hasta un codón stop altamente conservado en la posición 215. Este virus comparte características con los viroides, en este sentido la molécula de ARN genómico viral puede plegarse sobre si misma dando lugar a una estructura secundaria no ramificada altamente estable, prácticamente constituyendo un genoma de ARN de doble cadena cerrado por los extremos.^{5,32}

El genoma del VHD no codifica ninguna polimerasa propia, por lo que su estructura en quasiespecies no se puede explicar por la acumulación de errores en el proceso de replicación catalizado por este tipo de enzimas y por la falta de capacidad de corrección de errores de las mismas. La replicación del VHD sigue el mecanismo conocido como replicación en círculo rodante (análogo a los viroides) y se realiza mediante las ARN polimerasas del huésped, principalmente la ARN polimerasa II (ARN pol II). La isoforma corta del antígeno delta (S-HDAg) se une a esta polimerasa, propiciando su cambio de molde de ADN a ARN a coste de su pérdida de fidelidad.⁸¹ Además, es posible que las proteínas de edición genética del huésped también puedan modificar el genoma del VHD, más allá de la edición del nucleótido 1012 mediante ADAR-1.⁶⁵ Todo esto junto con una elevada tasa replicativa, parece explicar el espectro de diferentes mutantes en un mismo individuo conocidos como quasiespecies virales.²⁵⁵

La distribución de las poblaciones de mutantes en la quasiespecies es dinámica y se ve influenciada por el ambiente (respuesta inmune, alteraciones metabólicas del huésped...) o las intervenciones externas como los tratamientos antivirales. Además, la distribución poblacional también se ve influenciada por fenómenos de deriva genética, sobre todo durante la transmisión del virus. Todos estos procesos conducen a las formas génicas dominantes que se observan durante los cursos clínicos de la hepatitis delta crónica y pueden desempeñar un papel importante en el escape al sistema inmune del huésped o en la recaída clínica.^{100,109} Sin embargo, existen pocos estudios sobre la quasiespecies

del VHD y por lo tanto todavía no se dispone de datos muy concluyentes sobre sus implicaciones. Por ello, el objetivo principal de esta tesis ha sido estudiar la quasiespecies del VHD mediante secuenciación masiva (NGS), cuantificando la edición de su genoma, que da lugar a los antígenos S-HDAg y L-HDAg, y la conservación o variabilidad de las regiones funcionales del virus, correspondientes al HDAg y a la ribozima que es el único dominio de este virus con actividad enzimática propia y esencial para la replicación viral.

6.1 Métodos de estudio de la quasiespecies

Aunque la secuenciación por el método de Sanger tiene la capacidad de determinar la secuencia de fragmentos de ADN entre 300-900 nucleótidos,²⁵⁶ no permite determinar la secuencia de nucleótidos de haplotipos individuales, es decir, secuencias únicas que en la quasiespecies se agrupan en secuencias iguales entre sí en el conjunto de secuencias obtenidas. En lugar de eso, la secuencia obtenida está compuesta en cada posición por la media de los haplotipos presentes en la quasiespecies en una proporción $\geq 15-20\%$.²⁴² Por lo tanto esta técnica da una información limitada sobre las quasiespecies virales. Hasta hace pocos años, el método tradicional utilizado para analizar secuencias individuales, en un entorno de heterogeneidad poblacional genómica como es la quasiespecies viral, era la clonación molecular y su posterior secuenciación Sanger.²⁵⁷ Este método se ha utilizado como método de referencia para estudiar poblaciones virales, ya que sí permite estudiar haplotipos individuales, pero tiene algunos inconvenientes. Es una técnica costosa (tanto económicamente como en tiempo) y laboriosa que requiere de múltiples experimentos y complejos pasos, que acaban proporcionando una resolución limitada en lo que a espectro de mutaciones y frecuencias se refiere. Para obtener una representación significativa de una quasiespecies viral se necesita una gran cantidad de clones (> 100) lo que significativamente dificulta su uso.²⁵⁸ También, puede haber errores introducidos por las polimerasas o durante la reacción de secuenciación Sanger que pueden artefactuar los resultados.¹⁹⁷ En cambio, la NGS ha supuesto una importante mejora en el estudio de las quasiespecies virales ya que tiene la ventaja de tener la capacidad de analizar miles de haplotipos individuales en paralelo en un solo experimento de secuenciación. Eso le permite tener una considerable sensibilidad para detectar poblaciones minoritarias, en menos tiempo y a un coste económico inferior por secuencia obtenida que la clonación y posterior secuenciación Sanger^{198,199,214,258-260} A pesar de esto, la precisión y sensibilidad de la NGS para detectar

poblaciones del VHD no ha sido analizada hasta el momento, por este motivo, es necesario establecer unos controles (mezclas de clones).

6.1.1 Cuantificación de la edición del codón stop del HDAG en el genoma del VHD mediante NGS

En el primer estudio, se ha comparado el método de secuenciación Sanger de los productos de la clonación con el método de NGS. Para ello, se utilizan muestras de 4 pacientes que cumplen con los criterios de inclusión. En estos 4 pacientes se analiza el codón 196 y se considera que su edición es un evento binario (CUA o CCA, traducido a TAG o TGG, respectivamente), con frecuencias que exhiben una distribución binomial y, por lo tanto, son complementarias. La comparación de los intervalos de confianza al 95% de las frecuencias de codones no editados (CUA traducido a TAG, codón stop, que codifica el S-HDAG) demostró que los resultados obtenidos por ambos métodos son comparables y evidenció que la NGS es más precisa en la cuantificación de variantes que la secuenciación Sanger de los productos de la clonación, como se esperaba debido a su tamaño de muestra mucho mayor (se analizaron secuencias de decenas de clones frente a un mínimo de 2.085 *reads* por paciente). Por lo tanto, se puede concluir que la NGS sería un método útil para detectar proporciones de genomas del VHD no editados y editados al analizar muestras clínicas.

En los pacientes incluidos se observó un predominio de genomas no editados en las quasiespecies del VHD circulantes (una media del 64% de genomas no editados detectados mediante NGS y una media del 60% de genomas no editados detectados mediante secuenciación Sanger de los productos de la clonación). En un estudio longitudinal anterior de nuestro grupo, Homs *et al.*,¹⁰⁴ también se observaba este predominio de los genomas no editados (media del 67% de genomas no editados detectados mediante NGS). Este es un hallazgo razonable, ya que los genomas no editados codifican el S-HDAG, que se requiere para la replicación del ARN-VHD, mientras que los genomas editados codifican el L-HDAG, que inhibe la replicación.^{57,78} Sin embargo, se observó un porcentaje significativo de genomas editados (media del 36% de genomas editados detectados mediante NGS y media del 40% de genomas editados detectados mediante secuenciación Sanger de los productos de la clonación), parecido al observado en el estudio de Homs *et al.*¹⁰⁴ (media del 34% de genomas editados detectados mediante NGS). El considerable número de genomas editados plantea la cuestión de su papel en las quasiespecies del VHD circulantes, especialmente porque estos genomas no son capaces de replicarse.²⁶¹ Tanto el S-HDAG como el L-HDAG son esenciales para el ciclo de replicación viral,^{57,78} por lo tanto, los genomas editados

circulantes podrían servir para proporcionar el L-HDAg al ciclo viral antes de la estimulación de la expresión de la ADAR-1 del huésped, lo que haría posible el empaquetamiento de nuevos viriones del VHD al comienzo de la infección celular (ambos antígenos son necesarios para establecer la ribonucleoproteína) y mejoraría la propagación de la infección.⁷⁸ Por lo tanto, los genomas editados podrían tener un papel en el establecimiento y mantenimiento de la infección crónica, una hipótesis que requiere confirmación en futuros estudios. Además, esto sugiere que podría ser importante ajustar la carga viral de acuerdo con el porcentaje de genomas no editados (ya que son los que se replican eficientemente) al monitorizar pacientes con hepatitis delta crónica. Disponer de una técnica para cuantificar de manera fiable los genomas no editados y editados abre la puerta a futuros estudios para probar esta hipótesis.

En la comparación de la secuenciación Sanger de clones con la NGS, se ha demostrado que esta última es mucho más precisa para cuantificar las proporciones de haplotipos editados y no editados en la quasiespecies. Sin embargo, la precisión a la que la NGS es capaz de cuantificar estos haplotipos cuando están presentes en diferentes proporciones aún no se ha definido. Por tanto, para determinar la precisión de la NGS a diferentes proporciones de las poblaciones virales en las quasiespecies, se mezclaron productos de PCR derivados de clones editados y clones no editados en proporciones del 5% al 95% y se cuantificaron por NGS. Nuevamente, suponiendo una distribución binomial de frecuencias de codones editados y no editados, se compararon los porcentajes de las poblaciones no editadas detectadas en cada mezcla artificial con las esperadas en cada mezcla, y se observó una buena correlación entre ellas. Sin embargo, el análisis de las distancias entre las proporciones observadas y esperadas en cada mezcla de clones evidencia una tendencia de la NGS a subestimar ligeramente los porcentajes de codones no editados, ya que la línea de regresión estaba ligeramente por debajo del ajuste perfecto. Esta tendencia probablemente se deba a un sesgo de muy bajo nivel de la metodología de NGS. Así pues, los resultados de este primer estudio han mostrado la capacidad de la NGS para cuantificar con un buen nivel de precisión mezclas con proporciones entre el 5% y el 95%.

6.2 Análisis del patrón de cambios de nucleótidos

En el primer estudio, después de la validación de la cuantificación de las proporciones de genomas editados y no editados, se utilizó la NGS para analizar el fragmento del genoma del VHD amplificado (nucleótidos 910-1270), buscando evidencia de una tendencia a favor o en contra de un cambio nucleotídico particular o un tipo determinado de cambio

nucleotídico (transición o transversión). Los resultados obtenidos en los 4 pacientes incluidos en este estudio confirmaron los obtenidos en un estudio anterior de Homs *et al.*¹⁰⁴ Las transiciones representaron el 63% de los cambios nucleotídicos, una vez excluido el codón de edición y, curiosamente, los cambios de A:G o G:A (U:C o C:U en el ARN genómico) representaron el 74% del total de los cambios de tipo transición. Estas transiciones están asociadas con un patrón de desaminación: la desaminación en A:G (U:C) puede estar relacionada con la actividad de ADAR-1, similar al fenómeno de edición específico en el codón 196,⁷⁰ mientras que la desaminación en G:A (C:U) puede estar asociada con ciertos miembros de la familia de las enzimas catalíticas similares a polipéptidos de edición de ARNm apolipoproteína B (APOBEC).²⁶² Por lo tanto, nuestros resultados indican que la desaminación mediada por las actividades enzimáticas del huésped mediada por ADAR-1 y tal vez, ciertos miembros de la familia de APOBEC, además de su papel esencial en la edición del genoma en el codón 196, puede ser una fuente importante de acumulación de variabilidad en el genoma del VHD. Esta noción parece estar de acuerdo con los hallazgos de Giersch *et al.*¹⁰² en un modelo de infección del VHD en ratones humanizados, donde se observó un aumento significativo en los niveles de expresión de ciertos miembros de la familia APOBEC (por ejemplo, hAPOBEC3B y hAPOBEC3G) en ratones coinfectados por VHB-VHD frente a ratones mono infectados por VHB. Sin embargo, aún no se ha demostrado la edición del genoma del VHD por estas enzimas mutagénicas. Se ha reportado que la edición del genoma del VHD por ADAR-1 en sitios diferentes al nucleótido 1012 no parece ser significativa, de hecho se ha reportado que la probabilidad de la edición en esta posición es 600 veces más eficiente que en las otras 337 adenosinas del genoma.²⁶³ Por eso la posible edición del genoma viral por enzimas mutagénicas debería ser confirmada en modelos de infección *in vivo*.

6.3 Conservación y variabilidad en la quasiespecies viral en la región codificante del HDAg y la ribozima

En el segundo estudio, mediante la NGS se analizó el grado de conservación o variabilidad de la quasiespecies del VHD en las regiones analizadas del HDAg y de la ribozima. Además se determinó la existencia de patrones de mutaciones, a nivel aminoacídico, en la región del HDAg.

6.3.1 Análisis de la conservación y variabilidad en los aminoácidos del HDAg y en su región codificante

En primer lugar, el grado de conservación o variabilidad de la región del HDAg amplificada se determinó a nivel de nucleótidos y de aminoácidos. En ninguna de las muestra se observaron ni inserciones ni deleciones, de acuerdo con algunos estudios previos.²⁵¹ En el HDAg las dos regiones más conservadas a nivel de aminoácidos estaban codificadas por regiones incluidas en las más conservadas a nivel de nucleótidos (la primera región corresponden a los nucleótidos 1122-1091 y la segunda a los nucleótidos 994-970). Esto sugiere que la conservación de estas regiones nucleotídicas está asociada al mantenimiento de los aminoácidos que codifican. Por un lado, la primera región conservada consta de los aminoácidos del 160-169 que pertenecen a la región de unión de la ARN pol II (aminoácidos 150-195).^{80,81} Este motivo de aminoácidos (160-169) también se ha encontrado conservado en otros estudios mediante secuenciación Sanger.^{251,264-266} Por tanto, estos hallazgos por NGS confirman, con un nivel de sensibilidad mucho más elevado lo que ya se ha observado por otras técnicas. Además se ha reportado que la región entre los aminoácidos 161-166 es altamente conservada y actúa como dominio inmunogénico,²⁶⁷ sin embargo, el alto grado de conservación de una región hace cuestionar su papel como dominio inmunogénico. Un estudio de Yamaguchi *et al.*⁸¹ reportó que D187, I188 y F194 eran los aminoácidos críticos para la interacción del S-HDAg con la ARN Pol II. Curiosamente, D187 e I188 pertenecen a la región del S-HDAg codificada por la región más variable a nivel de nucleótidos, 1024-1048. Así, mientras que D187 está conservado en todos los pacientes analizados, I188 (que en la secuencia de referencia utilizada sería V188, no está conservado, solo se encuentra I188 en todos los haplotipos de 2 muestras de un paciente y de 3 muestras de 3 pacientes diferentes. F194 sólo no está conservado en las 3 muestras de un paciente pero con frecuencias de cambió del 1,1%, 0,4% y 0,5% (datos no mostrados). Además en ese mismo estudio se observó que S170, A180, T182 no son críticos para la interacción del S-HDAg con la ARN pol II. Los resultados a nivel de conservación en el presente estudio parecen confirmar estos resultados previos, la secuencia de referencia utilizada muestra el aminoácido T180 y las otras dos posiciones pertenecen al Patrón de mutaciones 1 (S170N y T182H), que aunque solo está presente en un paciente, en la primera muestra incluida se encuentra representado en el 100% de la quasiespecies.

Por otro lado, la segunda región conservada consta de los aminoácidos 199-212 dónde se incluyen los aminoácidos del 203-209 codificada por la región conservada a nivel de nucleótidos (994-970) que pertenece a la señal de ensamblaje viral (aminoácidos 195-

214) que comprende la señal de exportación nuclear (aminoácidos 198-210) y la señal de isoprenilación (aminoácidos 211-214). Hay otros estudios que también han encontrado conservada la parte de esta región entre los aminoácidos 199-203.^{264,265} En esta región hay dos posiciones importantes, la P205 y la C211, que en este estudio se encuentran conservadas en todas las muestras de todos los pacientes, al igual que en otros estudios.^{251,264-266} La posición de la P205 es crítica para la relocalización del L-HDAg desde el núcleo al citoplasma y está involucrada en el ensamblaje del VHD. Además, desempeña un papel esencial en el transporte de la RNP fuera del núcleo.^{268,269} La posición C211, es donde tiene lugar una modificación post-traducciona, la isoprenilación, necesaria para que el virus interactúe con el HBsAg. Además, a esta posición la farnesil transferasa celular le une covalentemente un farnesilo, lo cual es importante para inhibir la replicación viral y al mismo tiempo promover la encapsidación del complejo RNP y la liberación de las partículas virales.⁹³⁻⁹⁵ Se ha observado que una mutación en esta posición elimina tanto la isoprenilación como la formación de partículas virales.⁹³

En cambio, las dos regiones más variables encontradas a nivel de aminoácidos y a nivel de nucleótidos, no coinciden. Como se ha comentado anteriormente, la región más variable a nivel de nucleótidos (los nucleótidos 1024-1048) se localiza en la parte del ORF delta que codifica el dominio de unión a la ARN pol II (aminoácidos 150-195) del HDAg. La región más variable a nivel de aminoácidos del 112-121 pertenece a la región *helix-loop-helix* (aminoácidos 108-137) del S-HDAg. Esta región estructural consiste en dos α -hélices unidas por un bucle. Se cree que puede contribuir en la oligomerización del HDAg y en la interacción del HDAg con otras proteínas.^{252,253} Otro estudio también ha observado variaciones en esta región en muestras de genotipo 1.²⁶⁴ Sin embargo, es de destacar que la posición L115, importante para la replicación del VHD y necesaria para la actividad de unión al ARN *in vitro*,²⁵⁴ sí que está conservada en todas las muestras de todos los pacientes analizados, como se ha observado en otro estudio mediante secuenciación Sanger.²⁵¹

En cuanto al análisis de las mutaciones encontradas en la región del HDAg analizada (posiciones aminoacídicas 111-214 del L-HDAg), se han encontrado un total de 98 mutaciones que afectan a 54 posiciones, más de la mitad de las 105 analizadas, lo que indica la gran variabilidad del fragmento del HDAg estudiado. Una vez encontradas las mutaciones existentes en esta región, se seleccionaron las que sus frecuencias mostraban una $\rho > 0,8$ y un p -valor $< 0,05$ según la prueba de Spearman, lo cual les da una elevada probabilidad de transmitirse siempre juntas en un mismo haplotipo. En total

se identificaron 19 mutaciones que podían agruparse en 5 patrones de mutaciones aminoacídicas.

El patrón de mutaciones 1 es el más complejo, comprende hasta 9 mutaciones diferentes; E141K, R139E, V149P, S159Q, T182H, E184D, A202S a las que en algunos casos se puede sumar la mutación E135G y después la mutación S170N. Este patrón está presente en 1 paciente de los 12 analizados (8,3%), en el que se observa que a lo largo del seguimiento, la frecuencia de este patrón de mutaciones disminuye. Las mutaciones T182H y S170N aunque se encuentran en el dominio de interacción con la ARN pol II (aminoácidos 150-195) no parecen ser críticas para que esta interacción se produzca. La fosforilación de T182 ha sido observada *in vitro*, aunque no se conoce con certeza su funcionalidad.⁸⁹ Sin embargo, dada su falta de conservación no parece que T182 sea un sitio esencial de fosforilación. Otros estudios previos han reportado las mutaciones R139E y E141K,^{251,264} cambios a priori muy radicales teniendo en cuenta las características físico químicas de los aminoácidos implicados. Estas dos posiciones, junto a la V149P, están incluidas en uno de los dominios de unión a ARN (aminoácidos 97-146).²⁵⁴ Las mutaciones dentro de este dominio puede indicar un papel menos importante para la unión al ARN, en comparación con los otros dominios de unión a ARN presentes en el HDAG.²⁵¹ Sin embargo, el cambio de una arginina (R, aminoácido con carga positiva) por un ácido glutámico (E, aminoácido con carga negativa) en la posición 139 junto al cambio de E por lisina (K, aminoácido con carga positiva) en la posición 141 parece sugerir un mecanismo de compensación de la carga eléctrica que podría permitir mantener la funcionalidad de este dominio. No obstante, el cambio V149P parece sugerir un cambio muy radical a nivel de la estructura, puesto que la aparición de una Prolina (P) sugeriría la fijación de un loop en la estructura proteica. Se puede especular que este cambio podría estar asociado al mantenimiento de la funcionalidad del dominio, pero sería necesario un modelo estructural de la proteína para inferir las posibles consecuencias de este cambio. Asimismo, las posiciones S159Q y E184D, esta segunda “casi sinónima” por las características de los aminoácidos, pertenecen a la región de unión a la ARN pol II (aminoácidos 150-195), y podrían afectar a posiciones no esenciales funcionalmente. La mutación A202S se encuentra en la señal de empaquetamiento, en una posición en la que estudios previos también han descrito mutaciones,^{103,251,264-266,270} entre ellas esta misma mutación A202S.²⁷⁰ Por último, algunas mutaciones se encuentran dentro de regiones epitópicas. Por ejemplo, los cambios E141K, R139E, V149P, S159Q, T182H, E184D, E135G y S170N se encuentran en un epítipo inmunogénico de células T CD8+²⁵¹ y los cambios S159Q, T182H, E184D y S170N se encuentran en un epítipo de

células B.¹⁰⁹ Estas mutaciones en regiones epitópicas, como todas las que se comentarán en los siguientes patrones, podrían permitir al virus a escapar de la respuesta inmune del huésped, como se ha especulado también en otros estudios.^{100,109,271} Así pues, posiblemente las mutaciones que conforman el patrón 1 respondan a cambios conformacionales en regiones epitópicas del HDAg, para escapar a la respuesta inmune del huésped, manteniendo, al menos parcialmente, la funcionalidad de los diferentes dominios funcionales de esta proteína. Por lo tanto, la dinámica de sus proporciones a través del seguimiento del paciente en el que se ha identificado probablemente dependa de cambios en esta respuesta inmune a lo largo de su seguimiento.

El patrón de mutaciones 2, comprende las mutaciones P213T y S159A que en una elevada proporción de casos se le suma la mutación G158A; es el patrón más abundante, está presente en 7 pacientes de los 12 analizados (58%). En la mayoría de los pacientes el patrón tiene una frecuencia elevada en todas sus muestras, de más del 85%. En cambio en un paciente es baja, de menos del 15% y en otro paciente se mantiene elevada en la primera y segunda muestra pero disminuye considerablemente de la segunda a la tercera muestra. La mutación P213T afecta a la señal de ensamblaje de las partículas virales y más concretamente a la señal de isoprenilación. El cambio de P por un aminoácido con propiedades tan diferentes como Treonina (T) no parece afectar a la replicación viral. No obstante, como se ha comentado anteriormente, la posición más importante de la señal de isoprenilación es C211, donde tiene lugar la isoprenilación, y la posición más importante de la señal de ensamblaje es P205. Así pues, P213 podría no ser esencial ni para el ensamblaje de partículas virales ni para la isoprenilación de L-HDAg, a pesar del gran cambio estructural que podría representar el cambio de P por V. Las mutaciones S159A y G158A están situadas en posiciones no esenciales de la región de unión de la ARN pol II (aminoácidos 150-195) y pertenecen a un epítipo inmunogénico de células T CD8+ y a un epítipo de células B.^{109,251} Eso sugiere la posibilidad que las diferentes tendencias de las frecuencias de este patrón de mutaciones entre los pacientes que lo presentan podría ser debida a las diferencias entre sus respuesta inmunes frente al virus. La mutación P213T aunque no está dentro del epítipo inmunogénico citado anteriormente podría colaborar en el escape al sistema inmune.

El patrón de mutaciones 3, comprende las mutaciones A117G y Q192R, está presente en 3 de los 12 pacientes analizados (25%). Dos de ellos tienen el patrón en frecuencias muy elevadas, de más del 95% en todas las muestras del seguimiento. En cambio, aunque el otro paciente (paciente 7) muestra el patrón en la misma frecuencia que los otros dos en

la primera muestra, en la segunda muestra el patrón prácticamente desaparece mientras que en la tercera vuelve a seleccionarse intensamente y sus proporciones aumentan hasta cerca del 70% de la quasiespecies. La práctica desaparición del patrón en la segunda muestra coincide con un aumento de linfocitos y monocitos (datos no mostrados), por lo que podría suponerse una activación de la inflamación, por lo tanto, de nuevo, un cambio en la respuesta inmune. La mutación A117G, pertenece a la región *helix-loop-helix* (aminoácidos 108-137) del S-HDAg y la Q192R a la región de interacción con la ARN pol II (aminoácidos 150-195); mutaciones en dichas regiones ya se han comentado anteriormente. Además, ambas mutaciones se encuentran en regiones epitópicas: la A117G está dentro del epítipo de células restringido por el complejo mayor de la histocompatibilidad II y dentro de un *cytotoxic T ligand epitope* (aminoácidos 106-121)¹⁰⁹ y la Q192 dentro de un epítipo de células B y de un epítipo inmunogénico de células T CD8+.^{109,251} Así pues, estas mutaciones pueden estar relacionadas con un mayor *fitness* de las poblaciones virales en determinados momentos, ya que les pueden permitir escapar a la presión del sistema inmunitario.

El patrón de mutaciones 4, comprende las mutaciones R139K y Q172L, y está presente en 2 de los 12 pacientes analizados (16,7%). En un paciente con frecuencias altas (paciente 12), superiores al 85% de la quasiespecies a lo largo del seguimiento, y en el otro con frecuencias bajas (paciente 1), por debajo del 5%. Curiosamente, ambos pacientes también presentaron el patrón 2, la evolución del cual siguió tendencias opuestas a las del patrón 4. El paciente 1 con frecuencias bajas del patrón 4 mostró frecuencias altas del patrón 2 y el paciente 12, con frecuencias altas del patrón 4, aunque en la primera y segunda muestras también mostró frecuencias altas del patrón 2, de la segunda a la tercera muestra se denota una elevación de la frecuencia del patrón 4 coincidiendo con una gran disminución de la frecuencia del patrón 2. La mutación R139K se encuentra en un dominio de unión a ARN (aminoácidos 97-146) y también se ha observado en otro estudio.²⁶⁴ El cambio de R por K en un dominio de unión a ARN no parece indicar un cambio significativo de la capacidad de este dominio para interactuar con los ácidos nucleicos, dada la carga positiva de ambos aminoácidos. La mutación Q172L pertenece a la región de interacción con la ARN pol II (aminoácidos 150-195). Además, estas dos mutaciones están dentro de un epítipo inmunogénico de células T CD8+²⁵¹ y la Q172L además está dentro de un epítipo de células B.^{109,251} Esta última mutación implica el cambio de un aminoácido polar como la Glutamina (Q) por uno de apolar como la Leucina (L), parece sugerir un cambio muy radical a nivel de la estructura que puede causar un cambio en la conformación del epítipo de células B. Q172 no se ha

descrito como un residuo esencial para que tenga lugar la interacción entre el S-HDAg y la ARN pol II, sin embargo, sería necesario un modelo estructural de la proteína para predecir las posibles consecuencias funcionales de este cambio. Las tendencias opuestas a lo largo del seguimiento que parecen seguir las frecuencias de este patrón respecto al patrón 2, podrían ser debidas a la selección alternativa de variantes con uno u otro patrón para escapar de la presión del sistema inmune en distintos momentos, en los pacientes 1 y 12.

El patrón de mutaciones 5, comprende las mutaciones R132K e I198L y está presente en 2 de los 12 pacientes analizados (16,7%). En el paciente 3 las frecuencias de este patrón son muy bajas y en el 7 la frecuencia relativa fluctúa a lo largo del tiempo, siguiendo una tendencia casi idéntica a la del patrón 3 en este mismo paciente. Así pues, en el paciente 7, en la primera muestra se observa una frecuencia muy elevada de este patrón, en la segunda muestra esta frecuencia disminuye hasta alrededor de un 5% de la quasiespecies y en la tercera muestra sus proporciones aumentan hasta cerca del 90% de la quasiespecies. Las fluctuaciones en la frecuencia de los patrones 3 y 6 observadas en este paciente, coinciden con una elevación de linfocitos y monocitos en la segunda muestra (datos no mostrados), por lo que podría suponerse una activación de la inflamación. La mutación R132K pertenece a la región *helix-loop-helix* (aminoácidos 108-137) del S-HDAg y a un epítipo inmunogénico de células T CD8+²⁵¹ y la mutación I198L pertenece a la señal de empaquetamiento viral (aminoácidos 195-214) y a la señal de exportación nuclear (aminoácidos 198-210). También se han descrito mutaciones en la posición 198 en estudios previos.^{103,251,264-266,270} Dada la naturaleza de los aminoácidos que cambian, R por K en la posición 132 e Isoleucina (I) por L, no son esperables cambios funcionales o conformacionales importantes debido a estos aminoácidos, por lo tanto quedarían por aclarar los motivos de la acusada fluctuación de las proporciones de este patrón a lo largo del seguimiento del paciente 7.

Las mutaciones aminoacídicas en regiones funcionales del HDAg merecen una mayor investigación a gran escala respaldada por estudios clínicos exhaustivos, para determinar su influencia en la supervivencia y/o la patogenicidad del virus. Además, la diversidad en la secuencia de aminoácidos del HDAg en los epítipos inmunogénicos, puede estar relacionada con el escape viral de las respuestas inmunes y podría servir como punto de partida para mejorar las estrategias terapéuticas.

6.3.2 Análisis de la conservación y variabilidad en la región de la ribozima

Por último, el grado de conservación o variabilidad de la región de la ribozima se determinó únicamente a nivel de nucleótidos, ya que se trata de una región no codificante. Debido al mecanismo de replicación en círculo rodante que utiliza el VHD, que da lugar a “concatameros” con varias moléculas del ARN viral unidas, la actividad de la ribozima es crucial para producir la escisión de dichas moléculas de ARN de longitud unitaria.^{96,272} Por eso, en el presente estudio, se parte de la hipótesis de que la secuencia de la ribozima debería ser poco variable. Los resultados obtenidos parecen confirmar esta hipótesis, ya que la mayoría de las posiciones de esta región están en torno al 100% de conservación (contenido de información alrededor de 2 bits). Las mutaciones que se comentan a continuación fueron mayoritariamente transiciones y algunas están en tan baja proporción que casi ni se aprecian en el logo de toda la secuencia de la ribozima, que se muestra en el apartado 5.2.3 *Análisis de la conservación y variabilidad de la región de la ribozima estudiada*. Cabe destacar que la ribozima, como región autocatalítica que es, actúa siempre sobre la misma molécula de ARN que la contiene. Esto no sería así de ser una actividad enzimática debida a una proteína, como ocurre en los demás virus. Por ejemplo, en el VHB se pueden encontrar variantes deficitarias incluso de la polimerasa viral pero que pueden ser replicadas por la transcomplementación por polimerasas codificadas por otros componentes de la quasispecies. Por lo tanto, todas las variantes de la ribozima del VHD identificadas en este estudio deben ser capaces de catalizar la escisión de moléculas de ARN de longitud unitaria, con mayor o menor eficiencia, lo que podría dar lugar a las diferentes proporciones en las que se han encontrado estas mutaciones. En este sentido se puede especular que a menor frecuencia de una variante menor eficiencia replicativa de la misma o sea menor eficiencia de la correspondiente ribozima, no obstante este hecho debe confirmarse con pruebas fenotípicas que están actualmente en proceso, mediante la modelización por mutagénesis dirigida de las variantes detectadas en este estudio.

La estructura secundaria de la ribozima consta de regiones de doble cadena (P1, P1.1, P2, P3 y P4), regiones monocatenarias (J1/2, J1.1/4 y J4/2) y regiones en forma de bucle (L3 y L4).⁹⁷ En este estudio se ha observado que la región más variable se encuentra en la región P4, al igual que otros estudios previos,²⁵¹ especialmente debido a las mutaciones T60C, C61T, C62T y T69C, incluidas dentro del segmento más variable de la ribozima, entre las posiciones G59 y A70, que corresponden a las posiciones nucleotídicas 746-757. Se ha sugerido en varios estudios que la región P4 tiene una

función de estabilización estructural sin participar directamente en la catálisis, ya que, su eliminación completa no inhibe la actividad de la ribozima.^{273,274}

Por otro lado, la región L3, se ha sugerido como una región conservada, ya que es esencial para la actividad de escisión de la ribozima o contribuye a su actividad óptima.²⁷³⁻²⁷⁵ Sin embargo, en este estudio la posición T23 mostró un nivel de variabilidad considerable. En esta posición se ha detectado la mutación T23C en esta región, la cual ya había sido descrita previamente.²⁵¹ Aunque pertenece a la región L3, se ha sugerido que esta mutación sea menos relevante a nivel funcional que las mutaciones en posiciones próximas de la ribozima como T20, C21 y G25,²⁵¹ que supondrían una gran pérdida de la actividad de la ribozima.²⁷⁴⁻²⁷⁷ Las posiciones T20 y C21, que previamente se habían asignado a la región L3, actualmente no se sitúan ahí, T20 no pertenece a ninguna región y C21 pertenece a la región P1.1.⁹⁶ Además, en la región L3, en este estudio también se observa la mutación T27C.

En la región P1, se requiere el apareamiento de bases estable para garantizar una actividad eficiente en la autoescisión.²⁷⁸ Posiblemente por este motivo, parte de esta región se haya incluido dentro del fragmento más conservado de la ribozima (entre G28 y C58, que corresponden a las posiciones nucleotídicas 715-745) Además esta región contiene el sitio de autoescisión.²⁷⁹ Estudios *in vitro* mostraron que los cambios en P1 o J4/2 resultaron en la pérdida de la actividad de la ribozima severamente o completamente reducida o una interacción entre los iones metálicos (Mg^{2+}) y el ARN también reducida.^{272,275-277,260,281} En este estudio se han identificado las posiciones G6 y T37 como las más variables, con las mutaciones G6A y T37C (esta última en el fragmento más conservado de la ribozima). La mutación en T37 se espera que reduzca drásticamente la actividad de la ribozima.^{275-277,260} Sin embargo, en otro estudio²⁵¹ también se identificaron mutaciones en esta posición. El seguimiento de los pacientes con la mutación T37 de ese estudio mostraron un cambio al fenotipo salvaje a lo largo del tiempo, por tanto eso demuestra que la replicación del VHD no fue abolida como se podría anticipar según los estudios *in vitro* nombrados anteriormente.²⁵¹ En cuanto a mutaciones en la posición G6, no se han encontrado datos.

En la región J4/2, en este estudio se observa la mutación G76C, una de las pocas transversiones observadas, en proporciones residuales. Pero la posición importante de esta región es la C75, que se encuentra conservada en todas las muestras. La C75 tiene un papel importante para la actividad catalítica, ya que las mutaciones en esta posición suponen una pérdida total de la actividad de la ribozima. Además se cree que puede

interactuar con los iones de Mg^{2+} durante la autoescisión, actuando de catalizador ácido-base.^{274,275,277,278,282} Además otros estudios proponen que C75 actúa como la base general que activa el grupo 2'-hidroxilo del nucleótido -1 para el ataque nucleofílico.²⁷⁸

En la región J1/4 también destaca el nivel de variabilidad de la posición G40. En este estudio se ha identificado la mutación G40A, aunque la posición importante de esta región es la G38, que sí que se observa conservada en todas las muestras y está incluida en el fragmento más conservado de la ribozima. Esta posición desempeña un papel importante, ya que, las mutaciones en esa posición reducen la actividad de autoescisión de la ribozima a niveles muy bajos. Además también se cree que puede interactuar con los iones de Mg^{2+} durante la autoescisión.^{275,277,283}

El resto de mutaciones observadas en otras regiones de la ribozima se encuentran todas en proporciones marginales. En la región P2 en este estudio se observa la mutación C84T. Esta región es importante para estabilizar la estructura de autoescisión, aunque esta estructura podría formarse en ausencia de P2, pero si la región P2 está intacta facilita la autoescisión al estabilizar la estructura.²⁷⁸ La región P3 se considera esencial para la actividad de autoescisión, mutaciones en esta región resultan en niveles muy bajos de la actividad de la ribozima.^{273,278} En este estudio se observa la mutación C19T. La región L4 es más bien estructural,²⁷⁴ en este estudio se observan las mutaciones C58G y G59C, que junto con la mutación G76C son las únicas transversiones que se han identificado en la ribozima, todas ellas en proporciones muy bajas. La región J1/2 se cree que no tiene una función importante,²⁸⁴ aunque en este estudio está totalmente conservada, por lo que merecería la pena volver a explorar la esencialidad de esta región en futuros estudios *in vitro*.

6.4 Lecciones aprendidas sobre la quasiespecies del VHD

Los resultados del primer estudio demuestran que la NGS es altamente sensible y precisa para el estudio cuantitativo de las quasiespecies del VHD circulantes en muestras clínicas y es una buena alternativa a la secuenciación Sanger de los productos obtenidos por la clonación molecular. Además, los resultados han proporcionado información relevante sobre la quasiespecies del VHD: el porcentaje relativamente alto de genomas editados circulantes y el sesgo hacia las transiciones A:G y G:A (U:C o C:U en el ARN genómico) aluden a mecanismos potencialmente importantes para explicar, al menos parcialmente, la quasiespecies del VHD y por lo tanto su virología. Por otro lado, los

resultados del segundo estudio proporcionan una visión global de la conservación y variabilidad de dos de las regiones funcionales del genoma del VHD, como son la región del HDAg y de la ribozima, evidenciando una mayor conservación de la región de la ribozima pero con la presencia de variantes a bajo nivel, que deben ser estudiadas en el futuro. Estas dos regiones necesitan un mayor estudio con un mayor número de pacientes para que análisis de este tipo sean útiles como punto de partida para mejorar las estrategias terapéuticas o para profundizar en el conocimiento de la virología del VHD.

CONCLUSIONES

7. CONCLUSIONES

1. La secuenciación masiva es un método útil para detectar proporciones de genomas del VHD no editados y editados al analizar muestras clínicas. Además es un método más sensible y preciso que la secuenciación Sanger de productos de clonación para cuantificar poblaciones virales en la quasiespecies del VHD, como por ejemplo las poblaciones de genomas editados y no editados. Mediante esta metodología se ha podido confirmar que los genomas no editados son mayoritarios en la quasiespecies del VHD circulante.
2. El patrón de cambios nucleotídicos en el genoma del VHD muestra una frecuencia más elevada de cambios de tipo transición que de tipo transversión. Dicho patrón sugiere que el fenómeno de edición del genoma del VHD por enzimas mutagénicas (no solamente por ADAR-1) podría no ser exclusivo de la posición de edición en la región codificante del antígeno delta (HDAg) analizada.
3. En el estudio de la conservación y variabilidad del HDAg se ha observado que las dos regiones más conservadas a nivel de aminoácidos estaban codificadas por las dos regiones más conservadas a nivel de nucleótidos. Estas regiones pertenecen a dominios esenciales para la replicación del VHD (unión de la ARN polimerasa II y morfogénesis de las partículas virales). Esto sugiere que la conservación de la secuencia nucleotídica se asociaría a la preservación de la función esencial del HDAg en la replicación viral. En contraste, la región más variable a nivel de aminoácidos no estaba codificada por la región más variable a nivel de nucleótidos. Esta región pertenece a un dominio con funcionalidad estructural (dominio *helix-loop-helix*) lo que cuestiona que su papel sea esencial, o que la conservación de su secuencia aminoacídica sea imprescindible para realizar este papel.
4. Los patrones de mutaciones de aminoácidos individuales observados se localizan en general en regiones epitópicas. Dichas mutaciones podrían aún así mantener la funcionalidad de los dominios funcionales del HDAg a los que pertenecen. Las diferentes tendencias de las frecuencias de estos patrones de mutaciones observadas en los pacientes a lo largo del seguimiento longitudinal podrían estar relacionadas con el escape del sistema inmunitario.

5. La secuencia del genoma viral correspondiente a la ribozima, como se esperaba, mostró un elevado nivel de conservación en su secuencia nucleotídica. Sin embargo, se observaron cambios de nucleótidos mayoritariamente de tipo transición, que no afectan a posiciones descritas como esenciales para la función de esta región, que es la escisión de los monómeros de ARN. Por las características de esta actividad enzimática, asociada a una secuencia nucleotídica, este hecho parece indicar que cualquier variante detectada en circulación tendría una ribozima funcional que permitiría al genoma viral replicarse, aunque por la baja frecuencia, quizá con una menor efectividad.

LÍNEAS DE FUTURO

8. LÍNEAS DE FUTURO

Los dos estudios incluidos en esta tesis estudian la quasispecies del VHD en dos de las regiones más importantes de su genoma: el HDAg, la única proteína que codifica el VHD; y la ribozima, el único elemento viral con actividad catalítica y de vital importancia para la replicación y procesamiento del genoma viral. Los resultados obtenidos abren las puertas a nuevas líneas de investigación básicas y clínicas sobre la hepatitis delta crónica.

8.1 Líneas clínicas

En primer lugar, en relación a las líneas de investigación más asociadas a la clínica o con implicaciones en el tratamiento antiviral de los pacientes, se ha visto que la secuenciación masiva es una metodología útil para cuantificar poblaciones del VHD circulantes no editadas y editadas en muestras clínicas, con alta sensibilidad y precisión. A pesar de que habitualmente circula una proporción mayor de poblaciones no editadas que de las editadas, estas últimas se han detectado en proporciones considerables (alrededor del 37% de los genomas circulantes), lo cual parece significativo ya que estas poblaciones son ineficientes replicándose.²⁶¹ Teniendo en cuenta este dato sería interesante buscar las implicaciones clínicas de estas diferentes proporciones además de realizar el estudio en un número mayor de pacientes, con características clínicas y virológicas homogéneas que permitieran realizar estudios estadísticos significativos. Para ello, se podrían buscar correlaciones entre las proporciones de genomas no editados y editados con el *baseline-event-anticipation score* (BEA score). Este sistema de puntuación se desarrolló en base a variables asociadas con el desarrollo de complicaciones clínicas relacionadas con el hígado, tal como la edad, el sexo, la región de origen, los valores de bilirrubina, los valores de plaquetas y el índice internacional normalizado (INR).²⁸⁵ Además, los genomas editados que codifican para el L-HDAg son esenciales en el establecimiento y mantenimiento de la infección crónica, sin embargo estos genomas son prácticamente ineficientes replicándose (al no ser capaces de codificar el S-HDAg, que es esencial para la replicación como se ha comentado en el apartado 6. *Discusión*). Por tanto, podría ser importante estudiar cómo ajustar la carga viral de acuerdo con el porcentaje de genomas no editados al monitorizar pacientes con hepatitis delta crónica.

En cuanto al segundo estudio, los datos sobre conservación y variabilidad obtenidos en la región codificante del HDAg y la secuencia de aminoácidos de este antígeno podrían utilizarse para seleccionar posibles dianas para terapias génicas antivirales. Por ejemplo, ya que la secuencia codificante del HDAg está incluida en los 800 nucleótidos del ARNm sintetizado por el VHD,²⁴ una terapia de bloqueo basada en *small interference RNA* (ARNsi) dirigidos contra esas regiones hiper-conservadas podría resultar una valiosa herramienta terapéutica. Para establecer el papel de estos patrones de mutaciones en el escape al sistema inmune, un primer paso podría ser relacionar sus frecuencias a lo largo del seguimiento de los pacientes con los niveles de transaminasas o de su carga viral, e identificarse dichos patrones de mutaciones en diferentes grupos de pacientes para analizar si se encuentran relacionados con la severidad de la enfermedad.

Por último, la gran conservación de la región de la ribozima y el hecho de que sea totalmente esencial para la replicación del virus, hace plantear que sea una buena diana terapéutica. Por ejemplo, podría ser útil para diseñar oligos antisentido, incluyendo la reciente tecnología *gapmers*, que a diferencia del *small interference RNA* (ARNsi), serían capaces de entrar en el núcleo celular e inhibir la actividad de la ribozima.¹⁹⁴ Las regiones de la ribozima altamente conservadas descritas en este proyecto podrían ser útiles para empezar a realizar pruebas fenotípicas y encontrar una estrategia de bloqueo de la ribozima.

8.2 Líneas básicas

Estas líneas están más asociadas a entender la virología del VHD y su interacción con el sistema inmune. Para comprobar el efecto fenotípico de las diferentes proporciones de genomas no editados y editados podrían realizarse estudios fenotípicos en un sistema de co-transfección o co-infección de construcciones autorreplicantes de VHD y VHB (o que al menos expresen el antígeno de superficie de este último) en líneas celulares capaces de sustentar su replicación. Dicho sistema se ha puesto a punto recientemente por nuestro grupo y abrirá las puertas para este estudio así como otros que se comentarán a continuación.

En segundo lugar, se ha observado que la edición puede que no tenga lugar solamente en la posición de edición y que se han observado más cambios de tipo transición que de tipo transversión en la región del HDAg analizada y entre estos cambios el 74% fueron de A:G o G:A (U:C o C:U en el ARN genómico). Esto sugiere la posible edición del genoma

viral por parte de las enzimas mutagénicas del sistema inmune innato (APOBEC, ADAR-1...), que podría confirmarse en modelos de infección *in vivo*.

En cuanto al segundo estudio, los diferentes patrones de mutaciones encontrados en la región del HDAG se podrían modelizar *in vitro* para así analizar su importancia dentro de las diferentes regiones funcionales en las que se encuentran. Los efectos funcionales de las mutaciones de estos patrones se podrían estudiar mediante el sistema de co-transfección o co-infección de construcciones autorreplicantes de VHD y VHB puesto a punto por nuestro grupo, en líneas celulares capaces de sustentar su replicación. Asimismo, estos patrones también podrían modelizarse *in vivo* para estudiar su posible papel en el escape a la respuesta inmune, utilizando modelos capaces de establecer una respuesta inmune frente a la infección por VHD.

El mismo sistema de co-transfección o co-infección de construcciones autorreplicantes de VHD y VHB también podría ser utilizado para realizar pruebas *in vitro* para analizar la capacidad de autoescisión de las diferentes variantes encontradas en la región de la ribozima y así analizar la eficiencia replicativa de cada una de ellas. Este análisis podría complementarse con estudios estructurales sobre la estructura secundaria de la región de la ribozima, analizando el plegamiento del ARN viral de las diferentes variantes caracterizadas, para ver si las mutaciones nucleotídicas encontradas tendrían un efecto en el cambio de su estructura secundaria, lo cual podría afectar a su funcionalidad

BIBLIOGRAFÍA

BIBLIOGRAFÍA

1. Rizzetto M. The delta agent. *Hepatology*. 1983; 3: 729-737.
2. Taylor JM. Virology of hepatitis D virus. *Sem Liver Dis*. 2012; 32: 195-200.
3. Rizzetto M, Canese MG, Aricò S, Crivelli O, Trepo C, Bonino F et al. Immunofluorescence detection of new antigen-antibody system (delta/anti-delta) associated to hepatitis B virus in liver and in serum of HBsAg carriers. *Gut*. 1977; 18 (12): 997-1003.
4. Rizzetto M, Hoyer B, Canese MG, Shih JW, Purcell RH, Gerin JL. delta Agent: association of delta antigen with hepatitis B surface antigen and RNA in serum of delta-infected chimpanzees. *Proc Natl Acad Sci U S A*. 1980; 77 (10): 6124-6128.
5. Wang KS, Choo QL, Weiner AJ, Ou JH, Najarian RC, Thayer RM et al. Structure, sequence and expression of the hepatitis delta (delta) viral genome. *Nature*. 1986; 323 (6088): 508-514.
6. Kos A, Dijkema R, Arnberg AC, van der Meide PH, Schellekens H. The hepatitis delta (delta) virus possesses a circular RNA. *Nature*. 1986; 323 (6088): 558-560.
7. Makino S, Chang MF, Shieh CK, Kamahora T, Vannier DM, Govindarajan S et al. Molecular cloning and sequencing of a human hepatitis delta (delta) virus RNA. *Nature*. 1987; 329 (6137): 343-346.
8. Kuo MY, Goldberg J, Coates L, Mason W, Gerin J, Taylor J. Molecular cloning of hepatitis delta virus RNA from an infected woodchuck liver: sequence, structure, and applications. *J Virol*. 1988; 62 (6): 1855-1861.
9. Wu JC, Chen CM, Sheen IJ, Lee SD, Tzeng HM, Choo KB. Evidence of Transmission of Hepatitis D Virus to Spouses From Sequence Analysis of the Viral Genome. *Hepatology*. 1995; 22 (6): 1656-1660.
10. Niro GA, Casey JL, Gravinese E, Garrubba M, Conoscitore P, E Sagnelli E et al. Intrafamilial Transmission of Hepatitis Delta Virus: Molecular Evidence. *J Hepatol*. 1999; 30 (4): 564-569.
11. Kucirka LM, Farzadegan H, Feld JJ, Mehta SH, Winters M, Glenn JS et al. Prevalence, Correlates, and Viral Dynamics of Hepatitis Delta Among Injection Drug Users. *J Infect Dis*. 2010; 202 (6): 845-852.
12. World Health Organization [Internet]. c2019 [cited 2020 Apr 23]. Available from: <https://www.who.int/news-room/fact-sheets/detail/hepatitis-d>
13. Chen HY, Shen DT, Ji DZ, Han PC, Zhang WM, Ma JF et al. Prevalence and Burden of Hepatitis D Virus Infection in the Global Population: A Systematic Review and Meta-Analysis. *Gut*. 2019; 68 (3): 512-521.
14. Wedemeyer H, Negro F. Devil Hepatitis D: An Orphan Disease or Largely Underdiagnosed? *Gut*. 2019; 68 (3): 381-382.
15. Miao Z, Zhang S, Ou X, Li S, Ma Z, Wang W et al. Estimating the Global Prevalence, Disease Progression, and Clinical Outcome of Hepatitis Delta Virus Infection. *J Infect Dis*. 2020; 221 (10): 1677-1687.

16. Servant-Delmas A, Le Gal F, Gallian P, Gordien E, Laperche S et al. Increasing Prevalence of HDV/HBV Infection Over 15 Years in France. *J Clin Virol.* 2014; 59 (2): 126-128.
17. Wranke A, Pinheiro LM, Parana R, Lobato C, Hamid S, Ceausu E et al. Clinical and Virological Heterogeneity of Hepatitis Delta in Different Regions World-Wide: The Hepatitis Delta International Network (HDIN). *Liver Int.* 2018; 38 (5): 842-850.
18. Hsu SC, Wu JC, Sheen IJ, Syu WJ. Interaction and replication activation of genotype I and II hepatitis delta antigens. *J Virol.* 2004; 78 (6): 2693-2700.
19. Radjef N, Gordien E, Ivaniushina V, Gault E, Anaïs P, Drugan T et al. Molecular phylogenetic analyses indicate a wide and ancient radiation of African hepatitis delta virus, suggesting a deltavirus genus of at least seven major clades. *J Virol.* 2004; 78 (5): 2537-2544.
20. Le Gal F, Gault E, Ripault MP, Serpaggi J, Trinchet JC, Gordien E et al. Eighth major clade for hepatitis delta virus. *Emerg Infect Dis.* 2006; 12 (9): 1447-1450.
21. Su CW, Huang YH, Huo TI, Shih HH, Sheen IJ, Chen SW et al. Genotypes and viremia of hepatitis B and D viruses are associated with outcomes of chronic hepatitis D patients. *Gastroenterology.* 2006; 130(6): 1625-1635.
22. Gomes-Gouvêa MS, Soares MCP, Bensabath G, Carvalho-Mello IMVG, Brito EMF, Souza OSC et al. Hepatitis B Virus and Hepatitis Delta Virus Genotypes in Outbreaks of Fulminant Hepatitis (Labrea Black Fever) in the Western Brazilian Amazon Region. *J Gen Virol.* 2009; 90 (Pt 11): 2638-2643.
23. Rizzetto M. Hepatitis D Virus: Introduction and Epidemiology. *Cold Spring Harb Perspect Med.* 2015; 5 (7): a021576.
24. Hughes SA1, Wedemeyer H, Harrison PM. Hepatitis delta virus. *Lancet.* 2011; 378 (9785): 73-85.
25. Heermann KH, Goldmann U, Schwartz W, Seyffarth T, Baumgarten H, Gerlich WH. Large surface proteins of hepatitis B virus containing the pre-s sequence. *J Virol.* 1984; 52 (2): 396-402.
26. Gudima S, He Y, Meier A, Chang J, Chen R, Jarnik M et al. Assembly of hepatitis delta virus: particle characterization, including the ability to infect primary human hepatocytes. *J Virol.* 2007; 81 (7): 3608-36017
27. Yan H, Zhong G, Xu G, He W, Jing Z, Gao Z et al. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *Elife.* 2012; 1:e00049.
28. Ni Y, Lempp FA, Mehrle S, Nkongolo S, Kaufman C, Fälth M et al. Hepatitis B and D viruses exploit sodium taurocholate co-transporting polypeptide for species-specific entry into hepatocytes. *Gastroenterology.* 2014; 146 (4): 1070-1083.
29. Sureau C, Guerra B, Lee H. The middle hepatitis B virus envelope protein is not necessary for infectivity of hepatitis delta virus. *J Virol.* 1994; 68 (6): 4063-4066.
30. Abou-Jaoudé G, Sureau C. Entry of hepatitis delta virus requires the conserved cysteine residues of the hepatitis B virus envelope protein antigenic loop and is blocked by inhibitors of thiol-disulfide exchange. *J Virol.* 2007; 81 (23): 13057-13066.

31. Sureau C, Guerra B, Lanford RE. Role of the Large Hepatitis B Virus Envelope Protein in Infectivity of the Hepatitis Delta Virion. *J Virol.* 1993 Jan;67(1):366-72.
32. Gudima S, Chang J, Moraleda G, Azvolinsky A, Taylor J. Parameters of human hepatitis delta virus genome replication: the quantity, quality, and intracellular distribution of viral proteins and RNA. *J Virol.* 2002; 76 (8): 3709-3719.
33. Taylor JM. Hepatitis Delta Virus. *Virology.* 2006; 344 (1): 71-76.
34. Lamas Longarela O, Schmidt TT, Schöneweis K, Romeo R, Wedemeyer H, Urban S et al. Proteoglycans act as cellular hepatitis delta virus attachment receptors. *PLoS One.* 2013; 8 (3): e58340.
35. Sureau C, Salisse J. A Conformational Heparan Sulfate Binding Site Essential to Infectivity Overlaps With the Conserved Hepatitis B Virus A-Determinant. *Hepatology.* 2013; 57 (3): 985-994.
36. Verrier ER, Colpitts CC, Bach C, Heydmann L, Weiss A, Renaud M et al. A targeted functional RNA interference screen uncovers glypican 5 as an entry factor for hepatitis B and D viruses. *Hepatology.* 2016; 63 (1): 35-48.
37. Watashi K, Urban S, Li W, Wakita T. NTCP and Beyond: Opening the Door to Unveil Hepatitis B Virus Entry. *Int J Mol Sci.* 2014; 15 (2): 2892-2905.
38. Yan H, Li W. Sodium taurocholate cotransporting polypeptide acts as a receptor for hepatitis B and D virus. *Dig Dis.* 2015; 33 (3): 388-396.
39. Xia YP, Yeh CT, Ou JH, Lai MM. Characterization of nuclear targeting signal of hepatitis delta antigen: nuclear transport as a protein complex. *J Virol.* 1992; 66 (2): 914-921.
40. Chou HC, Hsieh TY, Sheu GT, Lai MM. Hepatitis delta antigen mediates the nuclear import of hepatitis delta virus RNA. *J Virol.* 1998; 72 (5): 3684-3690.
41. Lai MM. RNA replication without RNA-dependent RNA polymerase: surprises from hepatitis delta virus. *J Virol.* 2005;79 (13): 7951-7958.
42. Lempp FA, Ni Y, Urban S. Hepatitis delta: insights into a peculiar pathogen and novel treatment options. *Nat Rev Gastroenterol Hepatol.* 2016; 13 (10): 580-589
43. Greco-Stewart VS, Schissel E, Pelchat M. The hepatitis delta virus RNA genome interacts with the human RNA polymerases I and III. *Virology.* 2009 Mar; 386 (1): 12-15.
44. Greco-Stewart VS, Miron P, Abraham A, Pelchat M. The human RNA polymerase II interacts with the terminal stem-loop regions of the hepatitis delta virus RNA genome. *Virology.* 2007; 357 (1): 68-78.
45. Hsieh SY, Chao M, Coates L, Taylor J. Hepatitis Delta Virus Genome Replication: A Polyadenylated mRNA for Delta Antigen. *J Virol.* 1990; 64 (7): 3192-3198.
46. Lo K, Hwang SB, Duncan R, Trousdale M, Lai MM. Characterization of mRNA for Hepatitis Delta Antigen: Exclusion of the Full-Length Antigenomic RNA as an mRNA. *Virology.* 1998; 250 (1): 94-105.
47. Gudima S, Wu SY, Chiang CM, Moraleda G, Taylor J. Origin of Hepatitis Delta Virus mRNA. *J Virol.* 2000; 74 (16): 7204-7210.

48. Li YJ, Macnaughton T, Gao L, Lai MM. RNA-templated Replication of Hepatitis Delta Virus: Genomic and Antigenomic RNAs Associate With Different Nuclear Bodies. *J Virol.* 2006; 80 (13): 6478-6486.
49. Alfaiate D, Dény P, Durantel D. Hepatitis Delta Virus: From Biological and Medical Aspects to Current and Investigational Therapeutic Options. *Antiviral Res.* 2015; 122: 112-129.
50. Sharmeen L, Kuo MY, Dinter-Gottlieb G, Taylor J. Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J Virol.* 1988; 62 (8): 2674-2679.
51. Wu HN, Lin YJ, Lin FP, Makino S, Chang MF, Lai MM. Human hepatitis delta virus RNA subfragments contain an autocleavage activity. *Proc Natl Acad Sci U S A.* 1989; 86 (6): 1831-1835.
52. Lai MM. The Molecular Biology of Hepatitis Delta Virus. *Annu Rev Biochem.* 1995; 64: 259-86.
53. Reid CE, Lazinski DW. A Host-Specific Function Is Required for Ligation of a Wide Variety of Ribozyme-Processed RNAs. *Proc Natl Acad Sci USA.* 2000; 97 (1): 424-429.
54. Sharmeen L, Kuo MY, Taylor J. Self-ligating RNA Sequences on the Antigenome of Human Hepatitis Delta Virus. *J Virol.* 1989; 63 (3): 1428-1430.
55. Chen PJ, Kalpana G, Goldberg J, Mason W, Werner B, Gerin J et al. Structure and replication of the genome of the hepatitis delta virus. *Proc Natl Acad Sci U S A.* 1986; 83 (22): 8774-8778.
56. Macnaughton TB, Lai MM. Large Hepatitis Delta Antigen Is Not a Suppressor of Hepatitis Delta Virus RNA Synthesis Once RNA Replication Is Established. *J Virol.* 2002; 76 (19): 9910-9919.
57. Taylor JM. Hepatitis D Virus Replication. *Cold Spring Harb Perspect Med.* 2015; 5 (11) :a021568.
58. Harichandran K, Shen Y, Tsoris SS, Lee SC, Casey JL. Hepatitis Delta Antigen Regulates mRNA and Antigenome RNA Levels During Hepatitis Delta Virus Replication. *J Virol.* 2019; 93 (8): e01989-18.eticulo.
59. Chai N, Chang HE, Nicolas E, Han Z, Jarnik M, Taylor J. Properties of Subviral Particles of Hepatitis B Virus. *J Virol.* 2008; 82 (16): 7812-7817.
60. O'Malley B, Lazinski DW. Roles of Carboxyl-Terminal and Farnesylated Residues in the Functions of the Large Hepatitis Delta Antigen. *J Virol.* 2005; 79 (2): 1142-1153.
61. Komla-Soukha I, Sureau C. A Tryptophan-Rich Motif in the Carboxyl Terminus of the Small Envelope Protein of Hepatitis B Virus Is Central to the Assembly of Hepatitis Delta Virus Particles. *J Virol.* 2006; 80 (10): 4648-4655.
62. Wang CJ, Chen PJ, Wu JC, Patel D, Chen DS. Small-form Hepatitis B Surface Antigen Is Sufficient to Help in the Assembly of Hepatitis Delta Virus-Like Particles. *J Virol.* 1991; 65 (12): 6630-6636.
63. Barrera A, Guerra B, Notvall L, Lanford RE. Mapping of the Hepatitis B Virus pre-S1 Domain Involved in Receptor Recognition. *J Virol.* 2005; 79 (15): 9786-9798.

64. Huang C, Chang SC, Yu IC, Tsay YG, Chang MF. Large Hepatitis Delta Antigen Is a Novel Clathrin Adaptor-Like Protein *J Virol*. 2007; 81 (11): 5985-5994.
65. Mentha N, Clément S, Negro F, Alfaiate D. A Review on Hepatitis D: From Virology to New Therapies. *J Adv Res*. 2019; 17: 3-15.
66. Weiner AJ, Choo QL, Wang KS, Govindarajan S, Redeker AG, Gerin JL et al. A single antigenomic open reading frame of the hepatitis delta virus encodes the epitope(s) of both hepatitis delta antigen polypeptides p24 delta and p27 delta. *J Virol*. 1988; 62 (2): 594–599.
67. Luo GX, Chao M, Hsieh SY, Sureau C, Nishikura K, Taylor J. A specific base transition occurs on replicating hepatitis delta virus RNA. *J Virol*. 1990; 64 (3): 1021–1027.
68. Casey JL and Gerin JL. Hepatitis D virus RNA editing: specific modification of adenosine in the antigenomic RNA. *J Virol*. 1995; 69 (12): 7593–7600.
69. Polson AG, Bass BL, Casey JL. RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature*. 1996; 380 (6573): 454-456.
70. Wong SK and Lazinski DW. Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. *Proc Natl Acad Sci USA*. 2002; 99 (23): 15118–15123.
71. Hartwig D, Schoeneich L, Greeve J, Schütte C, Dorn I, Kirchner H, Hennig H. Interferon-alpha Stimulation of Liver Cells Enhances Hepatitis Delta Virus RNA Editing in Early Infection. *J Hepatol*. 2004; 41 (4): 667-672.
72. Hartwig D, Schütte C, Warnecke J, Dorn I, Hennig H, Kirchner H, Schlenke P. The Large Form of ADAR 1 Is Responsible for Enhanced Hepatitis Delta Virus RNA Editing in Interferon-Alpha-Stimulated Host Cells. *J Viral Hepat*. 2006; 13 (3): 150-157.
73. Chao M, Hsieh SY and Taylor J. Role of two forms of hepatitis delta virus antigen: evidence for a mechanism of self-limiting genome replication. *J Virol*. 1990; 64 (10): 5066–5069.
74. Kuo MY, Chao M, Taylor J. Initiation of Replication of the Human Hepatitis Delta Virus Genome From Cloned DNA: Role of Delta Antigen. *J Virol*. 1989; 63 (5): 1945-1950.
75. Yamaguchi Y, Filipovska J, Yano K, Furuya A, Inukai N, Narita T et al. Stimulation of RNA Polymerase II Elongation by Hepatitis Delta Antigen. *Science*. 2001; 293 (5527): 124-127.
76. Huang WH, Mai RT, Lee YH. Transcription Factor YY1 and Its Associated Acetyltransferases CBP and p300 Interact With Hepatitis Delta Antigens and Modulate Hepatitis Delta Virus RNA Replication. *J Virol*. 2008; 82 (15): 7313-7324.
77. Abeywickrama-Samarakoon N, Cortay JC, Sureau C, Alfaiate D, Levrero M, Dény P. Hepatitis Delta Virus Replication and the Role of the Small Hepatitis Delta Protein S-HDAg. *Med Sci (Paris)*. 2018; 34 (10): 833-841.
78. Chang FL, Chen PJ, Tu SJ, Wang CJ, Chen DS. The Large Form of Hepatitis Delta Antigen Is Crucial for Assembly of Hepatitis Delta Virus. *Proc Natl Acad Sci USA*. 1991; 88 (19): 8490-8494.

79. Modahl LE, Lai MM. The Large Delta Antigen of Hepatitis Delta Virus Potently Inhibits Genomic but Not Antigenomic RNA Synthesis: A Mechanism Enabling Initiation of Viral Replication. *J Virol.* 2000; 74 (16): 7375-7380.
80. Greco-Stewart V, Pelchat M. Interaction of Host Cellular Proteins With Components of the Hepatitis Delta Virus. *Viruses.* 2010; 2 (1): 189-212.
81. Yamaguchi Y, Mura T, Chanarat S, Okamoto S, Handa H. Hepatitis Delta Antigen Binds to the Clamp of RNA Polymerase II and Affects Transcriptional Fidelity. *Genes Cells.* 2007; 12 (7): 863-875.
82. Mu JJ, Wu HL, Chiang BL, Chang RP, Chen DS, Chen PJ. Characterization of the phosphorylated forms and the phosphorylated residues of hepatitis delta virus delta antigens. *J Virol.* 1999; 73 (12): 10540-10545.
83. Hwang SB, Lee CZ, Lai MM. Hepatitis delta antigen expressed by recombinant baculoviruses: comparison of biochemical properties and post-translational modifications between the large and small forms. *Virology.* 1992; 190 (1): 413-22.
84. Huang WH, Chen CW, Wu HL, Chen PJ. Post-translational modification of delta antigen of hepatitis D virus. *Curr Top Microbiol Immunol.* 2006; 307: 91-112.
85. Tan KP, Shih KN, Lo SJ. Ser-123 of the Large Antigen of Hepatitis Delta Virus Modulates Its Cellular Localization to the Nucleolus, SC-35 Speckles or the Cytoplasm. *J Gen Virol.* 2004; 85 (Pt 6): 1685-1694.
86. Chen YS, Huang WH, Hong SY, Tsay YG, Chen PJ. ERK1/2-mediated Phosphorylation of Small Hepatitis Delta Antigen at Serine 177 Enhances Hepatitis Delta Virus Antigenomic RNA Replication. *J Virol.* 2008; 82 (19): 9345-9358.
87. Hong SY, Chen PJ. Phosphorylation of Serine 177 of the Small Hepatitis Delta Antigen Regulates Viral Antigenomic RNA Replication by Interacting With the Processive RNA Polymerase II. *J Virol.* 2010; 84 (3): 1430-1438.
88. Yeh TS, Lo SJ, Chen PJ, Lee YH. Casein Kinase II and Protein Kinase C Modulate Hepatitis Delta Virus RNA Replication but Not Empty Viral Particle Assembly. *J Virol.* 1996 Sep;70(9):6190-8.
89. Chen CW, Tsay YG, Wu HL, Lee CH, Chen DS, Chen PJ. The double-stranded RNA-activated kinase, PKR, can phosphorylate hepatitis D virus small delta antigen at functional serine and threonine residues. *J Biol Chem.* 2002; 277 (36): 33058-33067.
90. Mu JJ, Tsay YG, Juan LJ, Fu TF, Huang WH, Chen DS et al. The small delta antigen of hepatitis delta virus is an acetylated protein and acetylation of lysine 72 may influence its cellular localization and viral RNA synthesis. *Virology.* 2004; 319 (1): 60-70.
91. Li YJ, Stallcup MR, and Lai MM. Hepatitis Delta Virus Antigen Is Methylated at Arginine Residues, and Methylation Regulates Subcellular Localization and RNA Replication. *J Virol.* 2004; 78 (23): 13325–13334.
92. Tseng CH, Cheng TS, Shu CY, Jeng KS and Lai MM. Modification of Small Hepatitis Delta Virus Antigen by SUMO Protein. *J Virol.* 2010; 84 (2): 918–927.

93. Glenn JS, Watson JA, Havel CM, White JM. Identification of a prenylation site in delta virus large antigen. *Science*. 1992; 256 (5061): 1331-1333.
94. Hwang SB, Lai MM. Isoprenylation mediates direct protein-protein interactions between hepatitis large delta antigen and hepatitis B virus surface antigen. *J Virol*. 1993; 67 (12): 7659-7662.
95. Hwang SB, Lai MM. Isoprenylation masks a conformational epitope and enhances trans-dominant inhibitory function of the large hepatitis delta antigen. *J Virol*. 1994; 68 (5): 2958-2964.
96. Riccitelli N, Lupták A. HDV family of self-cleaving ribozymes. *Prog Mol Biol Transl Sci*. 2013; 120: 123-171.
97. Ferré-D'Amaré AR, Zhou K, Doudna JA. Crystal structure of a hepatitis delta virus ribozyme. *Nature*. 1998; 395 (6702): 567-574.
98. Rosenstein SP, Been MD. Self-cleavage of Hepatitis Delta Virus Genomic Strand RNA Is Enhanced Under Partially Denaturing Conditions. *Biochemistry*. 1990; 29 (35): 8011-8016.
99. Liu X, Bushnell DA, Kornberg RD. RNA Polymerase II Transcription: Structure and Mechanism. *Biochim Biophys Acta*. 2013; 1829 (1): 2-8.
100. Wu JC, Chiang TY, Shiue WK, Wang SY, Sheen IJ, Huang YH et al. Recombination of hepatitis D virus RNA sequences and its implications. *Mol Biol Evol*. 1999; 16 (11): 1622-1632.
101. Chang J, Taylor J. In Vivo RNA-directed Transcription, With Template Switching, by a Mammalian RNA Polymerase. *EMBO J*. 2002; 21 (1-2): 157-164.
102. Giersch K, Allweiss L, Volz T, Helbig M, Bierwolf J, Lohse AW et al. Hepatitis Delta Co-Infection in Humanized Mice Leads to Pronounced Induction of Innate Immune Responses in Comparison to HBV Mono-Infection. *J Hepatol*. 2015; 63 (2): 346-353.
103. Dény P. Hepatitis delta virus genetic variability: from genotypes I, II, III to eight major clades? *Curr Top Microbiol Immunol*. 2006; 307: 151-71.
104. Homs M, Rodriguez-Frias F, Gregori J, Ruiz A, Reimundo P, Casillas R et al. Evidence of an Exponential Decay Pattern of the Hepatitis Delta Virus Evolution Rate and Fluctuations in Quasispecies Complexity in Long-Term Studies of Chronic Delta Infection. *PLoS One*. 2016; 11 (6): e0158557.
105. Domingo E, Martínez-Salas E, Sobrino F, de la Torre JC, Portela A, Ortín J et al. The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance--a review. *Gene*. 1985; 40 (1): 1-8.
106. Imazeki F, Omata M, Ohto M. Heterogeneity and evolution rates of delta virus RNA sequences. *J Virol*. 1990 ;64 (11): 5594-5599.
107. Wu JC, Chiang TY, Sheen IJ. Characterization and Phylogenetic Analysis of a Novel Hepatitis D Virus Strain Discovered by Restriction Fragment Length Polymorphism Analysis. *J Gen Virol*. 1998; 79 (Pt 5): 1105-1013.
108. Domingo E, Gomez J. Quasispecies and its impact on viral hepatitis. *Virus Res*. 2007; 127 (2): 131-150.

109. Wang SY, Wu JC, Chiang TY, Huang YH, Su CW, Sheen IJ. Positive selection of hepatitis delta antigen in chronic hepatitis D patients. *J Virol.* 2007; 81 (9): 4438-4444.
110. Abt I, Souquet M, Angot G, Mabon R, Dallot S, Thébaud G et al. Functional Transcomplementation Between Wheat Dwarf Virus Strains in Wheat and Barley. *Viruses.* 2019; 12 (1): 34.
111. Farci P, Niro GA. Clinical features of hepatitis D. *Semin Liver Dis.* 2012; 32 (3): 228–236.
112. Fattovich G, Giustina G, Christensen E, Pantalena M, Zagni I, Realdi G, et al. Influence of hepatitis delta virus infection on morbidity and mortality in compensated cirrhosis type B. *Gut.* 2000; 46 (3): 420–426.
113. Negro F. Hepatitis D virus coinfection and superinfection. *Cold Spring Harb Perspect Med.* 2014; 4 (11) :a021550.
114. Smedile A, Dentico P, Zanetti A, Sagnelli E, Nordenfelt E, Actis GC et al. Infection With the Delta Agent in Chronic HBsAg Carriers. *Gastroenterology.* 1981; 81 (6): 992-997.
115. Fattovich G, Boscaro S, Noventa F, Pornaro E, Stenico D, Alberti A et al. Influence of Hepatitis Delta Virus Infection on Progression to Cirrhosis in Chronic Hepatitis Type B. *J Infect Dis.* 1987; 155 (5): 931-935.
116. Pasetti G, Calzetti C, Degli Antoni A, Ferrari C, Penna A, Fiaccadori F. Clinical Features of Hepatitis Delta Virus Infection in a Northern Italian Area. *Infection.* 1988; 16 (6): 345-348.
117. Sagnelli E, Felaco FM, Filippini P, Pasquale G, Peinetti P, Buonagurio E et al. Influence of HDV Infection on Clinical, Biochemical and Histological Presentation of HBsAg Positive Chronic Hepatitis. *Liver.* 1989; 9 (4): 229-234.
118. Mathurin P, Thibault V, Kadidja K, Ganne-Carrié N, Moussalli J, El Younsi M et al. Replication Status and Histological Features of Patients With Triple (B, C, D) and Dual (B, C) Hepatic Infections. *J Viral Hepat.* 2000; 7 (1): 15-22.
119. Yurdaydin C, Idilman R, Bozkaya H, Bozdayi AM. Natural History and Treatment of Chronic Delta Hepatitis. *J Viral Hepat.* 2010; 17 (11): 749-756.
120. Giersch K, Dandri M. Hepatitis B and Delta Virus: Advances on Studies about Interactions between the Two Viruses and the Infected Hepatocyte. *J Clin Transl Hepatol.* 2015; 3 (3): 220-229.
121. Wu JC, Chen PJ, Kuo MY, Lee SD, Chen DS, Ting LP. Production of Hepatitis Delta Virus and Suppression of Helper Hepatitis B Virus in a Human Hepatoma Cell Line. *J Virol.* 1991; 65 (3): 1099-1104.
122. Lütgehetmann M, Mancke LV, Volz T, Helbig M, Allweiss L, Bornscheuer T et al. Humanized Chimeric uPA Mouse Model for the Study of Hepatitis B and D Virus Interactions and Preclinical Drug Evaluation. *Hepatology.* 2012 Mar;55(3):685-94. doi: 10.1002/hep.24758. Epub 2012 Jan 30.
123. Alfaiate D, Lucifora J, Abeywickrama-Samarakoon N, Michelet M, Testoni B, Cortay JC et al. HDV RNA Replication Is Associated With HBV Repression and

- Interferon-Stimulated Genes Induction in Super-Infected Hepatocytes. *Antiviral Res.* 2016; 136: 19-31.
124. Pollicino T, Raffa G, Santantonio T, Gaeta GB, Iannello G, Alibrandi A et al. Replicative and Transcriptional Activities of Hepatitis B Virus in Patients Coinfected With Hepatitis B and Hepatitis Delta Viruses. *J Virol.* 2011; 85 (1): 432-439.
 125. Schaper M, Rodriguez-Frias F, Jardi R, Tabernero D, Homs M, Ruiz G et al. Quantitative Longitudinal Evaluations of Hepatitis Delta Virus RNA and Hepatitis B Virus DNA Shows a Dynamic, Complex Replicative Profile in Chronic Hepatitis B and D. *J Hepatol.* 2010; 52 (5): 658-664.
 126. Raffa G, Niro GA, Raimondo G, Pollicino T. Hepatitis delta virus exerts an epigenetic control over hepatitis B virus transcription/replication activity. *Hepatology*, vol. 54, p. 1083A–1083A
 127. Williams V, Brichler S, Radjef N, Lebon P, Goffard A, Hober D et al. Hepatitis delta virus proteins repress hepatitis B virus enhancers and activate the alpha/beta interferon-inducible MxA gene. *J Gen Virol.* 2009; 90 (Pt 11): 2759-2767.
 128. Wang CC, Chang TC, Lin CW, Tsui HL, Chu PB, Chen BS et al. Nucleic Acid Binding Properties of the Nucleic Acid Chaperone Domain of Hepatitis Delta Antigen. *Nucleic Acids Res.* 2003; 31 (22): 6481-6492.
 129. Alves C, Cheng H, Tavanez JP, Casaca A, Gudima S, Roder H et al. Structural and Nucleic Acid Binding Properties of Hepatitis Delta Virus Small Antigen. *World J Virol.* 2017; 6 (2): 26-35.
 130. Chen M, Du D, Zheng W, Liao M, Zhang L, Liang G et al. Small Hepatitis Delta Antigen Selectively Binds to Target mRNA in Hepatic Cells: A Potential Mechanism by Which Hepatitis D Virus Downregulates Glutathione S-transferase P1 and Induces Liver Injury and Hepatocarcinogenesis. *Biochem Cell Biol.* 2019; 97 (2): 130-139.
 131. Tu T, Budzinska MA, Shackel NA, Urban S. HBV DNA Integration: Molecular Mechanisms and Clinical Implications. *Viruses.* 2017; 9 (4): 75.
 132. Freitas N, Lukash T, Gunewardena S, Chappell B, Slagle BL, Gudima SO. Relative Abundance of Integrant-Derived Viral RNAs in Infected Tissues Harvested From Chronic Hepatitis B Virus Carriers. *J Virol.* 2018; 92 (10): e02221-17.
 133. Hu B, Wang R, Fu J, Su M, Du M, Liu Yet al. Integration of Hepatitis B Virus S Gene Impacts on Hepatitis B Surface Antigen Levels in Patients With Antiviral Therapy. *J Gastroenterol Hepatol.* 2018; 33 (7): 1389-1396.
 134. Freitas N, Cunha C, Menne S, Gudima SO. Envelope Proteins Derived From Naturally Integrated Hepatitis B Virus DNA Support Assembly and Release of Infectious Hepatitis Delta Virus Particles. *J Virol.* 2014; 88 (10): 5742-5754.
 135. Goto T, Kato N, Yoshida H, Otsuka M, Moriyama M et al. Synergistic Activation of the Serum Response Element-Dependent Pathway by Hepatitis B Virus X Protein and Large-Isoform Hepatitis Delta Antigen. *J Infect Dis.* 2003; 187 (5): 820-828.

136. Asselah T, Loureiro D, Tout I, Castelnau C, Boyer N, Marcellin P et al. Future Treatments for Hepatitis Delta Virus Infection. *Liver Int.* 2020; 40 Suppl 1: 54-60.
137. Yurdaydin C, Abbas Z, Buti M, Cornberg M, Esteban R, Etzion O et al. Treating Chronic Hepatitis Delta: The Need for Surrogate Markers of Treatment Efficacy. *J Hepatol.* 2019; 70 (5): 1008-1015.
138. Rizzetto M, Rosina F, Saracco G, Bellando PC, Actis GC, Bonino F et al. Treatment of chronic delta hepatitis with alpha-2 recombinant interferon. *J Hepatol.* 1986; 3 Suppl 2: S229-33.
139. Wedemeyer H, Manns MP. Epidemiology, pathogenesis and management of hepatitis D: update and challenges ahead. *Nat Rev Gastroenterol Hepatol.* 2010; 7 (1): 31-40.
140. Wranke A, Heidrich B, Hardtke S, Wedemeyer H. Current Management of HBV/HDV Coinfection and Future Perspectives. *A. Curr Hepatol Rep.* 2015; 14: 284–292
141. Heller T, Rotman Y, Koh C, Clark S, Haynes-Williams V, Chang R et al. Long-term Therapy of Chronic Delta Hepatitis With Peginterferon Alfa. *Aliment Pharmacol Ther.* 2014; 40 (1): 93-104.
142. Wedemeyer H, Yurdaydin C, Dalekos GN, Erhardt A, Çakaloğlu Y, Değertekin H et al. Peginterferon plus adefovir versus either drug alone for hepatitis delta. *N Engl J Med.* 2011; 364 (4): 322-331.
143. Heidrich B, Yurdaydin C, Kabaçam G, Ratsch BA, Zachou K, Bremer B et al. Late HDV RNA relapse after peginterferon alpha-based therapy of chronic hepatitis delta. *Hepatology.* 2014; 60 (1): 87-97.
144. Wedemeyer H, Yurdaydin C, Ernst S, Caruntu FA, Curescu MG, Yalcin K et al. Prolonged therapy of hepatitis delta for 96 weeks with pegylated-interferon- α -2a plus tenofovir or placebo does not prevent HDV-RNA relapse after treatment: the HIDIT-2 study. *J Hepatol.* 2014; 60 (1): Suppl S2-S3.
145. Kaymakoglu S, Karaca C, Demir K, Poturoglu S, Danalioglu A, Badur S et al. Alpha interferon and ribavirin combination therapy of chronic hepatitis D. *Antimicrob Agents Chemother.* 2005; 49 (3): 1135-1138.
146. Niro GA, Ciancio A, Gaeta GB, Smedile A, Marrone A, Olivero A et al. Pegylated interferon alpha-2b as monotherapy or in combination with ribavirin in chronic hepatitis delta. *Hepatology.* 2006; 44 (3): 713-720.
147. Raison CL, Demetrashvili M, Capuron L, Miller AH. Neuropsychiatric adverse effects of interferon-alpha: recognition and management. *CNS Drugs.* 2005; 19 (2): 105-123.
148. Pugnale P, Paziienza V, Guilloux K, Negro F. Hepatitis delta virus inhibits alpha interferon signaling. *Hepatology.* 2009; 49 (2): 398-406.
149. Han Z, Nogusa S, Nicolas E, Balachandran S, Taylor J. Interferon impedes an early step of hepatitis delta virus infection. *PLoS One.* 2011; 6 (7): e22415.
150. Wedemeyer H. Hepatitis D revival. *Liver Int.* 2011; 31 Suppl 1: 140-144.

151. Ciancio A, Rizzetto M. Chronic hepatitis D at a standstill: where do we go from here? *Nat Rev Gastroenterol Hepatol*. 2014; 11 (1): 68-71.
152. Gilman C, Heller T, Koh C. Chronic Hepatitis Delta: A State-Of-The-Art Review and New Therapies. *World J Gastroenterol*. 2019; 25 (32): 4580-4597.
153. Li W, Urban S. Entry of hepatitis B and hepatitis D virus into hepatocytes: Basic insights and clinical implications. *J Hepatol*. 2016; 64 (Suppl 1): S32-S40.
154. Gripon P, Canine I, Urban S. Efficient inhibition of hepatitis B virus infection by acylated peptides derived from the large viral surface protein. *J Virol*. 2005; 79 (3): 1613-1622.
155. Volz T, Allweiss L, Ben MBarek M, Warlich M, Lohse AW, Pollok JM et al. The entry inhibitor Myrcludex-B efficiently blocks intrahepatic virus spreading in humanized mice previously infected with hepatitis B virus. *J Hepatol*. 2013; 58 (5): 861-867.
156. Bogomolov P, Alexandrov A, Voronkova N, Macievich M, Kokina K, Petrachenkova M et al. Treatment of chronic hepatitis D with the entry inhibitor myrcludex B: First results of a phase Ib/IIa study. *J Hepatol*. 2016; 65 (3): 490-498.
157. Wedemeyer H, Bogomolov P, Blank A, Allweiss L, Dandri-Petersen M, Bremer B et al. Final results of a multicenter, open-label phase 2b clinical trial to assess safety and efficacy of Myrcludex B in combination with Tenofovir in patients with chronic HBV/HDV co-infection. *J Hepatol*. 2018; 68 (Suppl 1): S3
158. Allweiss L, Dettmer C, Volz T, Giersch K, Alexandrov A, Wedemeyer H et al. Strong intrahepatic decline of hepatitis D virus RNA and antigen after 24 weeks of treatment with Myrcludex B in combination with Tenofovir in chronic HBV/HDV infected patients: Interim results from a multicenter, open-label phase 2b clinical trial. *J Hepatol*. 2018; 68 (Suppl 1): S90
159. Wedemeyer H, Schöneweis K, Bogomolov P, Voronkova N, Chulanov V, Stepanova et al. Interim results of a multicenter, open-label phase 2 clinical trial (MYR203) to assess safety and efficacy of Myrcludex B in combination with PEG-IFN α in patients with chronic HBV/HDV co-infection. *AASLD 2018*.
160. Wedemeyer H, Schöneweis K, Bogomolov P, Vorokova N, Chulanov V, Stepanova T et al. Safety and Efficacy of 10 mg Myrcludex B/IFN α Combination Therapy in Patients with Chronic HBV/HDV Co-Infection. *Z Gastroenterol*. 2020; 58(01): e2-e3
161. Giersch K, Bhadra OD, Volz T, Allweiss L, Riecken K, Fehse B et al. Hepatitis delta virus persists during liver regeneration and is amplified through cell division both in vitro and in vivo. *Gut*. 2019; 68 (1): 150-157.
162. Koh C, Da BL, Glenn JS. HBV/HDV Coinfection: A Challenge for Therapeutics. *Clin Liver Dis*. 2019; 23 (3): 557-572.
163. Koh C, Canini L, Dahari H, Zhao X, Uprichard SL, Haynes-Williams V et al. Oral prenylation inhibition with lonafarnib in chronic hepatitis D infection: a proof-of-concept randomised, double-blind, placebo-controlled phase 2A trial. *Lancet Infect Dis*. 2015; 15 (10): 1167-1174.

164. Yurdaydin C, Idilman R, Choong I, Kalkan C, Keskin O, Karakaya MF et al. Optimizing the prenylation inhibitor lonafarnib using ritonavir boosting in patients with chronic delta hepatitis. *J Hepatol.* 2015; 62 (Suppl 2): S252.
165. Rizzetto M, Ciancio A. The Prenylation Inhibitor, Lonafarnib: A New Therapeutic Strategy Against Hepatitis Delta. *Lancet Infect Dis.* 2015; 15 (10): 1119-1120.
166. Ghosal A, Chowdhury SK, Tong W, Hapangama N, Yuan Y, Su AD et al. Identification of human liver cytochrome P450 enzymes responsible for the metabolism of lonafarnib (Sarasar). *Drug Metab Dispos.* 2006; 34 (4): 628-635.
167. Yurdaydin C, Keskin O, Kalkan Ç, Karakaya F, Çalışkan A, Karatayli E et al. Optimizing lonafarnib treatment for the management of chronic delta hepatitis: The LOWR HDV-1 study. *Hepatology.* 2018; 67 (4): 1224-1236.
168. Koh C. A phase 2 study of Lonafarnib, ritonavir and peginterferon lambda for 24 weeks :interim end-of treatment results from the Lift HDV study. *AASLD* 2019.
169. Vaillant A. Nucleic acid polymers: Broad spectrum antiviral activity, antiviral mechanisms and optimization for the treatment of hepatitis B and hepatitis D infection. *Antiviral Res.* 2016; 133: 32-40.
170. Vaillant A, Juteau JM, Lu H, Liu S, Lackman-Smith C, Ptak R et al. Phosphorothioate Oligonucleotides Inhibit Human Immunodeficiency Virus Type 1 Fusion by Blocking gp41 Core Formation *Antimicrob Agents Chemother.* 2006; 50 (4): 1393-1401.
171. Beilstein F, Blanchet M, Vaillant A, Sureau C. Nucleic Acid Polymers Are Active against Hepatitis Delta Virus Infection *In Vitro.* *J Virol.* 2018; 92 (4) pii: e01416-17.
172. Noordeen F, Vaillant A, Jilbert AR. Nucleic Acid Polymers Prevent the Establishment of Duck Hepatitis B Virus Infection *In Vivo.* *Antimicrob Agents Chemother.* 2013 ;57 (11): 5299-5306.
173. Noordeen F, Scougall AC, Grosse A, Qiao Q, Ajilian BB, Reaiche-Miller G et al. Therapeutic Antiviral Effect of the Nucleic Acid Polymer REP 2055 Against Persistent Duck Hepatitis B Virus Infection. *PLoS One.* 2015; 10 (11): e0140909.
174. Guillot C, Martel N, Berby F, Bordes I, Hantz O, Blanchet M et al. Inhibition of hepatitis B viral entry by nucleic acid polymers in HepaRG cells and primary human hepatocytes. *PLoS One.* 2017; 12 (6): e0179697.
175. Blanchet M, Sinnathamby V, Vaillant A, Labonté L. Inhibition of HBsAg Secretion by Nucleic Acid Polymers in HepG2.2.15 Cells. *Antiviral Res.* 2019; 164:97-105.
176. Shamur MM, Mayer R, Vaillant A. Interaction of nucleic acid polymers with the large and small forms of the hepatitis delta antigen protein. *Hepatology.* 2017; 66:504A.
177. Bazinet M, Pantea V, Ceboatarescu V, Cojuhari L, Jimbei P, Vaillant A. REP 2139 monotherapy and combination therapy with pegylated interferon: Safety and

- potent reduction of HBsAg and HDV RNA in Caucasian patients with chronic HBV/HDV co-infection. *J Viral Hepat.* 2015; 22 (S2): 5-6
178. Poutay D, Sabra M, Abou Jaoudé G, Chemin I, Trepo C, Vaillant A et al. Nucleic acid polymers are efficient in blocking hepatitis delta virus entry in vitro. *J Hepatol.* 2015; 62 (Suppl 2): S276.
 179. Bazinet M, Pantea V, Ceboatarescu V, Cojuhari L, Jimbei P, Albrecht J et al. Update on the Safety and Efficacy of REP 2139 Monotherapy and Subsequent Combination Therapy with Pegylated Interferon Alpha-2A in Caucasian Patients with Chronic HBV/HDV Co-Infection. *J Hepatol.* 2016; 64 (2): S584-S585.
 180. Bazinet M, Pântea V, Ceboatarescu V, Cojuhari L, Jimbei P, Albrecht J et al. Safety and efficacy of REP 2139 and pegylated interferon alfa-2a for treatment-naive patients with chronic hepatitis B virus and hepatitis D virus co-infection (REP 301 and REP 301-LTF): a non-randomised, open-label, phase 2 trial. *Lancet Gastroenterol Hepatol.* 2017; 2 (12): 877-889.
 181. Bazinet M, Pantea V, Ceboatarescu V, Cojuhari L, Jimbei P, Vaillant A. Establishment of persistent functional remission of HBV and HDV infection following REP 2139 and pegylated interferon alpha 2a therapy in patients with chronic HBV/HDV co-infection: 18 month follow-up results from the REP 301-LTF. *J Hepatol.* 2018; 68 (Suppl 1): S509.
 182. Chan HL, Ahn SH, Chang TT, Peng CY, Wong D et al. Peginterferon Lambda for the Treatment of HBeAg-positive Chronic Hepatitis B: A Randomized Phase 2b Study (LIRA-B). *J Hepatol.* 2016; 64 (5): 1011-1019.
 183. Foster GR, Chayama K, Chuang WL, Fainboim H, Farkkila M, Adrian Gadano A et al. A Randomized, Controlled Study of Peginterferon lambda-1a/ribavirin ± Daclatasvir for Hepatitis C Virus Genotype 2 or 3. *Springerplus.* 2016; 5 (1): 1365.
 184. Pagliaccetti NE, Chu EN, Bolen CR, Kleinstein SH, Robek MD. Lambda and Alpha Interferons Inhibit Hepatitis B Virus Replication Through a Common Molecular Mechanism but With Different in Vivo Activities. *Virology.* 2010; 401 (2): 197-206.
 185. Radaeva S, Sun R, Pan HN, Hong F, Gao B. Interleukin 22 (IL-22) Plays a Protective Role in T Cell-Mediated Murine Hepatitis: IL-22 Is a Survival Factor for Hepatocytes via STAT3 Activation. *Hepatology.* 2004; 39 (5): 1332-1342.
 186. Zenewicz LA, Yancopoulos GD, Valenzuela DM, Murphy AJ, Karow M, Flavell RA. Interleukin-22 but Not interleukin-17 Provides Protection to Hepatocytes During Acute Liver Inflammation. *Immunity.* 2007; 27 (4): 647-659.
 187. Giersch K, Homs M, Volz T, Helbig M, Allweiss L, Lohse AW. Both Interferon Alpha and Lambda Can Reduce All Intrahepatic HDV Infection Markers in HBV/HDV Infected Humanized Mice. *Sci Rep.* 2017; 7 (1): 3757.
 188. Etzion O, Hamid SS, Lurie Y, Gane E, Bader N, Yardeni D. End of study results from LIMT HDV study: 36% durable virologic response at 24 weeks post-treatment with pegylated interferon lambda monotherapy in patients with chronic hepatitis delta virus infection. *J Hepatol.* 2019; 70 (1): E32.

189. Chang J, Taylor JM. Susceptibility of Human Hepatitis Delta Virus RNAs to Small Interfering RNA Action. *J Virol*. 2003; 77 (17): 9728-9731.
190. Yen L, Magnier M, Weissleder R, Stockwell BR, Richard C Mulligan RC. Identification of Inhibitors of Ribozyme Self-Cleavage in Mammalian Cells via High-Throughput Screening of Chemical Libraries. *RNA*. 2006; 12 (5): 797-806.
191. Singh S, Gupta SK, Nischal A, Khattri S, Nath R, Pant KK et al. Design of potential siRNA molecules for hepatitis delta virus gene silencing. *Bioinformatics*. 2012; 8 (16): 749-757.
192. Stokowa-Sołtys K, Gaggelli N, Nagaj J, Szczepanik W, Ciesiołka J, Wrzesinski J et al. High affinity of copper(II) towards amoxicillin, apramycin and ristomycin. Effect of these complexes on the catalytic activity of HDV ribozyme. *J Inorg Biochem*. 2013; 124: 26-34.
193. Rogers J, Chang AH, Von Ahsen U, Davies J. Inhibition of the self-cleavage reaction of the human hepatitis delta virus ribozyme by antibiotics. *J Mol Biol*. 1996; 259 (5): 916-925.
194. Abbas Z, Abbas M. Management of hepatitis delta: need for novel therapeutic options. *World J Gastroenterol*. 2015; 21 (32): 9461-9465.
195. Li H, Mao Q, Li Q. Inhibitory effect of replication and expression of HDV by antisense oligodeoxynucleotides in H1 delta 9 cell. *Zhonghua Gan Zang Bing Za Zhi* 1999, 7 (1): 13-14.
196. Gong L, Han Y, Chen L, Liu F, Hao P, Sheng J et al. Comparison of next-generation sequencing and clone-based sequencing in analysis of hepatitis B virus reverse transcriptase quasispecies heterogeneity. *J. Clin. Microbiol*. 2013; 51 (12): 4087–4094.
197. Ramírez C, Gregori J, Buti M, Tabernero D, Camós S, Casillas R et al. A comparative study of ultra-deep pyrosequencing and cloning to quantitatively analyze the viral quasispecies using hepatitis B virus infection as a model. *Antiviral Res*. 2013; 98 (2): 273–283
198. Homs M, Buti M, Quer J, Jardí R, Schaper M, Tabernero D et al. Ultra-deep pyrosequencing analysis of the hepatitis B virus preCore region and main catalytic motif of the viral polymerase in the same viral genome. *Nucleic Acids Res*. 2011; 39 (19): 8457–8471.
199. Homs M, Caballero A, Gregori J, Tabernero D, Quer J, Nieto L et al. Clinical application of estimating hepatitis B virus quasispecies complexity by massive sequencing: correlation between natural evolution and on-treatment evolution. *PLoS One*. 2014; 9 (11): e112306.
200. Roche Diagnostics [Internet]. F. Hoffmann - La Roche Ltd; [cited 2020 Mar 13]. Available from: <https://diagnostics.roche.com/global/en/products/params/cobas-hbv-test.html#productSpecs>
201. Roche Diagnostics [Internet]. F. Hoffmann - La Roche Ltd; [cited 2020 Mar 13]. Available from: <https://diagnostics.roche.com/global/en/products/params/electsys-hbsag-ii.html>

202. Roche Diagnostics [Internet]. F. Hoffmann - La Roche Ltd; [cited 2020 Mar 13]. Available from: <https://diagnostics.roche.com/global/en/products/params/electsys-hbe-ag.html>
203. DiaSorn [Internet]. Saluggia: DiaSorin; [cited 2020 Mar 13]. Available from: <http://hdv.diasorin.com/ese%20antiHDV%204330-A%2008-2018%20low.pdf>
204. LifeScience Roche [Internet]. Roche Molecular Systems, Inc; c1999 [cited 2020 Mar 13]. Available from: https://www.lifescience.roche.com/en_es/products/magna-pure-lc-total-nucleic-acid-isolation-kit.html#details
205. Sequencing Roche [Internet]. Roche Sequencing Corporate; [cited 2020 Mar 13]. Available from: <https://sequencing.roche.com/content/dam/rochesequence/US/Resources/PDFs/DNALibraryPrep/Brochure-%20SeqSuccess-WGS-Application-MC-US-01350.pdf>
206. Homs M, Giersch K, Blasi M, Lütgehetmann M, Buti M, Esteban R, et al. Relevance of a full-length genomic RNA standard and a thermal-shock step for optimal hepatitis delta virus quantification. *J Clin Microbiol.* 2014; 52: 3334-3338.
207. Paul-Ehrlich-Institute [Internet]. © Paul-Ehrlich-Institut; [cited 2020 Mar 13]. Available from: https://www.pei.de/SharedDocs/Downloads/EN/regulation-en/referencematerial/7657-12-ifu.pdf?__blob=publicationFile&v=3.
208. Salazar AM, Sandoval AS, Armendariz JS. *Electroforesis. Biología Molecular.* Mc Graw Hill; 2013.
209. QIAGEN [Internet]. QIAGEN; c2013 [cited 2020 Mar 14]. Available from: <https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/dna-clean-up/qiaquick-gel-extraction-kit/?clear=true#orderinginformation>
210. Agilent [Internet]. Santa Clara: Agilent Technologies, Inc; [cited 2020 Mar 14]. Available from: https://www.agilent.com/cs/library/usermanuals/public/ScreenTape_D1000_QG.pdf
211. Thermo Fisher Scientific [Internet]. ThermoFisher Scientific, Inc; [cited 2020 Mar 14]. Available from: <https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-Assets%2FLSG%2Fmanuals%2Fmp07581.pdf&title=UXVhbnQtaVQgUGljb0dyZVVuIGRzRE5BIFJlYWdlbnQgYW5kIEtpdHM=>
212. Thermo Fisher Scientific [Internet]. ThermoFisher Scientific, Inc; [cited 2020 Mar 14]. Available from: <http://tools.thermofisher.com/content/sfs/manuals/nd-1000-v3.8-users-manual-8%205x11.pdf>
213. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* 2016; 17 (6): 333-351
214. Metzker ML. Sequencing technologies - the next generation. *Nat. Rev. Genet.* 2010; 11: 31-46

215. Illumina Support Center [Internet]. Illumina, Inc; [cited 2020 Mar 14]. Available from: https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/translations/miseq-system-guide-1000000061014-esp.pdf
216. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005; 437: 376–380
217. Roche Sequencing Store [Internet]. F. Hoffmann - La Roche Ltd; [cited 2020 Mar 14]. Available from: https://rochesequencingstore.com/wp-content/uploads/2017/10/KAPA-Pure-Beads_KR1245-%E2%80%93v4.17.pdf
218. Roche Sequencing Store [Internet]. F. Hoffmann - La Roche Ltd; [cited 2020 Mar 14]. Available from: https://rochesequencingstore.com/wp-content/uploads/2017/10/KAPA-HiFi-HotStart-ReadyMix-PCR-Kit_KR0370-%E2%80%93v10.19.pdf
219. Roche Sequencing Store [Internet]. F. Hoffmann - La Roche Ltd; [cited 2020 Mar 14]. Available from: <https://sequencing.roche.com/en/products-solutions/by-category/library-preparation/library-quantification/kapa-library-quantification-kits.html>
220. Illumina Support Center [Internet]. Illumina, Inc; [cited 2020 Mar 14]. Available from: <https://support.illumina.com/bulletins/2017/02/what-is-the-phix-control-v3-library-and-what-is-its-function-in-.html>
221. Illumina [Internet]. Illumina, Inc; [cited 2020 Mar 14]. Available from: <https://www.illumina.com/products/by-type/sequencing-kits/cluster-gen-sequencing-reagents/miseq-reagent-kit-v3.html>
222. Illumina [Internet]. Illumina, Inc; [cited 2020 Mar 14]. Available from: https://www.well.ox.ac.uk/ogc/wp-content/uploads/2017/09/Illumina_Sequencing_Overview_15045845_D.pdf
223. Gregori J, Esteban JI, Cubero M, Garcia-Cehic D, Perales C, Casillas R et al. Ultra-deep pyrosequencing (UDPS) data treatment to study amplicon HCV minor variants. *PLoS One*. 2013; 8 (12): e83361
224. Gregori J, Salicrú M, Domingo E, Sanchez A, Esteban JI, Rodríguez-Frías F et al. Inference with viral quasispecies diversity indices: clonal and NGS approaches. *Bioinformatics*. 2014; 30(8): 1104-1111
225. Godoy C, Taberner D, Sopena S, Gregori J, Cortese MF, González C et al. Characterization of hepatitis B virus X gene quasispecies complexity in mono-infection and hepatitis delta virus superinfection. *World J Gastroenterol*. 2019; 25 (13): 1566-1579
226. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 2011; 27 (21): 2957-2963
227. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res*. 1998; 8 (3): 186-194
228. Cuadras CM. A distance based approach to discriminant analysis and its properties. *Univ. de Barcelona Math. Preprint series*. 1991; 90

229. Cuadras CM. Distance analysis in discrimination and classification using both continuous and categorical variables. In: *Statistical Data Analysis and Inference*. North Holland. 1989; 459-473
230. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*. 1980; 16 (2): 111-20
231. Bernard P, Couturier M. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J Mol Biol*. 1992; 226 (3): 735-45
232. Bernard P, Gabant P, Bahassi EM, Couturier M. Positive-selection vectors using the F plasmid ccdB killer gene. *Gene*. 1994; 148 (1): 71-74
233. Bernard P, Kézdy KE, Van Melderen L, Steyaert J, Wyns L, Pato ML et al. The F plasmid CcdB protein induces efficient ATP-dependent DNA cleavage by gyrase. *J Mol Biol*. 1993; 234 (3): 534-541.
234. Thermo Fisher Scientific [Internet]. ThermoFisher Scientific, Inc; [cited 2020 Mar 14]. Available from: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/topobluntseq_man.pdf
235. Thermo Fisher Scientific [Internet]. ThermoFisher Scientific, Inc; [cited 2020 Mar 14]. Available from: https://tools.thermofisher.com/content/sfs/vectors/pcr4blunttopo_map.pdf
236. QIAGEN [Internet]. QIAGEN; c2013 [cited 2020 Mar 14]. Available from: <https://www.qiagen.com/us/products/top-sellers/qiaprep-spin-miniprep-kit/>
237. Thermo Fisher Scientific [Internet]. ThermoFisher Scientific, Inc; [cited 2020 Mar 14]. Available from: <https://www.thermofisher.com/order/catalog/product/4376485#/4376485>
238. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977; 74 (12): 5463-5467.
239. Smith LM, Fung S, Hunkapiller MW, Hunkapiller TJ, Hood LE. The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis. *Nucleic Acids Res*. 1985; 13 (7): 2399-2412
240. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR et al. Fluorescence detection in automated DNA sequence analysis. *Nature*. 1986; 321 (6071): 674-679
241. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 1999; 41: 95-98
242. Zoulim F, Locarnini S. Management of treatment failure in chronic hepatitis B. *J Hepatol*. 2012; 56 Suppl 1: S112-22.
243. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2017. URL <https://www.R-project.org/>.

244. Schneider TD. Information content of individual genetic sequences. *J Theor Biol.* 1997; 189 (4): 427-441.
245. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 1990; 18 (20): 6097-6100
246. Shakil AO, Hadziyannis S, Hoofnagle JH, Di Bisceglie AM, Gerin JL, Casey JL. Geographic distribution and genetic variability of hepatitis delta virus genotype I. *Virology.* 1997; 234 (1): 160-167.
247. Suzuki R and Shimodaira H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics.* 2006; 22 (12): 1540-1542.
248. Suzuki R and Shimodaira H. Pvcust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling. 2015.
249. Felsenstein J. Confidence limits of phylogenies: An approach using the bootstrap. *Evolution.* 1985; 39 (4): 783-791.
250. Shimodaira H. Another calculation of the p-value for the problem of regions using the scaled bootstrap resampling. Tech. Rep. No. 2000-35. 2000.
251. Shirvani-Dastgerdi E, Amini-Bavil-Olyaei S, Alavian SM, Trautwein C, Tacke F. Comprehensive analysis of mutations in the hepatitis delta virus genome based on full-length sequencing in a nationwide cohort study and evolutionary pattern during disease progression. *Clin Microbiol Infect.* 2015; 21 (5): 510.e11-23.
252. Chang MF, Sun CY, Chen CJ, Chang SC. Functional motifs of delta antigen essential for RNA binding and replication of hepatitis delta virus. *J Virol.* 1993; 67 (5): 2529–2536.
253. Lee CZ, Lin JH, Chao M, McKnight K, Lai MM. RNA-binding Activity of Hepatitis Delta Antigen Involves Two Arginine-Rich Motifs and Is Required for Hepatitis Delta Virus RNA Replication. *J Virol.* 1993; 67 (4): 2221-2227.
254. Chang MF, Chen CH, Lin SL, Chen CJ, Chang SC. Functional Domains of Delta Antigens and Viral RNA Required for RNA Packaging of Hepatitis Delta Virus. *J Virol.* 1995; 69 (4): 2508-2514.
255. Tabernero D, Cortese MF, Buti M, Rodriguez-Frias F. HDV evolution-will viral resistance be an issue in HDV infection? *Curr Opin Virol.* 2018; 32: 100-107.
256. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol.* 1975; 94 (3): 441-448.
257. Lim SG, Cheng Y, Guindon S, Seet BL, Lee LY, Hu P et al. Viral quasi-species evolution during hepatitis Be antigen seroconversion. *Gastroenterology.* 2007; 133 (3): 951-958.
258. Rodriguez-Frias F, Buti M, Tabernero D, Homs M. Quasispecies structure, cornerstone of hepatitis B virus infection: mass sequencing approach. *World J. Gastroenterol.* 2013; 19 (41): 6995–7023.
259. Rodriguez-Frías F, Tabernero D, Quer J, Esteban JI, Ortega I, Domingo E et al. Ultra-deep pyrosequencing detects conserved genomic sites and quantifies linkage of drug-resistant amino acid changes in the hepatitis B virus genome. *PLoS One.* 2012; 7(5): e37874.

260. Rodriguez-Frias F, Tabernero D, Esteban R, Buti M. Study of hepatitis B virus quasispecies by ultrJ Virol
261. Macnaughton TB, Li YI, Doughty AL, Lai MM. Hepatitis Delta Virus RNA Encoding the Large Delta Antigen Cannot Sustain Replication Due to Rapid Accumulation of Mutations Associated With RNA Editing. *J Virol.* 2003; 77 (22): 12048-12056.
262. Cullen BR. Role and Mechanism of Action of the APOBEC3 Family of Antiretroviral Resistance Factors. *J Virol.* 2006; 80 (3): 1067–1076.
263. Polson AG, Ley HL, Bass BL, Casey JL. Hepatitis Delta Virus RNA Editing Is Highly Specific for the amber/W Site and Is Suppressed by Hepatitis Delta Antigen. *Mol Cell Biol.* 1998; 18 (4): 1919-1926.
264. Perveen S, Nasir MI, Shahid SM, Azhar A, Khan OY. Phylogenetic Analysis of HDV Isolates From HBsAg Positive Patients in Karachi, Pakistan. *Virology.* 2012; 9: 162.
265. Ghamari S, Alavian SM, Rizzetto M, Olivero A, Smedile A, Khedive A et al. Prevalence of Hepatitis Delta Virus (HDV) Infection in Chronic Hepatitis B Patients With Unusual Clinical Pictures. *Hepat Mon.* 2013; 13 (8): e6731.
266. Yacoubi L, Bricler S, Le Gal F, Triki H, Gordien E. Molecular Characterization of the Full-Length Genome Sequences of HDV Strains Circulating in Tunisia. *Arch Virol.* 2018; 163 (7): 1727-1731.
267. Wang JG, Jansen RW, Brown EA, Lemon SM. Immunogenic Domains of Hepatitis Delta Virus Antigen: Peptide Mapping of Epitopes Recognized by Human and Woodchuck Antibodies. *J Virol.* 1990; 64 (3): 1108-1116.
268. Lee CH, Chang SC, Wu CH, Chang MF. A novel chromosome region maintenance 1-independent nuclear export signal of the large form of hepatitis delta antigen that is required for the viral assembly. *J Biol Chem.* 2001; 276 (11): 8142-8148.
269. Shih HH, Shih C, Wang HW, Su CW, Sheen IJ, Wu JC. Pro-205 of Large Hepatitis Delta Antigen and Pro-62 of Major Hepatitis B Surface Antigen Influence the Assembly of Different Genotypes of Hepatitis D Virus. *J Gen Virol.* 2010; 91 (Pt 4): 1004-1012.
270. Behzadian F, Sabahi F, Karimi M, Sadeghizadeh M, Maghsoudi N, Forooshani RM et al. Molecular Phylogenetic Analysis of Iranian HDV Complete Genome. *Virus Genes.* 2005; 30 (3): 383-393.
271. Anisimova M, Yang Z. Molecular Evolution of the Hepatitis Delta Virus Antigen Gene: Recombination or Positive Selection? *J Mol Evol.* 2004; 59 (6): 815-826.
272. Golden BL. Two Distinct Catalytic Strategies in the Hepatitis δ Virus Ribozyme Cleavage Reaction. *Biochemistry.* 2011; 50 (44): 9424-9433.
273. Thill G, Vasseur M, Tanner NK. Structural and Sequence Elements Required for the Self-Cleaving Activity of the Hepatitis Delta Virus Ribozyme. *Biochemistry.* 1993; 32 (16): 4254-4262.

274. Been MD, Wickham GS. Self-cleaving ribozymes of hepatitis delta virus RNA. *Eur J Biochem.* 1997; 247 (3): 741-753.
275. Perrotta AT, Been MD. Core Sequences and a Cleavage Site Wobble Pair Required for HDV Antigenomic Ribozyme Self-Cleavage. *Nucleic Acids Res.* 1996; 24 (7): 1314-1321.
276. Perrotta AT, Been MD. The Self-Cleaving Domain From the Genomic RNA of Hepatitis Delta Virus: Sequence Requirements and the Effects of Denaturant. *Nucleic Acids Res.* 1990; 18 (23): 6821-6827.
277. Tanner NK, Schaff S, Thill G, Petit-Koskas E, Crain-Denoyelle AM, Westhof E. A three-dimensional model of hepatitis delta virus ribozyme based on biochemical and mutational analyses. *Curr Biol.* 1994; 4 (6): 488-498.
278. Perrotta AT, Been MD. A Pseudoknot-Like Structure Required for Efficient Self-Cleavage of Hepatitis Delta Virus RNA. *Nature.* 1991; 350 (6317): 434-436.
279. Rosenstein SP, Been MD. Hepatitis Delta Virus Ribozymes Fold to Generate a Solvent-Inaccessible Core With Essential Nucleotides Near the Cleavage Site Phosphate. *Biochemistry.* 1996; 35 (35): 11403-11413.
280. Belinsky MG, Britton E, Dinter-Gottlieb G. Modification Interference Analysis of a Self-Cleaving RNA From Hepatitis Delta Virus. *FASEB J.* 1993; 7 (1): 130-136.
281. Alam S, Grum-Tokars V, Krucinska J, Kundracik ML, Wedekind JE. Conformational Heterogeneity at Position U37 of an all-RNA Hairpin Ribozyme With Implications for Metal Binding and the Catalytic Structure of the S-turn. *Biochemistry.* 2005; 44 (44): 14396-14408.
282. Perrotta AT, Wadkins TS, Been MD. Chemical rescue, multiple ionizable groups, and general acid–base catalysis in the HDV genomic ribozyme. *RNA.* 2006; 12 (7): 1282–1291.
283. Kumar KP, Suh YA, Miyashiro H, Nishikawa F, Kawakami J, Taira K et al. Random mutations to evaluate the role of bases at two important single-stranded regions of genomic HDV ribozyme. *Nucleic Acids Res.* 1992; 20 (15): 3919–3924.
284. Puttaraju M, Perrotta AT, Been MD. A circular trans-acting hepatitis delta virus ribozyme. *Nucleic Acids Res.* 1993; 21 (18): 4253–4258.
285. Calle-Serrano B, Großhennig A, Homs M, Heidrich B, Erhardt A, Deterding K et al. Development and Evaluation of a Baseline-Event-Anticipation Score for Hepatitis Delta. *J Viral Hepat.* 2014; 21 (11): e154-63.

ANEXO

Esta tesis ha dado lugar a la siguiente publicación:

Autores: Sopena S, Godoy C, Tabernero D, Homs M, Gregori J, Riveiro-Barciela M, Ruiz A, Esteban R, Buti M, Rodríguez-Frías F.

Título: Quantitative characterization of hepatitis delta virus genome edition by next generation sequencing.

Revista: Virus Research

Fecha: Enero (2018)

Volumen: 243

Páginas: 52-59

Autores: Sopena S, Godoy C, Tabernero D, Homs M, Gregori J, Riveiro-Barciela M, Ruiz A, Esteban R, Buti M, Rodríguez-Frías F.

Título: Quantitative characterization of hepatitis delta virus genome edition by next generation sequencing.

Revista: Virus Research

Fecha: Enero (2018)

Volumen: 243

Páginas: 52-59

Autores: Sopena S, Godoy C, Tabernero D, Homs M, Gregori J, Riveiro-Barciela M, Ruiz A, Esteban R, Buti M, Rodríguez-Frías F.

Título: Quantitative characterization of hepatitis delta virus genome edition by next generation sequencing.

Revista: Virus Research

Fecha: Enero (2018)

Volumen: 243

Páginas: 52-59

Autores: Sopena S, Godoy C, Tabernero D, Homs M, Gregori J, Riveiro-Barciela M, Ruiz A, Esteban R, Buti M, Rodríguez-Frías F.

Título: Quantitative characterization of hepatitis delta virus genome edition by next generation sequencing.

Revista: Virus Research

Fecha: Enero (2018)

Volumen: 243

Páginas: 52-59

Autores: Sopena S, Godoy C, Tabernero D, Homs M, Gregori J, Riveiro-Barciela M, Ruiz A, Esteban R, Buti M, Rodríguez-Frías F.

Título: Quantitative characterization of hepatitis delta virus genome edition by next generation sequencing.

Revista: Virus Research

Fecha: Enero (2018)

Volumen: 243

Páginas: 52-59

Autores: Sopena S, Godoy C, Tabernero D, Homs M, Gregori J, Riveiro-Barciela M, Ruiz A, Esteban R, Buti M, Rodríguez-Frías F.

Título: Quantitative characterization of hepatitis delta virus genome edition by next generation sequencing.

Revista: Virus Research

Fecha: Enero (2018)

Volumen: 243

Páginas: 52-59

Autores: Sopena S, Godoy C, Tabernero D, Homs M, Gregori J, Riveiro-Barciela M, Ruiz A, Esteban R, Buti M, Rodríguez-Frías F.

Título: Quantitative characterization of hepatitis delta virus genome edition by next generation sequencing.

Revista: Virus Research

Fecha: Enero (2018)

Volumen: 243

Páginas: 52-59

Autores: Sopena S, Godoy C, Tabernero D, Homs M, Gregori J, Riveiro-Barciela M, Ruiz A, Esteban R, Buti M, Rodríguez-Frías F.

Título: Quantitative characterization of hepatitis delta virus genome edition by next generation sequencing.

Revista: Virus Research

Fecha: Enero (2018)

Volumen: 243

Páginas: 52-59
