

TRABAJO FIN DE GRADO

Activación de la proteína de envoltura tóxica del retrovirus endógeno (HERVW-ENV) inducida por SARS-CoV-2

SARS-CoV-2-induced activation of pathogenic endogenous retrovirus envelope HERV-W

Autor: Carlota Domingo Lamelas

Director: Johan Albert Schoorlemmer

Grado en Biotecnología

2023/2024

ÍNDICE

1. RESUMEN	4
2. INTRODUCCIÓN	5
2.1. Retrovirus. Diferencias exógeno/endógeno. ¿Qué son los retrovirus endógenos humanos?	6
2.2. Estructura de HERV.....	6
2.3. Inserción de HERV en el genoma y expresión	6
2.4. Copias de HERV con implicaciones patológicas y dificultades en su detección	7
2.5. Métodos de detección de HERV	8
2.5.1. Ensayo de real time qPCR y complicaciones en análisis de HERV	8
2.5.2. Análisis de HERV mediante enriquecimiento con sonda DNA específica de secuencia y secuenciación por Illumina	9
Justificación:.....	11
Objetivos:	11
3. MATERIALES Y MÉTODOS	12
3.1. Colección de muestras y extracción de RNA de muestras PAXGene	12
3.2. Evaluación de la cantidad/calidad del RNA mediante NanoDrop	13
3.3. Evaluación de la integridad de RNA mediante Tapestation	13
3.4. Síntesis de cDNA.....	14
3.5. Análisis de la expresión génica de elementos retrovirales mediante qPCR	15
4. RESULTADOS.....	16
4.1. Extracción de RNA de muestras de PAXGene	16
4.2. Análisis de la expresión del gen <i>HERV-W ENV</i> mediante qPCR en muestras extraídas por PAXGene	17
4.2.1. Análisis de la contaminación por DNA genómico, validación del método de extracción mediante qPCR.	18
4.2.2 Comparación de la expresión de <i>HERV-W ENV</i> entre pacientes COVID persistente y pacientes sanos que han superado la enfermedad (controles)	20
4.3. Evaluación de la especificidad del método de hibridación para las secuencias <i>HERV-W ENV</i>	20
5. DISCUSIÓN	23
5.1. Análisis de la expresión del gen <i>HERV-W ENV</i> mediante qPCR en muestras extraídas en tubos PAXGene	23
5.1.1. Análisis de la contaminación por DNA genómico y validación del método de extracción mediante qPCR.	23
5.1.2. Comparación de la expresión de <i>HERV-W ENV</i> entre pacientes COVID persistente y pacientes sanos que han superado la enfermedad	23
5.2. Evaluación de especificidad del método de hibridación para secuencias <i>HERV-W</i>	24

6. CONCLUSIONES.....	26
7. REFERENCIAS BIBLIOGRÁFICAS	26

1. RESUMEN

El COVID-19 es una enfermedad infecciosa causada por el síndrome respiratorio agudo severo coronavirus 2 (SARS-CoV-2). Se ha demostrado que la sobreexpresión de la proteína viral HERV-W-ENV, producida por retrovirus endógenos humanos se induce durante la infección por SARS-CoV-2 y se encuentra en niveles altos en la sangre de pacientes con COVID-19 agudo grave. Los retrovirus endógenos son secuencias genómicas repetitivas con origen viral, provenientes de antiguas infecciones que afectaron a la línea germinal (ocupan en torno a 8% del genoma). Hay dos familias de estos retrovirus endógenos, HERV-W y HERV-K, que están especialmente implicadas en la patología de enfermedades inflamatorias. El gran número de copias existente y la homología entre sus secuencias dificulta la diferenciación entre las copias patogénicas y las que no lo son, además de la detección y ubicación de las posiciones que ocupan en el genoma.

El presente trabajo describe intentos de determinar y analizar el RNA de *HERV-W ENV*. En primer lugar, describe los niveles de el RNA de *HERV-W ENV* sangre periférica de pacientes de COVID persistente. El método ideal para análisis de elementos retrovirales es “long-read sequencing” (PacBio y Nanopore). Para ello el grupo de investigación usa una sonda específica para *HERV-W ENV* con el fin de seleccionar las secuencias que se pretenden analizar por secuenciación. Para llegar a este punto es necesario optimizar los pasos anteriores del proceso. Este trabajo evalúa si la sonda realmente es específica para secuencias *HERV-W ENV* o si su uso también resulta en la selección de secuencias de la familia *HERV-K*.

ABSTRACT

COVID-19 is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Overexpression of the viral protein HERV-W-ENV, produced by human endogenous retroviruses, has been shown to be induced during SARS-CoV-2 infection. Moreover, it is found at high levels in blood of patients with severe acute COVID-19. Endogenous retroviruses are repetitive genomic sequences with viral origin from ancient germline infections (occupying about 8% of the genome). Two families of these endogenous retroviruses, HERV-W and HERV-K, are particularly implicated in the pathogenesis of inflammatory diseases. The large number of copies from these sequences and the homology between them, makes it difficult to differentiate between pathogenic and non-pathogenic copies, as well as to detect and locate their positions in the genome.

This essay describes attempts to determine and analyze *HERV-W ENV* RNA. First, it describes the levels of *HERV-W ENV* RNA in the peripheral blood of long COVID patients. The ideal method for retroviral element analysis is long-read sequencing (PacBio and Nanopore). For this, the research group uses a *HERV-W ENV* specific probe to select the sequences that are supposed to be analyzed by sequencing. To reach this point it is necessary to optimize the previous steps of the process. This work evaluates whether the probe is specific for *HERV-W ENV* sequences or whether its use also results in the selection of *HERV-K* family sequences.

2. INTRODUCCIÓN

El COVID-19 es una enfermedad infecciosa causada por el síndrome respiratorio agudo severo coronavirus 2 (SARS-CoV-2) (1). La propagación de este virus ocurre principalmente a través de aerosoles respiratorios, es decir, pequeñas partículas que se expulsan al hablar, toser o estornudar y que pueden ser inhaladas por otras personas (2). Debido a la facilidad con la que se transmite de persona a persona, en cuestión de pocos meses desde su identificación inicial, el virus se extendió globalmente, dando lugar a una pandemia a nivel mundial.

Los síntomas más comunes en la infección aguda son fiebre, tos, dificultad para respirar, fatiga, dolores musculares y dolor de cabeza. Entre los pacientes que han padecido COVID-19 se observan con frecuencia numerosos síntomas o síndromes post-infecciosos de larga duración, que pueden afectar a personas de todas las edades. A esto se le llama COVID persistente o “long COVID” (1).

El grupo de investigación, que trabaja dentro del proyecto europeo HERVCOV, se centra en el estudio de la relación entre la proteína viral HERV-W-ENV, producida por retrovirus endógenos humanos (HERV), y la infección por SARS-CoV-2 (1). Estudios previos realizados en los laboratorios de colaboradores han demostrado que esta proteína se induce durante la infección por SARS-CoV-2 y se encuentra en niveles altos en la sangre de pacientes con COVID-19 agudo grave (1,3).

Específicamente, se ha demostrado que la expresión del mRNA de *HERV-W ENV* patogénico y el porcentaje de células positivas a la proteína HERV-W ENV se expresan significativamente en las células sanguíneas de pacientes con COVID-19 frente a las de donantes sanos. El estudio fue realizado con muestras de sangre de 30 pacientes hospitalizados por COVID-19 y 70 donantes sanos. La representación de los datos de expresión de ENV de HERV-W se muestran en la figura 1 (4).

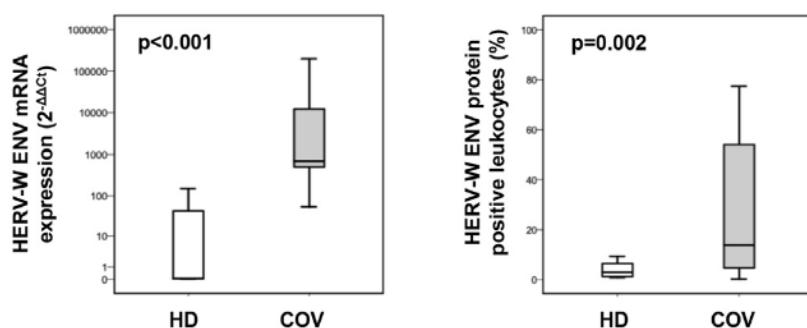


Figura 1: Representación de los datos de expresión obtenidos del análisis de leucocitos de pacientes COVID-19 y donantes sanos mediante real time q-PCR (mRNA) y citometría de flujo (proteína). Datos de Balestrieri et al., 2021 (4). Se muestra la expresión de HERV-W ENV en donantes sanos (HD) y pacientes COVID-19 (COV) en forma de mRNA (izquierda) y de proteína (derecha). El RNA procede de células sanguíneas (PBMCs, y granulocitos) de sangre periférica de los pacientes.

2.1. Retrovirus. Diferencias exógeno/endógeno. ¿Qué son los retrovirus endógenos humanos?

Un retrovirus es un virus cuyo material genético es RNA (5). Dentro de los retrovirus podemos realizar una distinción entre exógenos y endógenos. Los primeros son aquellos que realizan su ciclo vital mediante la infección de células somáticas y se transmiten de un organismo a otro por transmisión horizontal. Ejemplo de este tipo de virus son el virus de la inmunodeficiencia humana (VIH) y el virus de la leucemia de células T en humanos (HTLV) (6). Los retrovirus endógenos humanos son secuencias genómicas repetitivas con origen viral, provenientes de antiguas infecciones que afectaron a la línea germinal. Este proceso se estima que ha ocurrido a lo largo de los últimos millones de años (7–9). Los HERV se transmiten de forma vertical de generación en generación (10).

2.2. Estructura de HERV

Los HERV comparten la estructura genómica proviral con los retrovirus exógenos (figura 2). Normalmente, compuestos por dos secuencias terminales largas repetidas (LTRs por sus signos en inglés) que flanquean la región interna de los genes *gag*, *pro*, *pol* y *env* (figura 2), y que participan en el proceso de integración del material genético del virus en el genoma del huésped. Los genes virales codifican para los componentes estructurales de las partículas virales (figura 3). Estos componentes son matriz, cápside y nucleocápside (*gag*); la superficie de envoltura y subunidades transmembrana (*env*); así como para las enzimas implicadas en el ciclo de vida viral, la proteasa (*pro*), la transcriptasa inversa y la integrasa (*pol*) (7,9).

Los provirus de HERV albergan un sitio de unión a cebador (PBS) y un rasgo polipurínico (PPT) localizados entre el 5'LTR y *gag* y entre *env* y el 3'LTR, respectivamente (figura 2). Ambos tienen una importante función en la transcripción reversa (7).

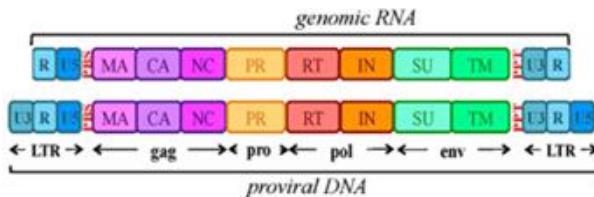


Figura 2: representación de la estructura del genoma de un retrovirus endógeno (7).

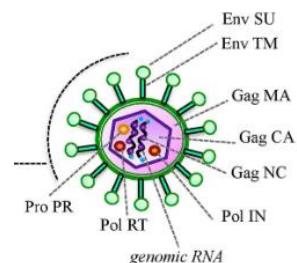


Figura 3: representación de la estructura de una partícula viral de un retrovirus (7).

2.3. Inserción de HERV en el genoma y expresión

Los HERV se agrupan en 31 familias basadas en la homología de sus secuencias, entre ellas HERV-W y HERV-K (11). Los HERV comprenden en torno al 8% del genoma humano. (1)

Cuando un retrovirus exógeno se introduce en el citoplasma de las células, su genoma de RNA es retrotranscritado en un DNA de doble cadena (dsDNA) por una retrotranscriptasa del propio virus, posteriormente el DNA proviral obtenido se integra en los cromosomas del huésped mediante una integrasa procedente del virus. Cuando el provirus se expresa, genera mRNA que codifica para las diferentes proteínas retrovirales. Estos mRNAs pueden volver a formar cDNA en presencia de una transcriptasa inversa y ser partícipes de nuevos eventos de integración (figura 4) (7).

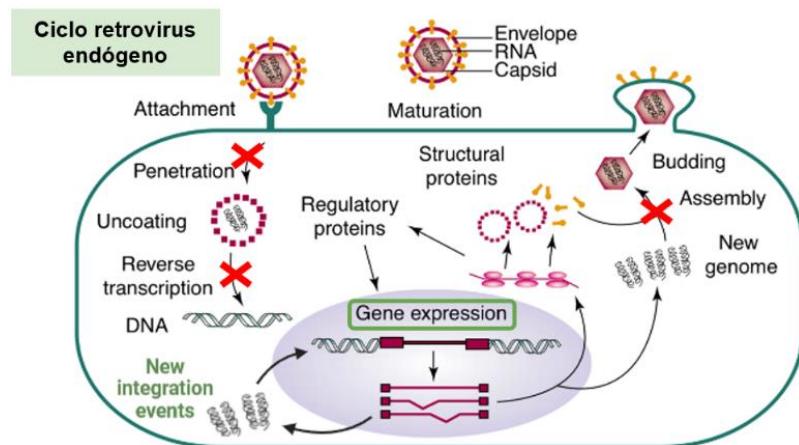


Figura 4: representación gráfica del ciclo vital de los retrovirus endógenos extraída del artículo Bruland 2003 (12) y posteriormente modificada. Las cruces rojas representan las vías que no siguen los retrovirus endógenos, pero que sí que llevarían a cabo los retrovirus exógenos.

Los retrovirus endógenos son una huella que queda en el genoma que deriva de la infección de un retrovirus exógeno. Desde su inserción en la línea germinal han ido evolucionando y mutando, lo que ha derivado en la pérdida de su función infectiva, es decir, no son capaces de formar partículas virales funcionales (7,13). Aunque la gran mayoría no se transcriben activamente en circunstancias biológicas (11), es bien sabido que las copias expresadas de determinadas familias de HERV intervienen en procesos fisiológicos y patológicos.

Por un lado, las proteínas de envoltura (ENV) de las familias HERV-FRD y HERV-W han sido relacionadas con el desarrollo placentario y la tolerancia inmunitaria materno-fetal de manera fisiológica (7,14,15). Los ejemplos mejor conocidos son los genes humanos *Sincitina 1/ERVW-1* y *Sincitina 2/ERVFRD-1*, que son productos de nuevos genes de envoltura de retrovirus endógenos humanos defectivos de las familias HER-W y HERV-FRD, respectivamente.

Por otro lado, la presencia de un agente infeccioso puede desencadenar la expresión de algunas copias de HERV y derivar en efectos patológicos (3,4). La expresión de proteínas de membrana de miembros de la familia HERV-K y la sobreexpresión de elementos HERV-W están implicadas en enfermedades inflamatorias (1,16–18), además de autoinmunes como la artritis reumatoide y la esclerosis múltiple (19).

En los últimos años, este campo se ha convertido en un tema importante debido a la aparición de la enfermedad COVID-19, que se ha relacionado con el aumento de los niveles de expresión de elementos retrovirales, como por ejemplo la proteína HERV-W ENV.

2.4. Copias de HERV con implicaciones patológicas y dificultades en su detección

Como se ha descrito en el apartado anterior, no todas las copias de HERV resultan estar implicadas en procesos patológicos, por ello, es de gran importancia ser capaces de identificarlas.

El genoma humano contiene un gran número de HERVs, cuyas secuencias tienen una alta homología. Tras su inserción en la línea germinal, los HERV son capaces de dar lugar a un número de copias muy elevado que se localizan en distintas partes del genoma. La tasa de mutación de la retro transposición, además de la transmisión vertical de generación en

generación hace que se vayan acumulando mutaciones durante largos períodos de tiempo (9). Las ligeras variaciones existentes entre las distintas copias de estos elementos retrovirales del genoma (figura 5) hacen muy difícil su identificación y saber dónde se ubican las copias patogénicas dentro del genoma (9).

Multiple Sclerosis-associated Retrovirus and HERV-W II

Sequence of MSRV clones	ATCTACCGTGACCCCTGGACCGGCTGCTAGACTATGCTGTATGTTAATGACATTGAA ATCTACCGCAGACCCCTGGACCGGCTGCTAGCCCCAGATCTGTATGTTAATGACATCAA ATCCACCGTGGACCCCTGGACCGGCTGCTAGCCCCATGCTCGATGTTAATGACATTGAA ATCTACCGTGACCCCTGGACCGGCTGCTAGCCCCATGCTGTATGTTAATGACATTGAA ATCTACCAACAGACCCCTGGACCGGCTGCTAGCCCCATGCTCAATGTTAATGACATTGAA ATCTACCATGGACCCCTGGACCGGCTGCTAGCCCCATGCTGTATGTTAATGACATTGAA ATCTGCCCGGACCCCTGGACCGGCTGCTAGCCCCATGGTCGATGTTAATGACATCAA ATCTACACGGACCCCTGGACCGGCTGCTAGCCCCATGCTCAATGTTAATGACATCGAA ATCTACCGTGACCCCTGGACCGGCTGCTAGCCCCATGCTCAATGTTAATGACATCAA ATCTACACGGACCCCTGGACCGGCTGCTAGCCCCATGCTCAATGTTAATGACATCAA ATCTACCATGGACCCCTGGACCGGCTGCTAGCCCCATGCTGTATGTTAATGACATTGAA
Human reference genome (version GRCh38.p12)	ATCTACCATGGACCCCTGGACCGGCTGCTAGCCCCATGCTGTATGTTAATGACATTGAA ATCTGCCCGGACCCCTGGACCGGCTGCTAGCCCCATGGTCGATGTTAATGACATCAA ATCTACACGGACCCCTGGACCGGCTGCTAGCCCCATGCTCAATGTTAATGACATCGAA ATCTACCGTGACCCCTGGACCGGCTGCTAGCCCCATGCTCAATGTTAATGACATCAA ATCTACACGGACCCCTGGACCGGCTGCTAGCCCCATGCTCAATGTTAATGACATCAA ATCTACCATGGACCCCTGGACCGGCTGCTAGCCCCATGCTGTATGTTAATGACATTGAA ATCTGCCCGGACCCCTGGACCGGCTGCTAGCCCCATGGTCGATGTTAATGACATCAA ATCTACACGGACCCCTGGACCGGCTGCTAGCCCCATGCTCAATGTTAATGACATCGAA ATCTACCGTGACCCCTGGACCGGCTGCTAGCCCCATGCTCAATGTTAATGACATCAA ATCTACACGGACCCCTGGACCGGCTGCTAGCCCCATGCTCAATGTTAATGACATCAA ATCTACCATGGACCCCTGGACCGGCTGCTAGCCCCATGCTGTATGTTAATGACATTGAA

→ Sequences of
HERV-W elements
(n~600)

Figura 5: comparación de las secuencias de MSRV (Multiple Sclerosis-associated Retrovirus) obtenidas a partir de mRNA expresado en casos de esclerosis múltiple, y comparación con secuencias de HERV-W extraídas del genoma de referencia (versión GRCh38.p12). Figura modificada de la tesis doctoral de Sofía Macías “Human endogenous retrovirus type W in multiple sclerosis” (20) y posteriormente modificada. Los asteriscos indican las posiciones en las que todas las secuencias son idénticas.

2.5. Métodos de detección de HERV

Existen diversos métodos disponibles para detectar las secuencias de elementos retrovirales.

2.5.1. Ensayo de real time qPCR y complicaciones en análisis de HERV

La real-time qPCR es un método estándar para la amplificación, detección y cuantificación de una secuencia diana específica a tiempo real (21). Se utilizan primers, que son secuencias cortas de DNA, que hibridan con regiones específicas del DNA diana y que son utilizados como molde para la amplificación que lleva a cabo la enzima DNA polimerasa.

La detección se consigue utilizando fluoróforos, que son moléculas que emiten fluorescencia al ser excitadas por luz. Estos fluoróforos pueden estar:

- Intercalados en el DNA: como el SYBR Green, que se une al DNA de doble cadena de manera no específica y emite fluorescencia (22).
- Unidos a sondas específicas: como las sondas TaqMan, que son secuencias de oligonucleótidos marcadas con un reportero fluorescente en un extremo y un quencher en el otro. Durante la amplificación, la sonda es degradada por la actividad exonucleasa de la DNA polimerasa, lo que separa el reportero del quencher y permite la emisión de fluorescencia (22).

Este es uno de los métodos más extendidos, es el de menor coste y más sencillo de realizar, por esta razón si es posible obtener resultados válidos se comenzará el análisis con este.

Complicaciones en el análisis de elementos retrovirales por real time qPCR

Habitualmente, los primers para detección de secuencias genómicas se realizan diseñando el primer directo en un exón y el reverso en el siguiente exón. Además, es conveniente que entre ambos exones se encuentre un intrón de gran tamaño. Durante el procesamiento del RNA, el

intrón se pierde, por tanto, las zonas de unión específica de los primers quedan próximas entre sí permitiendo que el amplicón resultante tenga el tamaño adecuado para ser detectado.

Mediante este diseño de primers podemos conocer el origen del material amplificado, ya que solo se producirá la amplificación del cDNA. Si hubiese gDNA procedente de una contaminación, tendrá el intrón, lo que imposibilitará su amplificación al ser el fragmento demasiado grande.

En la mayoría de los genes humanos se puede distinguir, con un diseño adecuado de primers, si la amplificación proviene de cDNA o gDNA. Lo que ocurre con los elementos retrovirales es que no tienen intrones, por lo que no podemos distinguir de igual manera si la señal de amplificación obtenida proviene de cDNA o de gDNA (contaminación).

Por lo tanto, el problema que surge al trabajar con elementos retrovirales es que en ensayos de qPCR no se puede distinguir cuál es el origen de la señal de amplificación, si DNA genómico (contaminación) o cDNA (procedente de la retro transcripción).

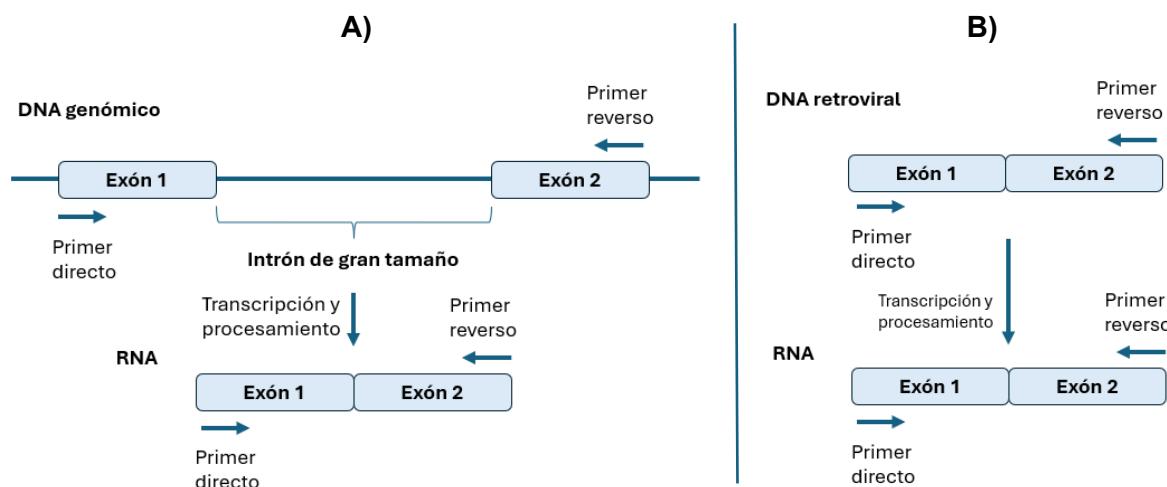


Figura 6: Comparación representativa de un método ordinario de diseño de primers para un ensayo qPCR (A) y método para un elemento retroviral endógeno (B).

Es fundamental por esta razón disponer de un buen método de extracción y aislamiento de RNA para que los resultados obtenidos sean válidos, porque en caso de haber contaminación de DNA genómico los resultados no se pueden interpretar.

2.5.2. Análisis de HERV mediante enriquecimiento con sonda DNA específica de secuencia y secuenciación por Illumina

El método ideal para analizar las secuencias de los HERV es la secuenciación, ya que además de que nos permite detectar diferencias en nucleótidos, también permite secuenciar de forma rápida, masiva y paralela millones de fragmentos de DNA. (23)

Dentro de los distintos métodos de secuenciación que existen en la actualidad, podemos diferenciarlos en 2 grupos en función del tamaño las lecturas que proporcionan. Por un lado, si las lecturas resultantes son cortas estamos hablando de “short-read sequencing” (hasta 300 pb), que es el caso de Illumina; por otro lado, si las lecturas obtenidas son largas (kb) estamos hablando de “long-read sequencing”, que sería el caso de PacBio o Nanopore.

Illumina se basa en la amplificación de los fragmentos de DNA generando clústers, y su posterior secuenciación por síntesis. Para ello se utilizan nucleótidos marcados con diferentes

fluorocromos que además presentan una modificación química que impide que se una más de uno a cada sitio de reacción (terminadores reversibles). Tras la incorporación de cada nucleótido se detecta la fluorescencia emitida correspondiente a su fluorocromo, y hay que eliminar la modificación química para permitir la unión del siguiente nucleótido (23). A pesar de que presenta múltiples ventajas, el análisis de las secuencias de los HERV y el mapeo en el genoma presenta complicaciones que derivan de disponer de short-reads.

Lo ideal para analizar los HERV es utilizar un método de “long-read sequencing”, porque cuanto más largas sean las lecturas, más complicado será encontrar regiones de secuencia comunes a dos o más lecturas sin que correspondan a una misma región genómica, y por tanto, más fácil será identificar y localizar las posiciones que ocupan estas secuencias en el genoma (24). Para ello el grupo de investigación usa una sonda específica para *HERV-W ENV* con el fin de seleccionar por hibridación (figura 7) las secuencias que se pretenden analizar por secuenciación. Para llegar a este punto es necesario optimizar los pasos anteriores del proceso. Este trabajo evalúa si el método de hibridación es eficaz, y si la sonda realmente es específica para secuencias *HERV-W ENV* o si su uso también resulta en la selección de secuencias de la familia *HERV-K*.

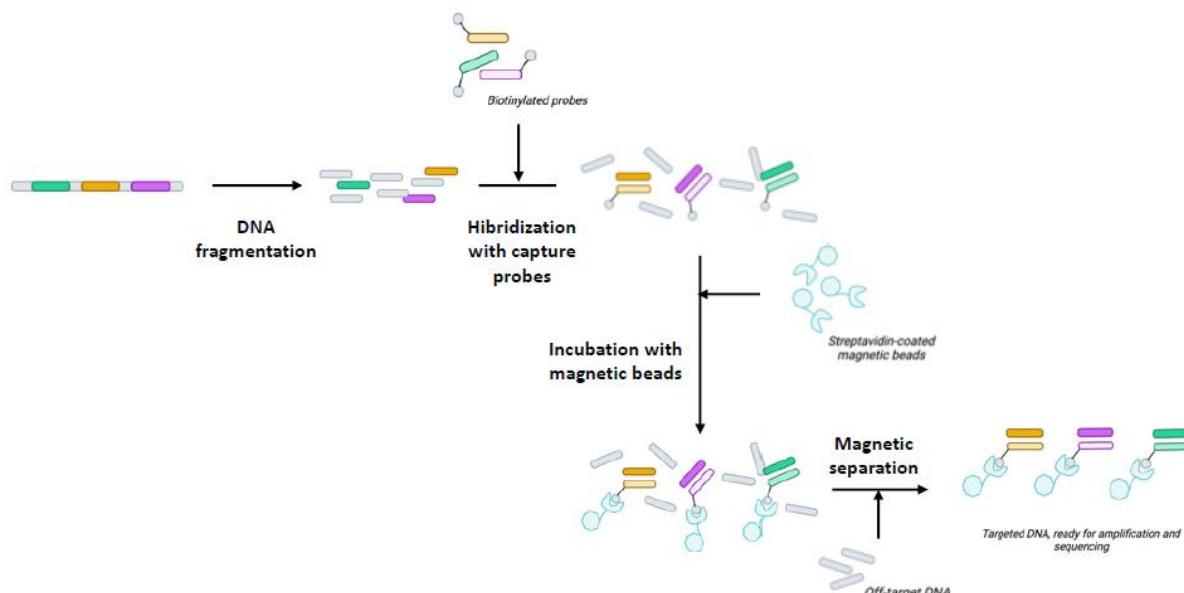


Figura 7: representación gráfica del proceso de enriquecimiento de secuencias diana (HERV-W ENV) mediante sondas específicas. El proceso comienza con la fragmentación del DNA procedente de la retro transcripción, después se hibridan los fragmentos con sondas específicas de secuencia (unidas a biotina), se incuba con bolas magnéticas (unidas a streptavidina). Se separan magnéticamente las sondas unidas a las secuencias de interés.

Justificación:

La Organización Mundial de la Salud (OMS) registró en diciembre de 2022 la definición oficial de “condición post COVID-19” (CPC), que se define como la persistencia o el desarrollo de nuevos síntomas tres meses después de la infección inicial por SARS-CoV-2, los cuales pueden durar al menos dos meses sin otra explicación (25). Los síntomas más característicos (según el Center for Disease and Prevention, USA) son alteraciones del gusto y olfato, dificultad para pensar o concentrarse, dolor de cabeza, problemas para dormir y depresión o ansiedad entre otros. Nos referiremos a este último como COVID persistente durante el desarrollo de este trabajo.

Como ya se ha descrito anteriormente, la proteína HERV-W ENV se induce durante la infección por SARS-CoV-2 y se encuentra en niveles altos en la sangre de pacientes con COVID-19 agudo grave. Sin embargo, no solo se ha asociado con COVID agudo, sino que también con COVID persistente, aunque en este caso hay menos estudios (26). Los síntomas derivan de la activación crónica del sistema inmune innato que llevan a cabo los HERV sobreexpresados (26). Se estima que en 2023 al menos 65 millones de personas en todo el mundo padecían COVID persistente, pero el número podría ser mayor, constituyendo un grave problema de salud a nivel mundial. (27)

Este trabajo consiste en analizar la expresión de *HERV-W ENV* en muestras de sangre de pacientes de COVID persistente y pacientes sanos que han superado la enfermedad (controles) con la finalidad de analizar si existen o no diferencias entre ellos a nivel de expresión de RNA. Además, considerando la problemática detección de ENV en general, y, por ende, de HERV-W y HERV-K en pacientes con COVID-19 persistente, es imprescindible tanto validar los métodos de detección por qPCR, como demostrar la presencia de niveles de HERV en pacientes.

Objetivos:

- Comprobar la efectividad del pretratamiento con DNase mediante qPCR y analizar si las muestras colectadas en tubos PAXGene están libres de contaminación por DNA.
- Analizar las diferencias de expresión de *HERV-W ENV* en muestras de sangre periférica de pacientes con COVID persistente y pacientes sanos que han superado la enfermedad (colectadas en tubos PAXGene).
- Comparar las diferentes formas de análisis de elementos retrovirales: por un lado, mediante real time qPCR y por otro mediante secuenciación por Illumina.
- Analizar si la sonda diseñada por el grupo de investigación es específica para secuencias *HERV-W*, o si también selecciona secuencias *HERV-K*.

3. MATERIALES Y MÉTODOS

3.1. Colección de muestras y extracción de RNA de muestras PAXGene

Colección de muestras. Se utilizan 48 muestras de sangre colectadas en tubos PAXGene procedentes del Biobanco del Sistema de Salud de Aragón. Las muestras provienen de personas que han firmado el consentimiento informado y dan autorización para que puedan ser utilizadas en investigación biomédica.

Los tubos PAXGene (fabricados y distribuidos por QIAGEN o BD para PreAnalytiX, ref: 762174) consisten en un sistema integrado para la extracción de sangre y la estabilización inmediata del RNA tanto de las células de la sangre, como del que está libre en el plasma. Reduce la degradación de RNA por RNasas (28). Contiene un aditivo que estabiliza el perfil de transcripto génico *in vivo* (29). Permite llevar a cabo ensayos de análisis del RNA manteniendo su integridad para análisis posteriores (28).

A pesar de que no son el mejor método de conservación de RNA, estos tubos permiten obtener muestras de forma fácil en atención primaria, es por ello por lo que se utilizan en este caso. Otros métodos más efectivos para mantener la integridad del RNA como, por ejemplo, utilizando el TMReactivio TRIzolTM, son más laboriosos, lo que dificulta el proceso de colección de muestras.

Las muestras en tubos PAXGene pueden almacenarse durante 3 días a 18-25°C o 5 días a 2-8°C. Si se requiere conservación durante más tiempo se mantienen hasta 4 años a temperaturas de entre -20°C y -70°C.

Extracción de RNA. Para la extracción de RNA a partir de muestras de sangre de PBMCs se ha utilizado el HiPure PAXGene Blood RNA Kit (Magen, ref: R416802). Las muestras de partida conservadas en tubos PAXgene Blood RNA deben estar la noche anterior a 4°C y ser incubadas 2h según el protocolo del fabricante. Primero se centrifugan los tubos PAXgene Blood RNA a 12000g, se quita el sobrenadante y se resuspende el pellet en 4 ml de agua libre de RNasas. Se centrifuga 10 minutos a 3000g y se vuelve a retirar el sobrenadante. A continuación, se añaden 350 µl de Buffer MBR1 y se disuelve de nuevo el pellet. Se pipetean las muestras a tubos de centrifuga de 1,5 ml. En este, se añaden, 300 µl de Buffer MBR2 y 40 µl de proteinasa K. Se mezcla para integrar bien los reactivos con la muestra y se incuba durante 20 min a 55°C en un incubador-agitador a 1400rpm. Durante este tiempo, se colocan las HiPure DNA Mini Columns en tubos colectores de 2 ml. Tras la incubación se transfiere el lisado a las columnas de DNA y se centrifuga 1 min a 12000g. Se descarta la columna y se guarda el fluido. Si queda un coágulo en el tubo se tiene que eliminar con una micropipeta antes de añadir 350 µl de etanol absoluto al tubo y mezclar con el fluido. Se colocan las HiPure RNA Mini Columns sobre nuevos tubos colectores y se añade 700 µl de la mezcla anterior sobre las columnas, se centrifuga a 12000g durante 1 min y en este caso se elimina el filtrado (líquido) y se reusan los tubos colectores. Se coloca de nuevo las columnas sobre estos tubos colectores y se repite este paso con la mezcla que quedaba sin transferir a la columna (solo caben 700 µl). En este instante, la muestra se encuentra en la membrana de la columna de RNA. Los pasos siguientes eliminan restos que hayan podido quedar en la membrana que no sean RNA para finalmente eluir solo el RNA que contenía la muestra. A continuación, se añade 350 µl de Buffer RW1, a la columna y se centrifuga a 12000g, 1 min y se descarta el fluido. Se reutiliza el tubo colector y se añade sobre la columna 100 µl de un mix de DNase I (contiene 10 µl de DNase I y 90 µl de DNase Buffer), se deja incubando 15 min a temperatura ambiente. Se añaden a la columna 500 µl de Buffer RW1 y se incuba 1 min, después

centrifugar a 12000g, 1 min. Se eliminan los tubos colectores con el contenido y se colocan las columnas sobre nuevos tubos colectores de 2 ml. Se añaden 650 µl de Buffer RW2 a la columna y se centrifuga a 12000g, 1 min y se descarta el eluido (reusamos el tubo colector). Se repite este paso y se centrifuga la columna vacía a 12000g, durante 2 min a temperatura ambiente para secar la matriz de la columna. Se transfiere la columna a un tubo de centrífuga de 1,5 ml. Se añade 40 µl de agua libre de RNasas y se centrifuga 12000g, 1 min para eluir el RNA. Se vuelven añadir 40 µl de agua libre de RNasas repitiendo el paso anterior. Por último, se incuba el eluido durante 5 minutos a 65°C en el incubador-agitación (sin agitación). Se guardan a -80°C.

3.2. Evaluación de la cantidad/calidad del RNA mediante NanoDrop

El espectrofotómetro NanoDrop-1000 (ThermoFisher) permite rápidamente disponer de una indicación de la concentración y la pureza del DNA, RNA y proteínas presentes en la muestra midiendo la absorbancia.

Los ratios que se obtienen del Nanodrop son indicadores importantes de la pureza y la calidad de los ácidos nucleicos (DNA y RNA):

- Ratio 260/280: se utiliza para evaluar la pureza de los ácidos nucleicos. Un valor óptimo para el DNA es de alrededor de 1.8 y para el RNA, cerca de 2.0. Un valor más bajo puede indicar contaminantes como proteínas o solventes orgánicos, mientras que un valor más alto puede sugerir la presencia de fenol, sales u otros contaminantes.
- Ratio 260/230: se utiliza para evaluar la pureza de los ácidos nucleicos, pero es más sensible a la presencia de contaminantes orgánicos e inorgánicos, como fenol, sales, y solventes orgánicos. Un valor óptimo está entre 1.8 y 2.2 para el ADN, y entre 2.0 y 2.2 para el ARN (30).

Para obtener unos valores de medición más precisos de la cantidad de RNA obtenida debe realizarse la medición el mismo día de la extracción.

3.3. Evaluación de la integridad de RNA mediante TapeStation

El método Tapestation se basa en analizar la integridad del RNA de las subunidades ribosomales, ya que al constituir la mayor parte del RNA, son una clara representación de la integridad de este. El resultado obtenido se muestra mediante RIN (RNA integrity number) (31).

Se evalúa la calidad de los RNAs previamente extraídos mediante TapeStation, para ello se utiliza el bioanalizador Tape Station 4150 (Agilent, Santa Clara, CA). Las muestras se diluyen 10ng/µl. Además, como control de calidad positivo se utiliza un RNA de buena calidad del laboratorio.

Metodología: El kit utilizado es el High Sensitivity RNA ScreenTape Analysis (Agilent). Para cada muestra se pipetea 1 µl del High Sensitivity RNA Sample Buffer (Ref: 5067-5580) y 2 µl de la muestra de RNA en tubos strip. Se mezcla utilizando un vortex IKA a 2000 rpm durante 1 min. Se da un spin de centrifugación quedando la muestra al fondo del tubo. A continuación, se lleva a cabo la desnaturización de las muestras:

- Se calientan las muestras a 72°C durante 3 min.
- Se colocan las muestras en hielo durante 2 min.

Después, se introduce la Screen Tape (Ref: 5067-5579) y las muestras en el dispositivo y se selecciona el modo de ensayo de RNA. En el software del dispositivo se selecciona el Ladder electrónico (se utiliza para correr las muestras), no se prepara en el laboratorio.

3.4. Síntesis de cDNA

Se utiliza el kit Mezcla Maestra SuperScript™ IV VILO™ (ThermoFisher, Ref: 11766050). Este kit incluye un paso de eliminación de DNA (ezDNase). La ezDNase es una nueva enzima termolábil específica de doble cadena que se utiliza para eliminar la contaminación de gDNA del RNA molde antes de la reacción de retro transcripción. Además, la enzima VILO permite la síntesis de cDNA de forma más rápida, sensible y reproducible (32).

Contenido del kit:

- Se utiliza SuperScript IV VILO Master Mix con la enzima ezDNase que es una solución optimizada para la síntesis de la primera hebra de cDNA en dos pasos de aplicación de RT-PCR.
- La master mix 5x incluye: SuperScript IV Reverse inhibitor, proteínas helper, proteínas estabilizantes, oligos (dT18), random primers (hexamers), $MgCl_2$, y dNTPs.
- El SuperScript IV VILO No RT Control contiene todos los componentes de SuperScript IV VILO Master Mix a excepción de la enzima transcriptasa reversa.

Metodología: El procedimiento comienza con la preparación de un mix de reacción para la digestión de gDNA. La cantidad de RNA de partida son 500 ng. Los volúmenes de los componentes de reacción vienen detallados por muestra en la tabla 1 (para 20 µL). Se agita la mezcla y se deja incubar el mix de reacción con las muestras durante 5 min a 37°C. Al acabar este tiempo se coloca en hielo.

Se necesita disponer de un control negativo para asegurarse de que no hay contaminación por gDNA. Cada una de las muestras se incuba simultáneamente por separado con la reacción No RT control (no tiene la enzima SuperScript IV VILO, o retro transcriptasa), y con la reacción RT completa (con la enzima retro transcriptasa).

Tabla 1: cantidades de cada componente del mix de reacción para la digestión de gDNA (por muestra).

Component	Volume
10X ezDNase Buffer	1 µL
ezDNase enzyme	1 µL
Template RNA (1 pg to 2.5 µg total RNA)	varies
Nuclease-free Water	to 10 µL

A continuación, se preparan las reacciones RT y No RT Control (en hielo) siguiendo las cantidades que aparecen en la tabla 2.

Tabla 2: cantidades de cada componente de las reacciones RT y No RT Control (por muestra).

Add the following components to the tube containing the 10-µL reaction mix from step 2 on ice.

Component	RT reaction	No RT Control reaction
SuperScript™ IV VILO™ Master Mix	4 µL	—
SuperScript™ IV VILO™ No RT Control	—	4 µL
Nuclease-free Water	6 µL	6 µL

Se introducen las muestras al termociclador con el siguiente programa: 10 min a 25°C (hibridación de los cebadores); 15 min a 50°C (transcripción reversa del RNA); 5 min a 85°C (inactivación de la enzima); 4°C.

3.5. Análisis de la expresión génica de elementos retrovirales mediante qPCR

El kit utilizado es TB Green Premix Ex Taq II™ (TII RNaseH Plus) (TAKARA, Ref: RR820A). Los primers utilizados en los ensayos de qPCR se muestran en la tabla 3.

Tabla 3: secuencias de los primers utilizados en los ensayos de qPCR.

DIANA	Secuencias primers	ORIGEN
ENV de HERV-K	Primer directo: 5'-CTGAGGCCATTGCAGGAGTT-3' Primer reverso: 5'-GCTGTCTCTCGGAGCTGTT-3'	Sigma
ENV de HERV-W	Primer directo: 5'-GTATGTCTGATGGGGTGGAG-3' Primer reverso: 5'-CTAGTCCTTGTAGGGCTAGAG-3'	IDT
RPL-19	Hs.PT.58.20491224	IDT

Se prepara una master mix con la pareja de primers específica para el gen *HER-W ENV* (igual para *HER-K ENV*), que será detectado utilizando SYBR, y otra master mix con la pareja de primers dirigida al gen *RPL-19*, que será detectado mediante el uso de una sonda.

Tabla 4: cantidades (μ l) a añadir de cada componente de la reacción de master mix (para una reacción).

Componentes reacción	SYBR	SONDA
Agua MQ	3,8	3,6
Mezcla 2x TB Green Premix Ex Taq II (TII RNase H Plus) Ref: RR820	10	-
Mezcla 2x EmeraldAmp® MAX PCR Master Mix Ref: RR390	-	10
ROX 50x (tipo II)	0,4	0,4
Primer forward	0,4	-
Primer reverse	0,4	-
Sonda + Pr (Assay) 20x	-	1

Las rectas de las reacciones fueron optimizadas previamente en el laboratorio sin necesidad de añadir el cofactor Mg²⁺.

Se añaden a cada pocillo 15 μ l de la mastermix correspondiente y 5 μ l de muestra cDNA (previamente diluido).

El equipo utilizado para la realización de este ensayo es el QuantStudio 5 (ThermoFisher). El programa del termociclador utilizado es el siguiente:

1. Hold Stage: 95°C, 4 min. 1 ciclo.
2. PCR Stage: 95°C, 15 seg → 60°C, 1 min. 40 ciclos.

Para la obtención de las melting curves como paso posterior a la amplificación a tiempo real:

3. Melting Curve Stage: 95°C, 15 seg; 60°C, 1 min; 95°C 15 seg.

Las melting curves son gráficas que se obtienen a partir de la fluorescencia emitida por el fluoróforo unido al DNA conforme se va aumentando la temperatura y se van separando las hebras del DNA. Los picos de fluorescencia resultantes indican la Tm de la curva, es decir, la temperatura a la que la mitad de las cadenas de DNA se encuentran disociadas. Un ensayo específico debe mostrar un pico de fusión único, correspondiente a un producto único. Esto nos permite detectar amplificaciones inespecíficas, puesto que si hay varios picos a distintas Tm es un indicativo de que hay varios productos de amplificación diferentes (33).

Se realizan los ensayos de qPCR de cada una de las muestras habiendo sido tratadas y sin tratar con enzima retro transcriptasa para poder comparar y determinar si la señal proviene de cDNA o gDNA (en las muestras sin retro transcribir no habrá amplificación a no ser que haya contaminación de DNA genómico).

Cada muestra se analiza por duplicado, los Ct de cada una resultan de la media de los Ct del duplicado. Para normalizar los resultados se utiliza el gen de referencia *RPL-19*.

Para comparar la expresión entre grupos se utiliza el método $\Delta\Delta Ct$. ΔCt se calcula mediante la diferencia entre los valores Ct de las muestras y el valor Ct medio de los genes de referencia. Para obtener el $\Delta\Delta Ct$, se calcula la diferencia entre el ΔCt y el valor Ct medio de las muestras del grupo control. El fold change se obtiene mediante la fórmula $2^{-\Delta\Delta Ct}$. La presencia de secuencias de DNA contaminantes se calcula comparando la expresión entre muestras paralelas de RNA con o sin transcriptasa reversa y se obtiene ΔCt en ambas condiciones recalculado como foldchange (34).

El grupo de investigación estudia la expresión de elementos retrovirales en tejido placentario. Es por ello por lo que se dispone de RNAs de placenta que contienen alta expresión de elementos retrovirales. Estos RNAs se utilizan en el presente trabajo como control positivo de la expresión de *HERV-W ENV*

4. RESULTADOS

4.1. Extracción de RNA de muestras de PAXGene

Mediante el protocolo que se sigue en materiales y métodos se extrajo el RNA de 48 muestras de sangre periférica colectada y conservada tubos PAXGene en diferentes tandas.

4.1.1. Evaluación calidad/cantidad del RNA mediante NanoDrop

Se analizan todas las muestras de RNA al finalizar la extracción. El ratio 260/280 calculado para cada una las muestras presenta un valor cercano a 2.0, que es un indicativo de la buena pureza del RNA. El ratio 260/230 difiere de los resultados óptimos, que deberían encontrarse entre 2.0 y 2.2. Este último ratio es más sensible a contaminantes orgánicos e inorgánicos, como por ejemplo, el etanol utilizado durante el proceso de extracción. A pesar de que pueda

haber restos de etanol, como el ratio 260/280 se encuentra en el rango esperado, concluimos que los RNA extraídos presentan la calidad suficiente para continuar con los análisis.

En el ANEXO I se pueden ver las tablas de resultados con las concentraciones, los valores de absorbancia y los ratios obtenidos a partir de las mediciones del NanoDrop, además de las gráficas de absorbancia correspondientes.

4.1.2. Análisis de la calidad y la integridad del RNA de las muestras extraídas de tubos PAXGene

Se evalúa la calidad de los RNAs mediante TapeStation siguiendo el protocolo descrito anteriormente. Para ello se seleccionan 8 muestras representativas.

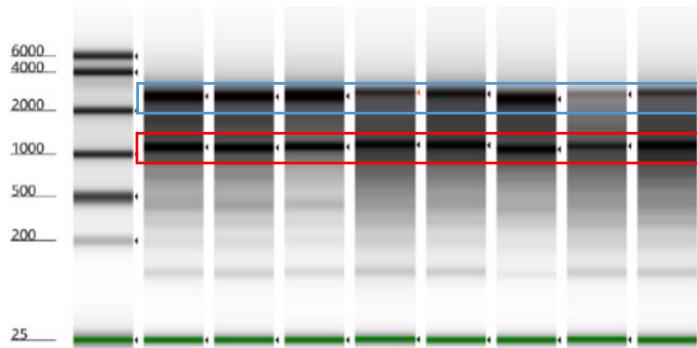


Figura 8: Análisis de la integridad del RNA mediante TapeStation. En la imagen se muestran las bandas electroforéticas resultado del ensayo, las bandas correspondientes a RNA ribosómico 18S y 28S aparecen recuadradas (rojo y azul respectivamente). La columna de la izquierda corresponde al patrón de pesos moleculares, el resto de las columnas son las 8 muestras de RNA analizadas.

RIN ^e							
7.2	7.2	7.9	5.3	6.0	6.5	4.6	5.5

Figura 9: Análisis de la integridad del RNA mediante TapeStation. Valores RIN obtenidos para cada una de las 8 muestras analizadas por TapeStation. El resultado se muestra mediante el valor RIN, que es calculado en una escala de 1-10. Un valor RIN alto, indica una buena integridad del RNA, y un valor RIN cercano a 1 indica que el RNA está muy degradado.

Todas las muestras presentan las bandas 28S y 18S (figura 8), además en todas se observa que la 28S es el doble del 18S. Los valores RIN se encuentran dentro del rango 4-10 (figura 9), lo que indica que las muestras presentan la calidad suficiente para poder ser utilizadas en los análisis posteriores.

4.2. Análisis de la expresión del gen HERV-W ENV mediante qPCR en muestras extraídas por PAXGene

Una vez evaluada la calidad de los RNA extraídos de muestras de sangre de PAXGene procedentes de pacientes de COVID persistente, se procede a analizar la expresión del gen *HERV-W ENV* mediante qPCR.

Para el ensayo se toman 47 muestras, de las cuales 24 son pacientes con COVID persistente y 23 son controles (pacientes que pasaron el COVID y se recuperaron).

4.2.1. Análisis de la contaminación por DNA genómico, validación del método de extracción mediante qPCR.

Se compara el resultado de la amplificación de *HERV-W ENV* de muestras con y sin RTasa, para determinar qué señal de *HERV-W ENV* proviene de cDNA (RNA), y si parte de la señal se debe a DNA genómico presente en la muestra (contaminación).

En la qPCR de las muestras de RNA que han sido previamente tratadas con enzima RTasa, para la amplificación con la pareja de primers de *HERVW-ENV*, los valores de Ct se encuentran comprendidos entre 27-30, mientras que para el gen de referencia *RPL-19* los valores de Ct detectados se encuentran entre 22-24. ΔCt indica la diferencia de ciclos entre la amplificación de *RPL-19* Y *HERV-W ENV*, de 5-7 ciclos.

En el caso de la qPCR de muestras que no fueron tratadas con RTasa, para la pareja de primers de *HERV-W ENV* los Ct se encuentran en valores de 32 a más de 40, siendo entre 34 y más de 40 para *RPL-19*. En este experimento los ΔCt presentan valores entre 0 y 5.

Tabla 5: resultados de qPCR de cDNAs procedentes de RNA de pacientes COVID persistente y controles.

A)

Retrotranscripción sin enzima						
Nº muestra	Tipo de muestra	Ct HERV-W ENV	Ct RPL-19	ΔCt	ΔΔCt	Fold Change
1	COVID	40,00	35,11	4,89	5,46	0,02
2	COVID	35,98	38,39	-2,41	-1,84	3,58
3	COVID	40,00	38,81	1,19	1,76	0,29
4	COVID	37,30	40,00	-2,70	-2,13	4,37
5	COVID	37,57	37,81	-0,24	0,33	0,79
6	COVID	35,06	40,00	-4,94	-4,37	20,62
7	COVID	36,07	37,18	-1,10	-0,53	1,45
8	COVID	34,62	40,00	-5,38	-4,81	28,04
9	COVID	35,98	36,86	-0,88	-0,31	1,24
10	COVID	35,57	36,58	-1,00	-0,43	1,35
11	COVID	40,00	36,98	3,02	3,59	0,08
12	COVID	40,00	34,57	5,43	6,00	0,02
13	COVID	36,29	36,94	-0,65	-0,08	1,06
14	COVID	40,00	39,70	0,30	0,87	0,55
15	COVID	40,00	35,42	4,58	5,15	0,03
16	COVID	40,00	37,30	2,70	3,27	0,10
17	COVID	35,02	40,00	-4,98	-4,41	21,23
18	COVID	40,00	35,12	4,88	5,45	0,02
19	COVID	40,00	40,00	0,00	0,57	0,67
20	COVID	40,00	39,94	0,06	0,63	0,65
21	COVID	40,00	35,32	4,68	5,25	0,03
22	COVID	40,00	40,00	0,00	0,57	0,67
23	COVID	40,00	36,36	3,64	4,21	0,05
24	COVID	40,00	36,72	3,28	3,85	0,07
25	Control	34,59	39,19	-4,59	-4,02	16,27
26	Control	36,86	40,00	-3,14	-2,57	5,93
27	Control	32,58	39,33	-6,75	-6,18	72,51
28	Control	34,17	40,00	-5,83	-5,26	38,28
29	Control	35,31	39,56	-4,25	-3,68	12,77
30	Control	40,00	40,00	0,00	0,57	0,67
31	Control	40,00	37,96	2,04	2,61	0,16
32	Control	35,46	37,03	-1,57	-1,00	2,00
33	Control	34,37	31,46	2,91	3,48	0,09
34	Control	34,05	31,34	2,71	3,28	0,10
35	Control	34,21	29,7	4,51	5,08	0,03
36	Control	35,77	37,51	-1,74	-1,17	2,25
37	Control	34,63	32,7	1,93	2,50	0,18
38	Control	34,89	31,73	3,16	3,73	0,08
39	Control	32,76	30,89	1,87	2,44	0,18
40	Control	35,61	30,07	5,54	6,11	0,01
41	Control	40,00	39,34	0,66	1,23	0,43
42	Control	34,72	39,34	-4,62	-4,05	16,62
43	Control	34,90	39,08	-4,18	-3,61	12,20
44	Control	40,00	38,15	1,85	2,42	0,19
45	Control	37,11	36,18	0,93	1,50	0,35
46	Control	40,00	40,00	0,00	0,57	0,67
47	Control	33,22	37,70	-4,48	-3,91	15,05

B)

Retrotranscripción con enzima						
Nº muestra	Tipo de muestra	Ct HERV-W ENV	Ct RPL-19	ΔCt	ΔΔCt	Fold Change
1	COVID	27,24	22,20	5,03	-0,73	1,66
2	COVID	28,27	22,57	5,70	-0,06	1,04
3	COVID	27,78	21,91	5,86	0,10	0,93
4	COVID	27,28	21,64	5,64	-0,12	1,09
5	COVID	27,67	22,52	5,15	-0,61	1,53
6	COVID	27,08	21,57	5,52	-0,24	1,19
7	COVID	28,31	21,68	6,64	0,88	0,54
8	COVID	28,21	23,53	4,68	-1,08	2,11
9	COVID	28,71	22,03	6,68	0,92	0,53
10	COVID	28,23	22,37	5,85	0,09	0,94
11	COVID	28,78	22,75	6,04	0,28	0,83
12	COVID	27,54	22,19	5,35	-0,41	1,33
13	COVID	27,65	22,33	5,32	-0,44	1,35
14	COVID	27,79	22,42	5,37	-0,39	1,31
15	COVID	29,35	22,81	6,54	0,78	0,58
16	COVID	26,89	21,91	4,98	-0,78	1,72
17	COVID	28,36	23,11	5,25	-0,51	1,42
18	COVID	27,86	22,12	5,74	-0,02	1,02
19	COVID	27,45	22,03	5,42	-0,34	1,26
20	COVID	28,29	22,76	5,53	-0,23	1,17
21	COVID	29,27	22,64	6,63	0,87	0,55
22	COVID	28,70	22,94	5,76	0,00	1,00
23	COVID	30,33	23,22	7,11	1,35	0,39
24	COVID	27,85	22,23	5,62	-0,14	1,10
25	Control	29,22	23,98	5,24	-0,52	1,44
26	Control	29,69	23,46	6,23	0,47	0,72
27	Control	28,45	22,74	5,71	-0,05	1,04
28	Control	29,38	23,62	5,76	0,00	1,00
29	Control	30,29	24,28	6,01	0,25	0,84
30	Control	29,20	23,19	6,01	0,25	0,84
31	Control	29,86	23,90	5,96	0,20	0,87
32	Control	29,90	23,76	6,14	0,38	0,77
33	Control	29,50	23,74	5,76	0,00	1,00
34	Control	30,53	24,87	5,66	-0,10	1,07
35	Control	29,21	23,30	5,92	0,16	0,90
36	Control	30,87	26,14	4,73	-1,03	2,04
37	Control	30,30	24,08	6,22	0,46	0,73
38	Control	29,44	23,11	6,33	0,57	0,67
39	Control	28,87	22,95	5,92	0,16	0,89
40	Control	30,09	22,90	7,19	1,43	0,37
41	Control	29,16	23,69	5,48	-0,28	1,22
42	Control	28,23	23,18	5,05	-0,71	1,64
43	Control	29,36	23,93	5,42	-0,34	1,26
44	Control	29,19	23,96	5,23	-0,53	1,44
45	Control	28,90	23,43	5,47	-0,29	1,22
46	Control	29,94	23,43	6,52	0,76	0,59
47	Control	30,03	24,49	5,54	-0,22	1,16

Se muestran las medias de los Ct correspondiente a cada muestra. Se observan los Ct de los cDNAs procedentes de RNAs sin tratar con enzima RTasa (A) y tratados con enzima RTasa (B). Resultados de la amplificación con parejas de primers dirigidos a *HERV-W ENV* y *RPL-19*. Cálculo de FoldChange para comparar expresión en pacientes COVID persistente (COVID) y pacientes control.

En los RNA sin tratar con enzima RTasa no se detecta amplificación de los genes diana, ni para *HERV-W ENV*, ni usando la pareja de primers específica del gen de referencia *RPL-19* (solo amplifica cDNA). Por lo tanto, la señal de amplificación de los RNAs tratados con RTasa proviene únicamente del cDNA sintetizado a partir de ellos. No se observa contaminación genómica.

Para este ensayo, se incluyeron como controles positivos dos muestras de RNAs de placenta. Analizando los Ct de los cDNAs obtenidos a partir de estos RNAs, se observa como estos aparecen en el ciclo 18, indicativo de que el material amplificado, cDNA de *HERV-W ENV*, presenta alta expresión en las muestras de placenta.

Tabla 6: resultados de qPCR de cDNAs procedentes de RNAs de placentas

Se muestran los Ct (media de los duplicados de cada muestra) de los cDNAs procedentes de RNAs de placenta tratados con enzima RTasa y sin tratar, amplificados con parejas de primers dirigidos a *HERV-W ENV* y *RPL-19*

Muestras	Ct HERV-W ENV	Ct RPL-19
Placenta1 RT	18,21940036	22,69930621
Placenta1 sin RT	37,36036658	40
Placenta2 RT	18,95105181	28,41604946
Placenta2 sin RT	40	40

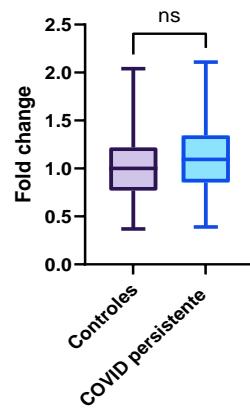
Estos RNAs de placenta nos sirven para establecer un verdadero control positivo. Al comparar las diferencias de expresión de *HERV-W ENV* entre las muestras de COVID persistente y las muestras de placenta, se observa claramente como la diferencia entre los Ct es muy grande (10 ciclos). Pues la media de Ct de las muestras de placenta es en torno a 18,5 y la de las muestras de COVID persistente es alrededor de 28.

En base a estos resultados podemos decir que en las muestras de COVID persistente los niveles de expresión de *HERV-W ENV* detectados son bajos.

4.2.2 Comparación de la expresión de *HERV-W ENV* entre pacientes COVID persistente y pacientes sanos que han superado la enfermedad (controles)

Una vez comprobado que no hay contaminación por gdNA y que la señal observada procede de cDNA de la retro transcripción, se pueden analizar las diferencias de expresión entre las muestras procedentes de pacientes de COVID persistente y pacientes sanos que han superado la enfermedad (controles).

Figura 10: representación gráfica de las diferencias de expresión de *HERV-W ENV* entre pacientes de COVID persistente y controles a partir de los datos obtenidos mediante real time qPCR. La normalidad se evaluó con el test de Shapiro-Wilk (cuando $n < 50$). Para la evaluación de la significación estadística de las diferencias, se realizó una prueba específica en función de la normalidad y del número de grupos de muestras comparados. Los datos están distribuidos normalmente: prueba t-student (2 grupos).



Según los resultados obtenidos (figura 10), no se observa una expresión diferencial de *HERV-W ENV* entre muestras de COVID persistente (casos) y muestras control.

4.3. Evaluación de la especificidad del método de hibridación para las secuencias *HERV-W ENV*

Se quiere evaluar si una sonda de hibridación para secuencias *HERV-W ENV*, que ha sido diseñada por el grupo de investigación, es específica para esta familia de retrovirus endógeno, y por tanto es capaz de discriminar las secuencias de esta familia respecto a las de *HERV-K*.

Se analizaron 4 muestras disponibles de RNA convertidos en cDNA y librería, antes y después del enriquecimiento con sonda (figura 7). La comparación de estas muestras se realiza mediante qPCR, analizando un amplicón situado dentro del gen *ENV* de *HERV-K*. En el caso ideal no debería de verse señal de amplificación de *ENV* de *HERV-K*. Pues la mayor parte del contenido de la muestra debería ser *ENV* de *HERV-W* si la sonda es realmente específica de estas secuencias.

Al analizar las melting curves del ensayo (figura 11), lo que se observa es que hay varios picos de fusión para cada una de las muestras. Esto es un indicativo de que la reacción no ha sido específica pues hay más de un producto de amplificación. Se concluye por los resultados obtenidos a partir de las melting curves que la PCR no es fiable y que, por tanto, no podemos evaluar si la sonda diseñada es específica para secuencias *HERV-W* mediante este método.

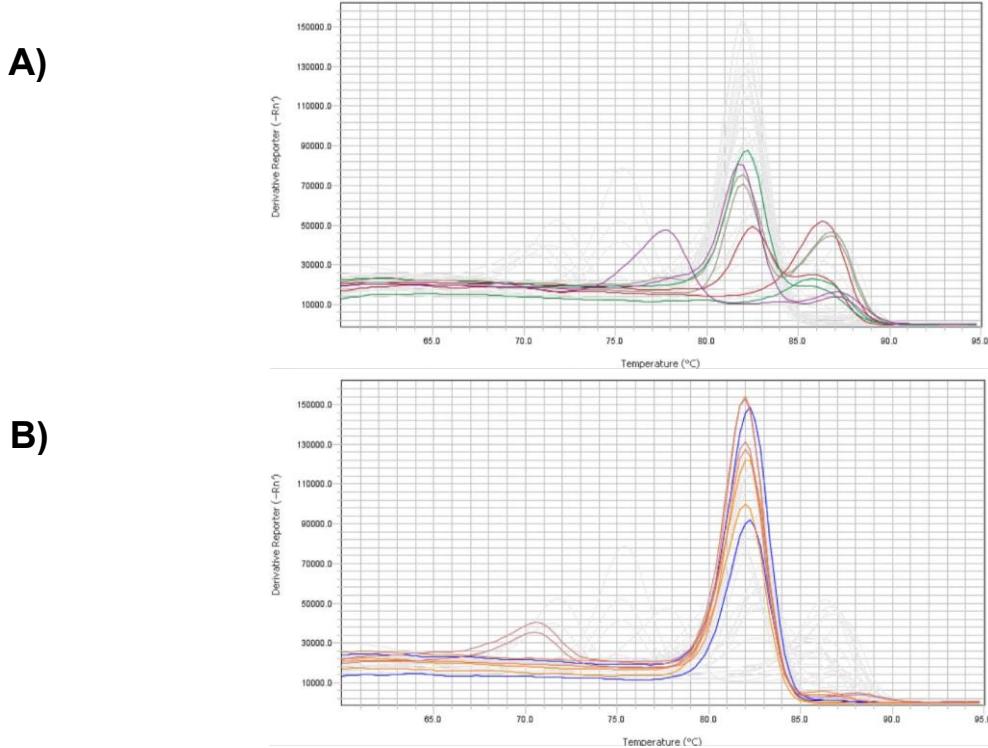


Figura 11: representación gráfica de las melting curves obtenidas de un ensayo de real time qPCR en el que se analizan las 4 muestras disponibles de RNA convertidos en cDNA y librería, después del enriquecimiento con sonda (A) y un ejemplo de amplificación específica de *HERV-W ENV* (B). En la imagen A puede observarse que aparecen dos picos de fusión para cada muestra (representadas en distintos colores). Por lo tanto, hay varios productos de amplificación, indicativo de un ensayo inespecífico. La imagen B muestra un ejemplo de cómo son las melting curves de un ensayo específico para *HERV-W ENV*.

De este ensayo podemos plantearnos la siguiente pregunta, si la real time qPCR no nos sirve para este análisis, ¿cómo podemos analizar la especificidad de la sonda? Una opción sería secuenciar el DNA seleccionado con sonda mediante Illumina. Colaboradores del grupo “ERV (Elementos Retrovirales Endógenos) en el Desarrollo y Patologías Asociadas”, realizaron dicho análisis. Lo que hicieron fue analizar muestras de DNA genómico convertido en librería (proceso de la figura 7). Las librerías estaban formadas por fragmentos de una longitud media de 5 kb, las secuencias de *HERV-W* seleccionadas previamente (figura 7). Las lecturas obtenidas (ver tabla 7) fueron mapeadas contra las secuencias *HERV-W* y *HERV-K*. Los resultados se muestran en la tabla 7.

Tabla 7: lecturas de Illumina mapeadas sobre HERV.

Muestra	Tamaño medio de los fragmentos antes generación librería	Lecturas totales	Lecturas mapeadas a <i>HERV-W</i>		Lecturas mapeadas a <i>HERV-K</i>	
			Nº	%	Nº	%
1	12 kb	1604024	10591	0,66	160	0,00
2	5 kb	2224134	161429	7	280	0,01
3	2 kb	3566868	161376	4,52	462	0,01

Se obtuvieron lecturas de 3 bibliotecas diferentes mediante secuenciación por Illumina y posteriormente se mapearon a las secuencias referencia de *HERV-W* y *HERV-K* utilizando el programa bioinformático Bowtie2. Se describen 3 bibliotecas diferentes, cada una con un tamaño medio de los fragmentos diferente (12 kb, 5kb, y 2 kb). Se indican las lecturas totales obtenidas (tras control de calidad) y el número de lecturas mapeadas a las secuencias de *HERV-K* (figura 11A) o *HERV-W* (figura 11B).

Como vemos en la tabla 7 el número de lecturas mapeadas a *HERV-W* es mayor que para *HERV-K* en las 3 librerías secuenciadas por Illumina. Tomando la librería 2 (muestra 2 en la tabla 7) como ejemplo, el número de lecturas mapeadas a *HERV-W* es de 161429 frente a las 280 mapeadas a *HERV-K*. Es decir, el 7% de las lecturas totales han sido mapeadas a *HERV-W*, en cambio tan solo un 0,01% de las lecturas han sido mapeadas a *HERV-K*. Estos resultados indican que la mayor parte del contenido de la librería de cDNA después de haber pasado por el proceso de hibridación (figura 7) son secuencias de *HERV-W*.

A)

Alineamiento con HERVK

Query: M03926:170:000000000-GDWDR:1:1101:14158:17039
Target: AY037928.1 Human endogenous retrovirus K113 complete genome
Model: affine:local:dna2dna
Raw score: 655
Query range: 0 -> 150
Target range: 8624 -> 8470

1 : GATTAACAGCATCCCAAGGCAGAAGAATTTCCTTAGTATAGAACAAAATGGAGTCTCCTAT : 62
||||||| ||| ||||||||||| ||| ||| ||| ||| ||| ||| ||| |||
8624 : GATTAACAGAACATCTCAAGGCAGAAGAATTTCCTTAATACATAACAAAATGGAGTCTCCTAT : 8563
||||||| ||| ||||||||||| ||| ||| ||| ||| ||| ||| |||
63 : GTCTACTTCTTCTACACAGACACAGTAACAATCTGATCTCT---G-TTTCCCCACATTTC : 120
||||||| ||| ||||||||||| ||| ||| ||| ||| |||
8562 : GTCTACTTCTTCTACACAGACACAGTAACAATCTGATCTCTTGCTTTCCCCACATTTC : 8501
||||||| ||| ||||||||||| ||| ||| ||| |||
121 : CCCCTTTCTTTCGACAAAACCAGGCCATCG : 150
||||||| ||| ||| ||| |||
8500 : CCCCTTTCTTTCGACAAAACCACCATCG : 8471
||||||| ||| ||| |||

B)
Alineamiento con HERVW

Query: M03926:170:000000000-GDWDR:1:1101:10001:6171
Target: HERVW provirus
Model: affine:local:dna2dna
Raw score: 691
Query range: 0 -> 149
Target range: 7208 -> 7357

1 : GTACTGGGTTTCTGTGAGAGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGC : 62
| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7209 : GCACCTGGGTTTCTGTGAGAGGGGGACTGAGAGACAGGACTAGCTGGATTTCCTAGGC : 7270
| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
63 : CAACTAAGAACCTAAGCCTAGGTGGGAAGGTGACTGCATCCACCTTAAACACGGGGCTT : 124
|||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
7271 : CAACGAAGAACCTAAGCCTAGCTGGGAAGGTGACTGCATCCACCTTAAACATGGGGCTT : 7332
| ||| ||| ||| ||| ||| ||| |||
125 : GCAACTTAGCTCACGCCGACCAAT : 149
|||| ||| ||| |||
7333 : GCAACTTAGCTCACACCCGACCAAT : 7357
|||| ||| |||

Figura 12: representación del mapeo de las lecturas obtenidas por Illumina a las secuencias de referencia de HERV-K y HERV-W utilizando Bowtie2. Se utilizó Exonrate, una herramienta en línea de tipo BLAST para comparaciones de secuencias (<https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonrate>), para visualizar las homologías entre una determinada lectura y las secuencias HERV-K (A) y HERV-W (B). Se muestran ejemplos típicos. “Query” indica una lectura concreta. La comparación se realizó con HERV-K (indicado como K113, para el caso A) o con el provirus HERV-W (para el caso B). También se indican las puntuaciones brutas y los rangos.

5. DISCUSIÓN

5.1. Análisis de la expresión del gen *HERV-W ENV* mediante qPCR en muestras extraídas en tubos PAXGene

5.1.1. Análisis de la contaminación por DNA genómico y validación del método de extracción mediante qPCR.

Trabajos previos en el laboratorio demostraron que las preparaciones habituales de RNA en tubos PAXGene (utilizando el protocolo del proveedor) estaban contaminadas con DNA. Dado el problema que suponen las contaminaciones de DNA en los análisis de elementos retrovirales mediante qPCR, se introdujeron cambios en el protocolo. Debido a que los elementos retrovirales no tienen intrones, ni con un diseño adecuado de primers podemos distinguir entre cDNA o gDNA, las muestras contaminadas no se pueden analizar. Con el fin de reducir al máximo las contaminaciones de DNA, un cambio que se llevó a cabo en el protocolo es el uso de un paso adicional de DNasa y/o el uso de un kit de PCR específico que contiene una combinación de enzimas que incluye el pretratamiento con DNasa. Las preparaciones descritas en este TFG se realizaron con la última versión de este protocolo. Era de vital importancia asegurarse de que la señal de PCR provenía de RNA retrotranscrita (cDNA) y no de DNA genómico, es decir que estas preparaciones estaban libres de DNA.

Para comprobar que el pretratamiento con DNasa resulta efectivo y asegurarnos de que la señal de amplificación es realmente representativa de la cantidad de RNA de las muestras y no de DNA contaminante, los ensayos de real time qPCR analizan paralelamente muestras tratadas con enzima RTasa y sin enzima.

En base a los resultados obtenidos se puede afirmar que el pretratamiento con DNasa ha resultado efectivo, pues mediante el ensayo de qPCR se ha observado que no ha habido amplificación de *HERV-W ENV* ni de *RPL-19* en muestras procedentes de RNA sin tratar con enzima RTasa. Se concluye que los RNAs extraídos de tubos PAXGene tras ser sometidos al pretratamiento con DNasa se encontraban libres de contaminación por DNA.

5.1.2. Comparación de la expresión de *HERV-W ENV* entre pacientes COVID persistente y pacientes sanos que han superado la enfermedad

Varios estudios han demostrado que la proteína HERV-W ENV inducida durante la infección por SARS-CoV-2, se encuentra en niveles altos en la sangre de pacientes con COVID-19 agudo grave (1,3,4). Sin embargo, también se ha relacionado con el COVID persistente (26). Dado que para este último no se han realizado apenas estudios, en este trabajo se pretendía analizar los niveles de expresión de RNA de *HERV-W ENV* en pacientes con COVID persistente (26).

Como se muestra en los resultados, mediante técnicas de qPCR no somos capaces de detectar diferencias de expresión de RNA de *HERV-W ENV* entre pacientes COVID persistente y pacientes sanos que han superado la enfermedad (controles) a partir de muestras de sangre periférica colectada en tubos PAXGene.

Los tubos PAXGene permiten obtener muestras de forma fácil en atención primaria, pero presentan algunas limitaciones. Un llenado insuficiente del tubo puede dar lugar a un rendimiento del producto que no sea suficiente, obtención de analíticos incorrectos, o una relación sangre-aditivo que no sea adecuada (29). Esto puede afectar a la estabilidad del RNA y por tanto a su conservación.

Ensayos posteriores en el laboratorio deberían analizar si mediante otros métodos de extracción se consigue obtener niveles de expresión más altos de RNA de *HERV-W ENV* que los obtenidos en este ensayo en pacientes con COVID persistente. Por ejemplo métodos basados en el TMReactivos TRIzolTM, que es capaz de aislar RNA de alta calidad (35).

Otra forma de evaluar la expresión de *HERV-W ENV* podría ser midiéndola como proteína en vez de RNA. Si se utilizan muestras de sangre frescas, las proteínas son más estables y mediante citometría de flujo puede analizarse la presencia de la proteína en células sanguíneas (4).

5.2. Evaluación de especificidad del método de hibridación para secuencias *HERV-W*

El método ideal para analizar los HERV es “long-read sequencing”. Como ya hemos explicado anteriormente, la secuenciación por Illumina, da lugar a múltiples problemas a la hora de analizar los HERV, debido a que proporciona “short-reads”. El problema de analizar las secuencias de elementos retrovirales a partir short-reads deriva del hecho de que los elementos retrovirales son secuencias repetidas ligeramente variables entre sí. Las secuencias repetidas siempre presentan problemas técnicos para los alineamientos de secuencia y programas de ensamblado (figura 13). Desde la perspectiva computacional, las repeticiones producen ambigüedades en el alineamiento y el mapeo que pueden causar errores en la interpretación del resultado (36). El mayor de los problemas ocurre cuando una lectura alinea con varias localizaciones del genoma de referencia (multi-reads). Aunque el tipo específico de repetición no influye directamente en el programa de mapeo de lecturas, puede influir en el análisis de datos que se basan en regiones únicas que flanquean las repeticiones (36).

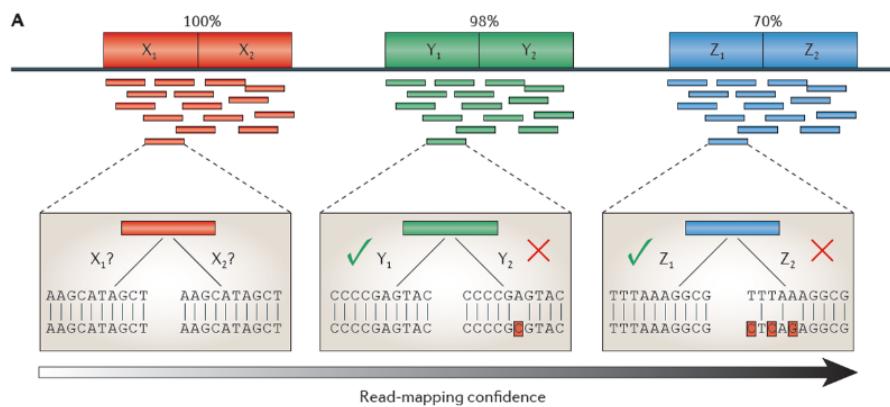


Figura 13: Ambigüedades en el mapeo de lecturas. Se muestra la confianza en la localización de la lectura frente a la similitud entre copias repetidas. A medida que aumenta la similitud entre dos copias de una repetición, disminuye la confianza en la ubicación de cualquier lectura dentro de la repetición. Se muestran tres repeticiones en tandem con dos copias cada una. Justo debajo de estas repeticiones en tandem se muestran las lecturas secuenciadas de estas regiones. Para cada repetición en tandem, se resalta y amplía una única lectura (36).

Para poder analizar los HERV por “long-read sequencing” y evitar los problemas de “short reads”, hay que optimizar los pasos previos del proceso. Para ello el grupo de investigación está desarrollando el método de hibridación con sonda específica de secuencias de *HERV-W ENV* (figura 7). Es fundamental comprobar la especificidad de la sonda y determinar si las secuencias que se están seleccionando son realmente la de *HERV-W ENV* y no las de la familia *HERV-K*.

La primera forma que se ha utilizado para analizar esta especificidad es mediante real time qPCR, sin embargo, los resultados de las melting curves muestran la amplificación de varios productos, por lo tanto, el ensayo no es específico y no podemos fiarnos de los resultados obtenidos. Se demuestra que mediante real time qPCR no podemos evaluar adecuadamente la especificidad de la sonda diseñada.

Si analizamos la homología entre *HERV-W* y *HERV-K* respecto a la sonda diseñada, se puede observar cómo esta sonda presenta una alta homología para la secuencia de *HERV-W* (figura A), mientras que para *HERV-K* apenas hay homología (figura B).

A)

Bait			ATGGCCCTCCCTTATCATACTTTCTCT	28
HERVW	5521	ATCTGTGCCTGCTCTTCAAGCAACAAACCGTGAATGGCCCTCCCTTATCATACTTTCTCT		5580
Bait	29	TTACTGTTCTTACCCCCCTTCGCTCTCACTGCACCCCCCTCCATGCTGCTGTACAACCA		88
HERVW	5581	TTACTGTTCTTACCCCCCTTCGCTCTCACTGCACCCCCCTCCATGCTGCTGTACAACCA		5640
Bait	89	GTAGCTCCCTTACCAAGAGTTCTATGGAGA-----		120
HERVW	5641	GTAGCTCCCTTACCAAGAGTTCTATNAAGAACGCCGCTTCCTGGAAATATTGATGCC		5700

B)

Bait -----
HERVK 721 GCTGACCCCTCTCCCCACAATTGTCTTGTGACCCCTGACACATCCCCCTTTGAGAACAC 780

Bait 17 -----ATACT----- 21
 |||||
HERVK 781 CCACAGATGATCAATAAATACTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTG 840

Bait 22 -----TTTCTCTTACTGT-TCTCT----- 40
 ||||||| ||||| | |||||
HERVK 841 AACGCTGGTTCCCCGGTTCCCTTATTTCTTCTATACTTTGTCTGTGTCTTTTC 900

Figura 13: representación del alineamiento de la sonda diseñada con secuencias de referencia de HERV-W (A) y HERV-K (B). Se puede ver el número de nucleótidos de las secuencias de HERV-W (120) y HERV-K (40) que emparejan con la sonda (bait). El alineamiento se realizó mediante BLAST.

El alineamiento de la figura 13 sugiere que no hay homología suficiente entre *HERV-W* y *HERV-K* para pensar que la sonda diseñada para la secuencia *ENV* de *HERV-W* iba a seleccionar secuencias de *HERV-K* también. Para comprobarlo, viendo que por real time qPCR no se han obtenido resultados fiables, el otro método que se ha utilizado para comprobar la especificidad de la sonda es mandar las librerías que han pasado por el proceso de hibridación a secuenciar por Illumina y analizar si las lecturas obtenidas proceden de *HERV-W* o de *HERV-K*. Como se ha descrito en los resultados, la sonda diseñada para la detección de secuencias *HERV-W ENV* es específica para estas secuencias y las librerías preparadas tras pasar el proceso de hibridación apenas contienen secuencias de *HERV-K*, su presencia es insignificante.

6. CONCLUSIONES

- El pretratamiento con DNase puesto a punto previamente en el laboratorio es útil para la eliminación del DNA que pueda estar contaminando las muestras de RNA extraídos a partir de tubos PAXGene.
- Mediante técnicas de qPCR no somos capaces de detectar diferencias de expresión de *HERV-W ENV* entre muestras de pacientes COVID persistente y controles.
- No se detecta niveles aumentados en COVID, pero si se detecta *HERV-W ENV*.
- Se vislumbra que para determinar si realmente *HERV-W ENV* se expresa o no en muestras de pacientes de COVID persistente, habrá que utilizar métodos diferentes a la qPCR. Puede que otros métodos de extracción de RNA y/o la preservación de RNA en otras condiciones permitan detectar la presencia.
- La especificidad de la sonda diseñada para *HERV-W ENV* y la eficacia del método de hibridación han sido comprobadas mediante la secuenciación por Illumina y análisis de homologías con *HERV-K* y *HERV-W*.

CONCLUSIONS:

- The DNase pre-treatment, previously developed in the laboratory, is useful for remove DNA that may be contaminating RNA samples extracted from PAXGene tubes.
- Using qPCR techniques, we are not able to detect differences between *HERV-W ENV* expression in samples from long COVID patients and controls.
- *HERV-W* is detected, increased levels of COVID are not.
- It appears that to determine whether or not *HERV-W ENV* is actually expressed in samples from persistent COVID patients, other methods instead of qPCR will have to be used. Other RNA extraction methods and/or preservation of RNA under other conditions may be able to detect the presence of *HERV-W ENV*.
- The specificity of the probe designed for *HERV-W ENV* and the efficacy of the hybridization method have been verified by Illumina sequencing and homology analysis with *HERV-K* and *HERV-W*.

7. REFERENCIAS BIBLIOGRÁFICAS

1. Charvet B, Brunel J, Pierquin J, Iampietro M, Decimo D, Queruel N, et al. SARS-CoV-2 awakens ancient retroviral genes and the expression of proinflammatory HERV-W envelope protein in COVID-19 patients. *iScience*. mayo de 2023;26(5):106604. Disponible en: <https://www.sciencedirect.com/science/article/pii/S2589004223006818>
2. Salazar D, Uzquiano M, Rivera G, Velasco E. Mecanismos de transmisión del SARS-CoV-2. *Acta Nova*. noviembre de 2020;9(5-6):773-92. Disponible en: http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S1683-07892020000200008
3. Charvet B, Brunel J, Pierquin J, Iampietro M, Decimo D, Queruel N, et al. SARS-CoV-2 induces human endogenous retrovirus type W envelope protein expression in blood lymphocytes and in tissues of COVID-19 patients [Internet]. medRxiv; 2022 [citado 15 de noviembre de 2023]. p. 2022.01.18.21266111. Disponible en: <https://www.medrxiv.org/content/10.1101/2022.01.18.21266111v2>
4. Balestrieri E, Minutolo A, Petrone V, Fanelli M, Iannetta M, Malagnino V, et al. Evidence of the pathogenic HERV-W envelope expression in T lymphocytes in association with the respiratory outcome of COVID-19 patients. *EBioMedicine*. abril de 2021;66:103341.

5. National Human Genome Research Institute. Retrovirus [Internet]. 2023 [citado 5 de marzo de 2024]. Disponible en: <https://www.genome.gov/es/genetics-glossary/Retrovirus>
6. Carrillo P, Díaz G. HIV y HTLV: dos retrovirus que interfieren con el sistema inmune. Rev Médica Hosp Nac Niños Dr Carlos Sáenz Herrera. enero de 2002;37(1-2):65-9.
7. Grandi N, Tramontano E. Human Endogenous Retroviruses Are Ancient Acquired Elements Still Shaping Innate Immune Responses. *Front Immunol.* 2018;9:2039.
8. Dewannieux M, Heidmann T. Endogenous retroviruses: acquisition, amplification and taming of genome invaders. *Curr Opin Virol.* 1 de diciembre de 2013;3(6):646-56.
9. Sentís C. Retrovirus endógenos humanos: Significado biológico e implicaciones evolutivas. *Arbor.* 30 de mayo de 2002;172(677):135-66.
10. Retrovirus - EMEI [Internet]. 2012 [citado 5 de marzo de 2024]. Disponible en: <https://epidemiologiamolecular.com/retrovirus/>
11. Oja M, Peltonen J, Blomberg J, Kaski S. Methods for estimating human endogenous retrovirus activities from EST databases. *BMC Bioinformatics.* 3 de mayo de 2007;8(Suppl 2):S11.
12. Bruland T. Studies of early retrovirus-host interactions. Viral determinants for pathogenesis and the influence of sex on the susceptibility to Friend murine leukaemia virus infection. 1 de enero de 2003;
13. de Frutos R. Elementos transponibles y organización del genoma de eucariotas. Valencia: Departamento de Genética, Facultad de Ciencias Biológicas. 1997
14. Mangeney M, Renard M, Schlecht-Louf G, Bouallaga I, Heidmann O, Letzelter C, et al. Placental syncytins: Genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. *Proc Natl Acad Sci U S A.* 12 de diciembre de 2007;104(51):20534.
15. Lavialle C, Cornelis G, Dupressoir A, Esnault C, Heidmann O, Vernochet C, et al. Paleovirology of 'syncytins', retroviral env genes exapted for a role in placentation. *Philos Trans R Soc B Biol Sci.* 19 de septiembre de 2013;368(1626):20120507.
16. Tönjes RR, Löwer R, Boller K, Denner J, Hasenmaier B, Kirsch H, et al. HERV-K: The Biologically Most Active Human Endogenous Retrovirus Family. *J AIDS J Acquir Immune Defic Syndr.* 1996;13:S261.
17. Dembny P, Newman AG, Singh M, Hinz M, Szczepk M, Krüger C, et al. Human endogenous retrovirus HERV-K(HML-2) RNA causes neurodegeneration through Toll-like receptors [Internet]. American Society for Clinical Investigation; 2020 [citado 29 de mayo de 2024]. Disponible en: <https://insight.jci.org/articles/view/131093/pdf>
18. Moreno-Martinez L, Macías-Redondo S, Strunk M, Guillén-Antonini MI, Lunetta C, Tarlarini C, et al. New Insights into Endogenous Retrovirus-K Transcripts in Amyotrophic Lateral Sclerosis. *Int J Mol Sci.* enero de 2024;25(3):1549.
19. Li F, Karlsson H. Expression and regulation of human endogenous retrovirus W elements. *APMIS Acta Pathol Microbiol Immunol Scand.* 2016;124(1-2):52-66.
20. Macías S. Human Endogenous Retrovirus Type W in Multiple Sclerosis [Internet]. Zaguan. Repositorio de la Universidad de Zaragoza. 2018 [citado 29 de mayo de 2024]. Disponible en: <https://zaguan.unizar.es/record/87031/files/pdf.pdf>
21. Diccionario de cáncer del NCI. Definición de RT-PCR. NCI [Internet]. 2011 [citado 5 de junio de 2024]. Disponible en: <https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/rt-pcr>

22. Ortega García M^av, Jiménez Mateo O, Sellek Cano R, Bassy Álvarez O, Granja Albarellos C, Cabria Ramos JC, et al. PCR cuantitativa en tiempo real para la amplificación de ADN de Burkholderia mallei: comparación con el método molecular recomendado por la OIE. Sanid Mil. junio de 2017;73(2):85-90.
23. Rubio S, Pacheco-Orozco RA, Gómez AM, Perdomo S, García-Robles R. Secuenciación de nueva generación (NGS) de ADN: presente y futuro en la práctica clínica. Javeriana. [Internet]. 2020;61(2):[citado 4 de junio de 2024]; Disponible en: [https://revistas.javeriana.edu.co/files-articulos/UMED/61-2%20\(2020\)/231062391008/](https://revistas.javeriana.edu.co/files-articulos/UMED/61-2%20(2020)/231062391008/)
24. Valderrama JM, Ortigosa F, Cañas RA. Métodos de secuenciación: tercera generación [Internet]. Encuentros en Biología. Universidad de Málaga. 2020;13(75):15-21 [citado 10 de junio de 2024]. Disponible en: https://encuentros.uma.es/assets/journals/13/175singles/175.4_secuenciacion.pdf
25. Cairoli E. ¿De qué hablamos cuando hablamos de pos-COVID-19?. Rev Clin Esp. diciembre 2021;221(10):614-6. Disponible en: <https://www.revclinesp.es/es-de-que-hablamos-cuando-hablamos-articulo-S0014256521001442?xhy=Dr56DrLjUdaMjzAgze452SzSInMN&rfr=truhgiz&y=kEzTXsahn8atJufRpNPuIGh67s1>
26. Giménez-Orenga K, Pierquin J, Brunel J, Charvet B, Martín-Martínez E, Perron H, et al. HERV-W ENV antigenemia and correlation of increased anti-SARS-CoV-2 immunoglobulin levels with post-COVID-19 symptoms. Front Immunol. 27 de octubre de 2022;13:1020064.
27. Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. Nat Rev Microbiol. marzo de 2023;21(3):133-46.
- 28.BD. Blood Collection Tubes [Internet]. [citado 18 de abril de 2024]. Disponible en: <https://www.bdbiosciences.com/en-us/products/blood-collection/blood-collection-tubes/paxgene-blood-rna-tube.762165>
29. PreAnalytix. Qiagen/BD. PAXgene Blood RNA Tube [Internet]. [citado 9 de junio de 2024]. Disponible en: https://static.bd.com/documents/eifu/ZMG_VDP40479_ES_A_01.pdf
30. ThermoScientific [Internet]. T123-TechnicalBulletin. NanoDropLite. Interpretation of Nucleic Acid 260/280 Ratios. [citado 5 de junio de 2024]. Disponible en: <https://assets.thermofisher.com/TFS-Assets/CAD/Product-Bulletins/T123-NanoDrop-Lite-Interpretation-of-Nucleic-Acid-260-280-Ratios.pdf>
31. Agilent Technologies [Internet]. RNA Integrity Number (RIN)-Standarization of RNA Quality Control. [citado 5 de junio de 2024]. Disponible en: <https://www.agilent.com/cs/library/applications/5989-1165EN.pdf>
32. Thermofisher [Internet]. Mezcla maestra SuperScript™ IV VILO™. [citado 31 de mayo de 2024]. Disponible en: <https://www.thermofisher.com/order/catalog/product/es/es/11766050>
33. Wittwer CT, Hemmert AC, Kent JO, Rejali NA. DNA melting analysis. Mol Aspects Med. 1 de junio de 2024;97:101268.
34. Rao X, Huang X, Zhou Z, Lin X. An improvement of the 2^(-delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. Biostat Bioinforma Biomath. agosto de 2013;3(3):71-85.
35. Fisherscientific [Internet]. Invitrogen™Reactiv TRIZOL™. [citado 11 de junio de 2024]. Disponible en: <https://www.fishersci.es/shop/products/ambion-trizol-reagent-2/p-4918750>
- 36.Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 29 de noviembre de 2011;13(1):36-46