



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

Construcción de un vector para la evaluación de
variantes génicas que modifican el splicing de
RNA

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

B I Ó L O G O

P R E S E N T A:

HUGO ABRAHAM GUEVARA JIMÉNEZ



DIRECTOR DE TESIS:
DR. FEDERICO CENTENO CRUZ
Ciudad Universitaria, CD. MX. 2025



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas

Tesis Digitales

Restricciones de uso

DERECHOS RESERVADOS ©

PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

1. Datos del alumno:

Guevara
Jiménez
Hugo Abraham
55 3990 9376
Universidad Nacional Autónoma de México
Facultad de Ciencias
317211977

2. Datos del asesor

Dr.
Federico
Centeno
Cruz

3. Datos del sinodal 1

Dra.
Itzel Guadalupe
Ramírez
Moreno

4. Datos del sinodal 2

Dra.
Miriam Erandi
Reyna
Fabian

5. Datos del sinodal 3

M. en C.
Luz Vianney
Cortes
González

6. Datos del sinodal 4

Dr.
Jorge Antonio
García
Álvarez

7. Datos del trabajo escrito

Construcción de un vector para la evaluación de variantes génicas que modifican el splicing
de RNA
50
2025

Agradecimientos

Personal

Mi mamá, la persona sin la cual no habría llegado hasta aquí. Gracias por todo el esfuerzo que has hecho por salir adelante y por darnos lo que no pudiste tener. Eres la persona más responsable que he conocido.

Espero poder darte todo lo que te mereces. Desde niño fui una persona curiosa y me generaba un gran interés comprender la existencia, las estrellas, los planetas, el lugar en el que vivimos y cómo es posible que todo ello pueda simplemente estar... tristemente es algo que no se paga muy bien, pero no quería dejar pasar la única oportunidad que tengo de estar en este plano para hacerlo. Solo me queda dar lo mejor de mí e intentarlo. Tal vez no lo logre, pero no iba a poder vivir sin saciar esta energía interna que ha estado en mí desde niño por entender una pequeña parte de todo lo maravilloso que existe en este universo.

A Litzy Vania, gracias por ser mentora en el laboratorio y guiarme en diferentes protocolos. Aprendí mucho gracias a ti, no solo en el laboratorio o con temas relacionadas a la carrera o a la tesis sino en aspectos de la vida. Eres una gran bióloga (y la mejor chef debo decir), muchas gracias por todo dentro y fuera del laboratorio, en verdad.

A Mariana, para mí fue un gesto de agradecimiento por tu amistad incondicional y no solo un acuerdo. Mi primera práctica de campo fue de las experiencias más bellas que he tenido, un montón de momentos que recordaré de un solo viaje. Tener tu amistad fue un apoyo que no se describe. Gracias por todas las veces que me escuchaste y por los consejos con ese toque de realidad. Eres la persona más noble que he conocido.

Hay tantas personas que, aunque no formaron parte de esta etapa, me ayudaron a ser la persona que soy hoy en día, a poder avanzar y llegar hasta aquí. Gracias a: Jorge Yebra, Jossette, Rafa, Emiliano, Abi Abigail, la profe Odette, Kari, Víctor, Miguel, Ángel, Caro, el profe Uziel, Lizeth, Fany.

Quizá muchos de ustedes no lo sepan, pero cada uno dejó en mí una enseñanza, un consejo, su amistad, alguna acción o algunas palabras que marcaron y me ayudaron a seguir adelante. Son personas con cualidades increíbles que agradezco haberme encontrado en algún momento de mi vida. Gracias.

Académico

Al Dr. Federico Centeno Cruz por aceptarme en el laboratorio del que forma parte y darme la oportunidad de ayudarlo en el servicio social a pesar de que no sabía nada de trabajar en el laboratorio cuando lo conocí.

Al Instituto Nacional de Medicina Genómica, por la disponibilidad de los recursos para poder llevar a cabo los experimentos realizados para los fines de esta tesis.

A los sinodales. La Dra. Itzel Guadalupe Ramírez Moreno, Dra. Luz Vianney Cortes González, Dra. Miriam Erandi Reyna Fabian y al Dr. Jorge Antonio García Álvarez, por sus comentarios y sugerencias para la tesis. Excelentes profesores en la carrera con quienes disfrute mucho de sus clases.

“El hecho de que la existencia sea absurda debería ser motivo de risa y no de llanto” — Frase que leí en algún lugar, alguna vez en mi vida, pero que aparentemente no tiene autor, aunque es una referencia a Camus y/o Sartre.

¿1/3?

Resumen

El corte y empalme (*splicing*) consiste en el retiro de intrones y la unión de exones como parte de la maduración del pre-mRNA. Para asegurar la correcta eliminación de los intrones existen elementos *cis* y factores *trans* del *splicing*. Entre los elementos *cis* reguladores se encuentran las secuencias de reconocimiento para eliminación de intrones, así como potenciadores y silenciadores intrónicos y exónicos de *splicing*. Cuando surge una variante en uno de estos sitios esta puede alterar el *splicing*, lo que ocasiona una alteración en la proteína. Detectar si una variante tendrá efecto en el *splicing* es complicado debido a que las secuencias de reconocimiento se encuentran en sitios intrónicos, por lo que, con análisis de exoma, a menos que estén dentro del rango detectable, serán pasadas por alto. No obstante, se ha demostrado que variantes en distancias de hasta 25 pb río arriba y río debajo de las secuencias de reconocimiento pueden generar un *splicing* aberrante. Es debido a esto que los minigenes son herramientas útiles para determinar el efecto de una variante en el *splicing*.

A pesar de esto, los minigenes en vectores de expresión no son de uso comercial. Si se desea adquirir uno, este debe diseñarse y sintetizarse, lo que puede tener un alto costo, especialmente si se sintetiza directamente en un vector de expresión para la evaluación del ARNm y de la proteína. Además, al diseñar un minigen, debe evaluarse cuidadosamente, ya que podrían eliminarse elementos *cis*, como potenciadores intrónicos. Existen una gran cantidad de algoritmos diseñados para el análisis de variantes, su efecto en el *splicing* y la detección de sitios de *splicing*. Sin embargo, estudios han mostrado que variantes predichas como no patogénicas, al evaluarse funcionalmente mostraron tener efecto en el *splicing*.

Aunado a esto, también se han identificado variantes que no estaban reportadas en bases de datos públicas como Clinvar o gnomAD, las cuales posteriormente fueron clasificadas como patogénicas. Es así que la construcción y validación de un minigen en un vector permitirá evaluar funcionalmente el efecto de variantes génicas en el *splicing*.

LISTA DE FIGURAS Y DE TABLAS

No. De Figura	Título	Página
1	Estructura general de un gen	2
2	Código genético	4
3	Relación entre el ADN, ARNm, codones, anticodones, ARNt y aminoácidos	5
4	Modificaciones químicas en nucleótidos del ARNm	7
5	Interacciones entre snRNA y pre-mARN.	9
6	Componentes principales del <i>splicing</i> .	9
7	<i>Splicing</i>	10
8	Tipos de <i>splicing</i> aberrantes causados por variantes.	13
9	Principales componentes de un plásmido	15
10	Diagrama de trabajo	19
11	Estructura minigen	20
12	Vector pUCDIT-Amp con minigen	20
13	Vector pcDNA3.1/myc-His C	21
14	<i>Primers</i> para pruebas	26
15	Restricción de plásmidos	31
16	Vector e inserto purificados	32
17	Placas de agar con células DH5a transformadas	32
18	Restricción enzimática y PCR de las clonas candidato	33
19	Alineamiento del minigen y la clona 4	34
20	Alineamiento del minigen y la clona 5	35
21	Células HEK293T transfectadas con vector de la clona 4	36
22	Producto de la RT-PCR	36
23	Electroferograma de la secuencia <i>forward</i> del ADNc amplificado por PCR	37
24	Alineamiento exones minigen y secuenciación amplicón ADNc	38
25	Vector validado con minigen	38

No. De Tabla	Título	Página
1	Principales componentes de <i>splicing</i>	8

LISTA DE ABREVIATURAS

ACMG	<i>American College of Medical Genetics and Genomics</i>
ADN	Ácido Desoxirribonucleico
ADNc	ADN complementario
ARN	Ácido Ribonucleico
ARNm	ARN mensajero
ARNt	ARN de transferencia
BPS	Secuencia de Punto de Ramificación
CDS	Secuencia Codificante
CMV	Citomegalovirus
CNV	Variantes en el Número de Copias
CSS	Sitio Canónico de <i>Splicing</i>
DMSO	Dimetilsulfóxido
dNTPs	Trifosfatos de Desoxirribonucleósidos
ESE	<i>Exonic Splicing Enhancer</i>
ESS	<i>Exonic Splicing Silencer</i>
GWAS	<i>Genome-Wide Association Study</i>
hnRNP	Ribonucleoproteína nuclear heterogénea
ISE	<i>Intronic Splicing Enhancer</i>
ISS	<i>Intronic Splicing Silencer</i>
LoF	<i>Loss of Function</i>
NCBI	<i>National Center for Biotechnology Information</i>
pb	Pares de Bases
PCR	<i>Polymerase Chain Reaction</i>
PPT	Tracto de Polipirimidinas
RBP	Proteínas de Unión ARN
RT-PCR	Reacción en Cadena de la Polimerasa con Transcripción Reversa
SAP	<i>Shrimp Alkaline Phosphatase</i>
SDV	Variantes que Interrumpen el <i>Splicing</i>
SF	Factor de <i>Splicing</i>
snRNA	ARN nuclear pequeño
snRNP	Ribonucleoproteína nuclear pequeña
SNV	Variantes de un Solo Nucleótido
SREs	Elementos Regulatorios de <i>Splicing</i>
ss	Sitio de <i>Splicing</i>
TF	Factor de Transcripción
TSS	<i>Transcription Start Site</i>
UTR	<i>Untranslated Region</i>
VUS	Variantes de Significado Incierto

Índice

I. Introducción:	1
1.1 Genoma: genes y reguladores de la expresión génica	1
1.2 Expresión génica	3
1.3 Modificaciones postranscripcionales del pre-ARNm	5
1.4 <i>Splicing</i>	7
1.5 Variantes génicas que modifican el <i>splicing</i> de ARN	11
1.6 Vectores de expresión y minigenes	14
II. Justificación	16
III. Objetivos	17
IV. Hipótesis	17
V. Antecedentes y marco contextual	18
VI. Material y métodos	19
6.1 Diseño del minigen	19
6.2 Vector para subclonación	21
6.3 Expansión de plásmidos	22
6.4 Subclonación en pcDNA3.1 myc / His C	23
6.5 Pruebas para determinar la presencia del inserto en el plásmido	26
6.6 Transfección en células HEK293T	28
6.7 RT-PCR	29
6.8 Purificación de producto PCR	30
VII. Resultados	31
VIII. Discusión	39
IX. Conclusión	43
X. Referencias	44

I. Introducción:

1.1 Genoma: genes y reguladores de la expresión génica

En las células el genoma está conformado por ADN, posee una compleja organización y numerosas regiones y procesos reguladores que determinan cómo se almacena, lee, procesa, expresa y transmite dicha información.^{1,2} En humanos y la mayoría de los eucariotas, la información contenida en el genoma se organiza en una estructura llamada cromatina. La cromatina consiste principalmente en ADN asociado a proteínas conocidas como histonas. Cuando el ADN se enrolla en las histonas forma nucleosomas, los cuales son la unidad básica de la cromatina.³ De esta manera se compacta y guarda el material genético en el núcleo, dando como resultado en los humanos la agrupación del genoma en 46 cromosomas: 22 pares autosómicos y 1 par sexual.⁴

La información contenida en el genoma se utiliza a través de secuencias conocidas como genes. Los genes son la unidad básica de información biológica, se forman por segmentos específicos de ADN que sirven para producir un transcrito, que puede ser un ARN o una cadena polipeptídica conocida como proteína.⁵ Sin embargo, aunque los genes son el medio por el cual se almacenan y transmiten los caracteres, dentro de ellos existen regiones que no se expresan. Los segmentos del gen que dan lugar a productos funcionales se les conoce como secuencias codificantes o exones (**Figura 1**).⁶ Los exones son las regiones específicas del gen utilizadas para la expresión génica, mientras que aquellas secuencias del gen que no codifican se les conoce como intrones.⁷ Además de esto, en el ARNm el exón inicial y terminal de un gen contienen regiones no codificantes conocidas como *UTRs* (*untranslated regions*). El resto de los exones de un gen corresponde a la secuencia codificadora (*coding sequence, CDS*), que es la que determina el producto final del gen.⁸

En humanos aproximadamente solo el 2% del genoma es codificante, por lo que anteriormente se pensaba que el resto era ADN basura, sin embargo, múltiples estudios y nuevas tecnologías han demostrado que más del 80% del genoma está involucrado en alguna actividad bioquímica regulatoria de la expresión génica.⁹⁻¹² Se ha identificado una variedad de factores asociados con la expresión génica que han sido clasificados en: elementos *cis* y factores *trans*. Los elementos *cis* regulatorios normalmente son secuencias presentes en el mismo cromosoma que el gen que

regulan, por lo que afectan la expresión génica únicamente en las regiones que se encuentren dentro de éste. Algunos ejemplos son: sitios de inicio de la transcripción (*Transcription Start Site*: TSS), promotores flanqueantes, aisladores (*insulators*) unidos a la proteína CTCF, potenciadores (*enhancers*) o silenciadores (*silencers*) de la transcripción, así como marcas epigenéticas o regiones no traducidas (*UTR*).¹³⁻¹⁶

Por otro lado, los factores *trans* regulatorios actúan de manera global, no son parte de la secuencia de ADN y regulan la expresión génica en cada par homólogo de cromosomas o una vez que se un gen ha sido transcrito.^{14, 17} Ejemplos de factores *trans* regulatorios incluyen factores de transcripción (*Transcription Factors*: TF), proteínas remodeladoras de la cromatina, ARNs no codificantes, como ARNs pequeños no codificantes (*Small noncoding RNAs*: *sncRNAs*) o ARNs largos no codificantes (*Long NonCoding RNA*: *lncRNA*), entre otros.^{12, 13, 19-22}

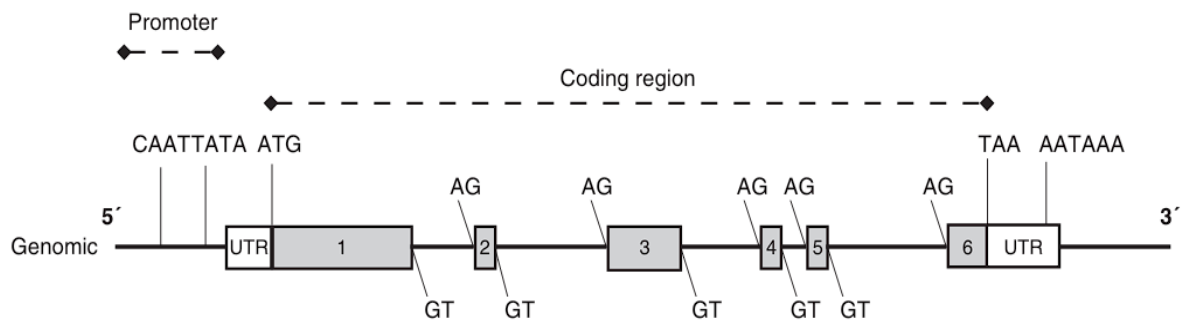


Figura 1. Estructura general de un gen. En el extremo 5' se observa una región promotora que incluye secuencias características como la caja TATA y la secuencia CAAT, esenciales para la iniciación de la transcripción. El gen está compuesto por una serie de exones, representados en bloques numerados, los cuales constituyen las secuencias codificantes. Se intercalan los intrones, representados en líneas, constituyen secuencias no codificantes que serán eliminadas durante el *splicing*. Los sitios de empalme entre exones e intrones están marcados por las secuencias "GT" al inicio del intrón y "AG" al final, lo que asegura la correcta eliminación de los intrones y la unión de los exones. Las UTR (Untranslated Regions) presentes en los extremos del ARNm, son regiones no traducidas del extremo 5' del exón inicial y del extremo 3' del exón final. (Imagen tomada de Dixon, 2010, figura 1.3)

1.2 Expresión génica

La expresión génica es el proceso mediante el cual la célula la información contenida en sus genes para generar productos funcionales, como ARNs o proteínas. El primer proceso es la transcripción, que consiste en la formación de una molécula de ARN a partir de un molde de ADN proveniente de un gen; seguido del procesamiento y maduración del ARN sintetizado durante la transcripción; luego sigue la traducción (cuando el producto es una proteína) y finalmente las modificaciones postraduccionales, necesarias para que las proteínas puedan llevar a cabo sus funciones las proteínas.¹⁵

La transcripción ocurre en el núcleo de la célula, constando de tres etapas: iniciación, elongación (proceso en el que la ARN polimerasa sintetiza el ARN) y terminación. Para que este proceso comience, primero debe desenrollarse una porción de cromatina por proteínas remodeladoras de la cromatina, lo que permite el acceso al ADN.²³⁻²⁵ Una vez que el ADN se vuelve accesible, se requiere de una enzima llamada ARN polimerasa (habiendo tres tipos, siendo la ARN polimerasa II responsable de sintetizar la mayoría de los genes que codifican proteínas). Esta enzima se une a la cadena antisentido en el promotor del gen con la ayuda de proteínas llamadas factores generales de transcripción.^{7, 26, 27} El final de la transcripción ocurre entre 500 - 2000 nucleótidos después de una secuencia de poliadenilación.²⁶ La cadena de ARN formada recibe el nombre de pre-ARNm o transcrito primario.²⁸

Debido a la interrupción que presentan los genes por la presencia de secuencias no codificantes, es necesario procesar el pre-ARNm. El tamaño de los exones humanos tiene una longitud de entre 50 a 300 pares de bases (pb), mientras que los intrones varían, yendo de docenas de pb hasta miles de bases. Este promedio es menor en eucariotas inferiores y mayor en eucariotas superiores.²⁹ Debido a esto el pre-ARNm debe someterse a una serie de reacciones con la para producir un ARN maduro o un ARN no codificante. Estas reacciones incluyen la adición de una 7-metilguanosina, un corte y empalme (*splicing*), poliadenilación y en algunas ocasiones la edición del mismo ARN.³⁰

Cuando el ARN ha madurado se denomina ARN mensajero (ARNm). Este ARNm sale del núcleo al citoplasma, donde es reconocido por un ribosoma para iniciar la síntesis de la proteína. La propiedad de un ARNm para dirigir la producción de una secuencia específica de aminoácidos se ha establecido a través de un conjunto de patrones descritos en un código

genético, en el que la información proveniente del ARNm se lee en grupos de tres nucleótidos consecutivos llamados codones. Cada codón especifica un aminoácido que se utiliza para formar una cadena polipeptídica naciente. Existen 64 combinaciones únicas de codones. Como se puede observar en la **Figura 2**, hay varias posibles combinaciones para un mismo aminoácido, lo que hace que el código genético sea redundante o degenerado, dado que hay 20 aminoácidos y 64 posibles combinaciones.⁶

AAA	Lys	CAA	Gln	GAA	Glu	UAA	STOP
AAG		CAG		GAG		UAG	
AAC	Asn	CAC	His	GAC	Asp	UAC	Tyr
AAU		CAU		GAU		UAU	
ACA		CCA		GCA		UCA	
ACG	Thr	CCG	Pro	GCG	Ala	UCG	Ser
ACC		CCC		GCC		UCC	
ACU		CCU		GCU		UCU	
AGA	Arg	CGA		GGA		UGA	STOP
AGG		CGG	Arg	GGG	Gly	UGG	Trp
AGC	Ser	CGC		GGC		UGC	Cys
AGU		CGU		GGU		UGU	
AUA	Ile	CUA		GUA		UUA	Leu
AUG	Met	CUG	Leu	GUG	Val	UUG	
AUC		CUC		GUC		UUC	Phe
AUU		CUU		GUU		UUU	

Figura 2. Código genético. Del lado izquierdo de las barras grises se presentan los codones y del lado derecho los respectivos aminoácidos que codifican de manera abreviada. Existen 61 codones que codifican para un aminoácido específico y tres codones que funcionan como señales de terminación para la adición de aminoácidos. Generalmente el codón AUG, que codifica para la metionina (Met), es el codón de inicio para la traducción. Las figuras marcadas en color azul representan secuencias que cambian de significado en la traducción mitocondrial. (Imagen modificada de Strachan, 2019, figura 1.29)

Los ribosomas trabajan con ARNs de transferencia (ARNt), cada uno asociado a un aminoácido específico, y a su vez, cada aminoácido posee un anticodón complementario a los codones del ARNm. De esta manera, cuando el ARNt reconoce al ARNm mediante complementariedad de bases. Los aminoácidos correspondientes se incorporan a la cadena polipeptídica naciente (**Figura 3**). Cuando un ARNm sale del núcleo, las subunidades del ribosoma, 40S y 60S, junto con otras moléculas como los factores eucarióticos de inicio de la traducción (eIFs), el metionil-ARNt y la subunidad 40s, se unen a la región 5'UTR del ARNm para formar el complejo ribosomal encargado de la traducción.^{31, 32, 33}

Cada aminoácido tiene propiedades fisicoquímicas específicas, y es el conjunto de todos ellos que, cuando se unen para formar una proteína, le confieren propiedades únicas que le permiten realizar sus funciones de manera correcta dentro de la célula o en un sistema biológico más complejo.^{34, 35, 36}

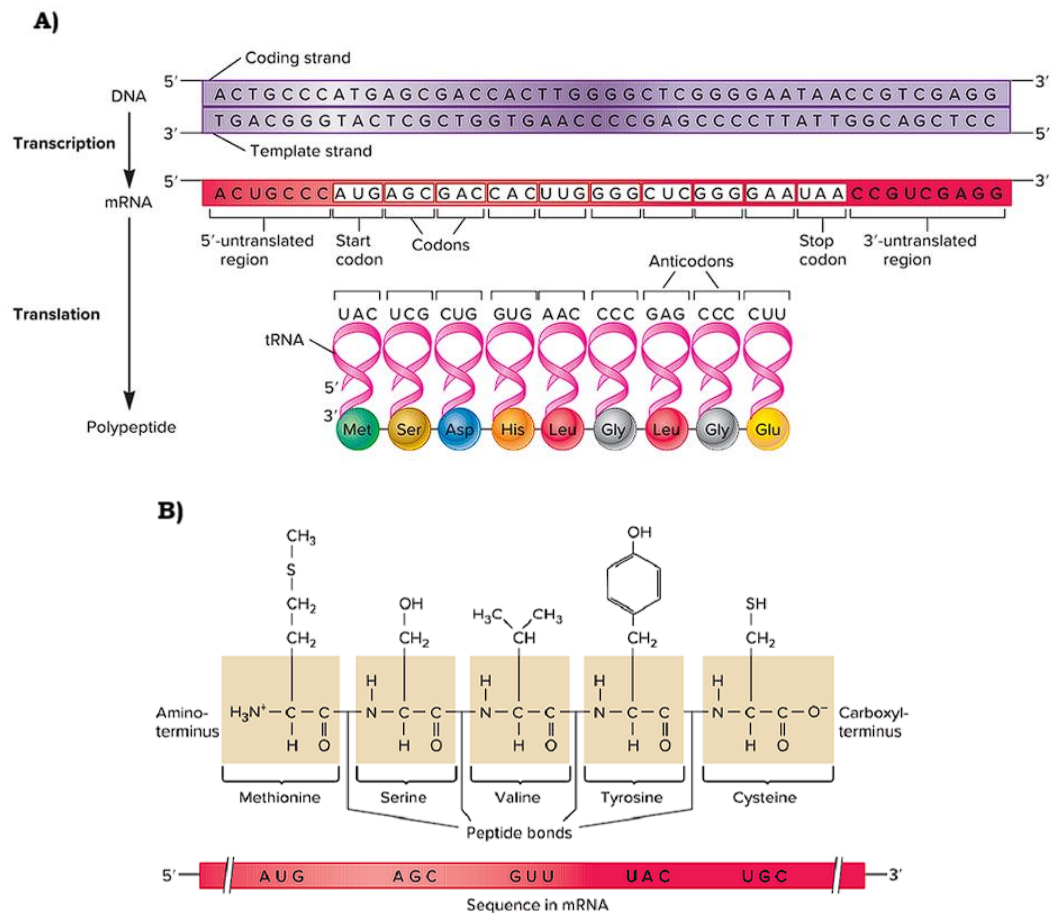


Figura 3. Relación entre el ADN, ARNm, codones, anticodones, ARNt y aminoácidos. (A) En morado se muestra una secuencia de ADN de doble cadena. La cadena sentido o codificante, orientada 5'-3'; la cadena antisentido o molde, orientada 3'-5', siendo esta última la que sirve como base para formar al ARNm representado en color rojo. Tanto la cadena codificante como el ARNm tienen la misma secuencia de bases (con excepción de que las T del ADN se cambian por U en el ARNm). El sitio de inicio de la traducción es en el codón AUG, para ello el ARNt representado en rosa, sirve como un conector entre el anticodón (que se une por complementariedad de bases al codón) y el aminoácido específico que ha de incorporarse a la cadena polipeptídica. (B) En beige se representan los aminoácidos codificados por una secuencia de nucleótidos de un ARNm. Cada aminoácido tiene un grupo funcional específico el cual le confiere características fisicoquímicas específicas. (Imagen modificada de Brooker, 2021, figura 13.3 y 13.4b).

1.3 Modificaciones postranscripcionales del pre-ARNm

Las modificaciones que ocurren en el ARN pueden tener lugar cuando este termina de sintetizarse (post-transcripcionales) o durante la síntesis (co-transcripcionales). Estas modificaciones en el ARN tienen como finalidad otorgarle estabilidad, modular el procesamiento, facilitar la exportación fuera del núcleo, determinar su localización celular y juegan un papel importante en la traducción.³⁷ El ARNm pasa por una gran cantidad de modificaciones bioquímicas, identificándose hasta 170 tipos diferentes³⁸, en las cuales se modifican nucleótidos o se les añaden marcas químicas. Entre las modificaciones más estudiadas, está la adición de una metilguanósina en el extremo 5', el corte y empalme

(*splicing*) y la poliadenilación en el extremo 3'. Además de estas modificaciones, se destacan otras como la N6-metiladenosina (M6A), 5-metilcitosina (M5C), N1-metiladenosina (M1A), N3-metilcitosina (M3C), pseudouridina (ψ), 2'-O-metilación (2'-O-Me), 5-hidroximetilcitosina (hm5C) (**Figura 4**), las cuales son importantes de mencionar, ya que cuando estas modificaciones ocurren de manera anormal generan condiciones como desórdenes mitocondriales, obesidad, diabetes, deficiencias neurológicas y enfermedades metabólicas dado que sirven como un puente para la interacción de proteínas con el ARNm.^{37, 39}

Entre las modificaciones que sufre el ARNm, la primera es la adición de una 7-metilguanosina (m7G o cap) en el extremo 5', esta se coloca tan pronto como se han sintetizado los primeros 25-30 nucleótidos del ARN naciente. Su función es proteger al ARN recién sintetizado para evitar su degradación, reclutar proteínas (entre ellas las que realizan el *splicing*), promover el inicio de la traducción, marcar los ARNm como "propios" para su protección ante mecanismos del sistema inmune dentro de la célula.⁴⁰ Para la adición del cap 5' se requiere de tres procesos enzimáticos: ARN trifosfatasa (TPasa), guaniltransferasa (GTasa) y guanina -N7-metiltransferasa (guanina-N7 Mtasa).⁴¹

El paso final de la transcripción es la escisión endonucleotídica, la cual se produce de 10 a 30 nucleótidos río abajo de una secuencia señal AAUAAA, seguido de la adición de adeninas en el extremo 3' (cola de poli A) que junto con la estructura del cap 5' potencian el inicio de la traducción. Esta modificación se realiza a través del reconocimiento de una secuencia señal AAUAAA, la cual es altamente conservada en mamíferos. Esta modificación es crucial para la estabilidad del ARN y evitar su degradación. La escisión requiere múltiples proteínas, entre ellas: el factor de especificidad de escisión/poliadenilación (**CPSF**), que se une a la secuencia AAUAAA; el factor de estimulación de la escisión (**CsTF**), que se une a sitios ricos en U/GU río abajo; y, en mamíferos dos factores de escisión más; CFIm y CFIIIm.^{40,42} La poli A polimerasa (**PAP**) agrega la cola de poli A la cual tiene una longitud de 50 - 250 nucleótidos al extremo 3' en un grupo -OH expuesto debido a la escisión.⁴³ Tanto CPSF como CsTF son necesarios para la escisión y la adenilación. CPSF reconoce al ARN y facilita el reclutamiento de PAP (específicamente la endonucleasa que realiza la escisión en mamíferos es la CPSF-73).^{42, 43}

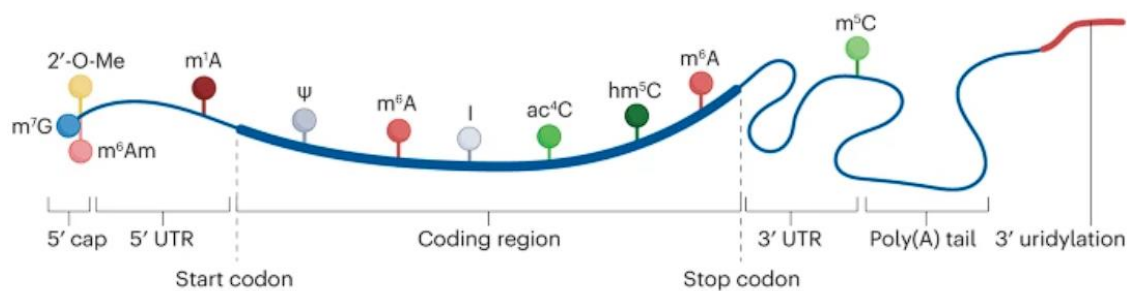


Figura 4. Modificaciones químicas en nucleótidos del ARNm. Se ejemplifican algunas modificaciones químicas que presenta un ARNm compuesto de una secuencia codificante (CDS) y sus regiones no traducidas (UTR) 5' y 3'. Entre estas modificaciones se encuentra el 5' cap el cual consta de dos modificaciones; la adición de un grupo metilo al oxígeno en la posición 2 en la ribosa del nucleótido (2'-O-Me) y si el primer nucleótido transcrito es una adenosina, se instala la marca N6,2'-O-dimetiladenosina (m6Am). N6-metilguanosina (m6A) ocurren lejos de sitios de empalme. Otras modificaciones en los ARNm eucariotas incluyen pseudouridina (Ψ) y de manera menos común N1-metiladenosina (m1A), 5-metilcitosina (m5C), 5-hidroximetilcitosina (hm5C), N4-acetilcitosina (ac4C) e inosina (I). La uridilación 3' del ARNm se indica en rojo. (Imagen modificada de Delaunay, S, et al, 2024).

1.4 Splicing

El *splicing* es un proceso que ocurre en el núcleo celular mediante el cual se eliminan los intrones de un pre-ARNm y se unen los exones. Se han identificado diferentes componentes necesarios para el *splicing*, siendo el espliceosoma uno de los principales. En la mayoría de los eucariotas existen dos tipos de espliceosomas: el mayor o dependiente U2 y el menor o dependiente de U12. El espliceosoma mayor (complejo principal), participa en el *splicing* de la mayoría de los genes, está conformado por más de 150 proteínas y cinco subunidades nucleares de ARN (*Small Nuclear RNAs*: snRNAs): U1, U2, U4, U5 y U6, que cuando se unen a proteínas reciben se denominan *snRNPs* (*Small Nuclear Ribonucleoprotein*).⁴⁴ Otros componentes esenciales para el *splicing* son secuencias consenso denominadas como sitios canónicos de *splicing* (*Canonical Splice Sites*: CSS), como los sitios de *splicing* (ss) ubicados en los límites de un exón y un intrón (**Figura 1**), la secuencia de punto de ramificación (*Branch Point Sequence*: BPS) y un tracto de polipirimidinas (*Polypyrimidine Tract*: PPT) cercano a BPS. También participan elementos reguladores de *splicing* (*Splicing Regulatory Elements*: SREs) clasificados de acuerdo con el efecto en el *splicing* y a su localización, y proteínas de unión a ARN (*RNA Binding Proteins*: RBPs), esenciales para el correcto funcionamiento del proceso de *splicing* (Tabla 1).^{44 - 52}

Tabla 1. Principales componentes de *splicing*.

Componente de <i>splicing</i>	Tipo de regulador
Espliceosoma	Trans
Sitio de <i>splicing</i> 5' (donador de <i>splicing</i> /5'ss)	Cis
Sitio de <i>splicing</i> 3' (aceptor de <i>splicing</i> /3'ss)	Cis
Secuencia de punto de ramificación (BPS)	Cis
Tracto de polipirimidinas (PPT)	Cis
Proteínas SR	Trans
Proteínas hnRNP	Trans
Potenciador Exónico de <i>Splicing</i> (ESE)	Cis
Potenciador Intrónico de <i>Splicing</i> (ISE)	Cis
Silenciador Exónico de <i>Splicing</i> (ESS)	Cis
Silenciador Intrónico de <i>Splicing</i> (ISS)	Cis

Los sitios de *splicing* son secuencias que se encuentran en los intrones de un gen, cerca de los límites entre exones e intrones. Se nombran de acuerdo con su posición relativa al intrón: el sitio de *splicing* 5' (5' ss) se encuentra al inicio del intrón, mientras que el sitio de *splicing* 3' (3' ss) está al final. El sitio de *splicing* 5' incluye una secuencia consenso GU (o GT en el ADN), y el sitio de *splicing* 3' contiene una secuencia consenso AG. Esto establece la regla GU-AG, donde GU sirve como sitio donador y AG como sitio aceptor en el proceso de *splicing*.⁴⁷

Además de los sitios de *splicing*, existe una secuencia intrónica conservada llamada sitio de ramificación, identificada con una adenina (A), localizada a no más de 40 nucleótidos río arriba del sitio de *splicing* 3'. Entre el sitio de *splicing* 3' y el punto de ramificación se encuentra un tracto de polipirimidinas (PPT) rico en uracilos (timinas en el ADN) o citocinas.⁵⁰ Cada uno de estos sitios es reconocido por proteínas de unión a ARN (RBPs) denominadas factores de *splicing* (*Splicing Factors*: SF), uniéndose a través de interacciones de ARN-ARN o de ARN-proteína (**Figura 5**). Entre las más destacadas están las proteínas SR o hnRNP (**Figura 6**), sin embargo, existen gran cantidad de RBPs que participan en el proceso.^{51,52}

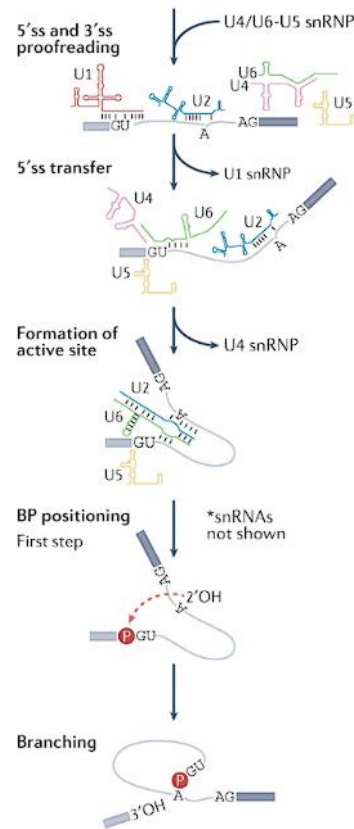


Figura 5. Interacciones entre snRNA y pre-ARNm. En el *splicing*, las interacciones de las proteínas de unión a ARN (RBPs) y de los snRNAs no ocurren únicamente con los sitios consenso (GU-AG), sino también con los sitios circundantes. Como se puede ver en representación, cada snRNA interactúa de tal forma que las regiones circundantes a los sitios de *splicing* también son clave, sin embargo, no se encuentran identificadas secuencias específicas para estos. (Imagen modificada de Rogalska, M.E, et al, 2023).

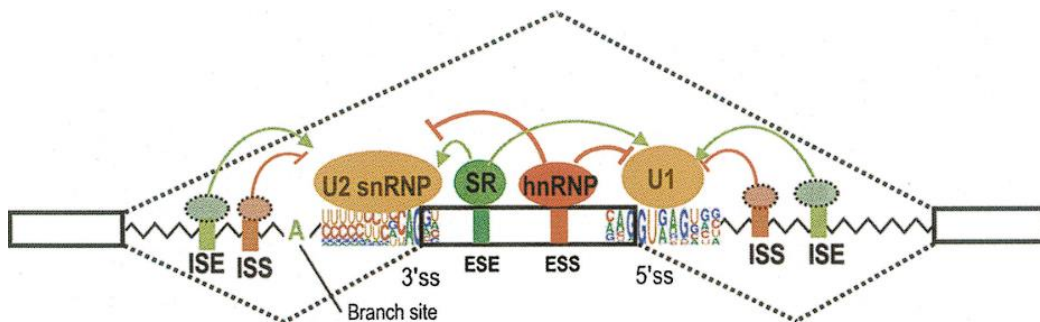


Figura 6. Componentes principales del *splicing*. Las proteínas SR y hnRNP son un tipo de proteína de unión a ARN (RBPs) que se unen a potenciadores y silenciadores de *splicing* respectivamente, tanto aquellos que son exónicos como intrónicos, facilitando o impidiendo la unión del espliceosoma. Las secuencias muestran los sitios de *splicing* (ss), el sitio de ramificación (BPS) y el tracto de polipirimidinas, el cual está localizado río arriba del 3'ss. (Imagen tomada de Wang, Z, et al, 2008).

Durante el proceso de *splicing* (**Figura 7**), se forman cuatro complejos entre el pre-ARNm y el espliceosoma. El primero es el complejo temprano (E). El snRNP UI reconoce y se une a la secuencia complementaria AG-GU en el sitio donador de *splicing* (extremo 5' del intrón). Al mismo tiempo, el factor de *splicing* SF1 se une al sitio de ramificación (BPS). SF1 es

reconocida y unida por la proteína U2AF, misma que también se une al tracto de polipirimidina ubicado entre el punto de ramificación y el sitio de *splicing* 3' (3'ss) del intrón.⁵³ Posteriormente, SF1 es desplazada del punto de ramificación por snRNP U2, formando el complejo dependiente de ATP (A). La interacción entre el punto de ramificación y la proteína snRNP U2 es estabilizada por helicasas específicas de ARN (Prp5 y Sub2) que desplazan a SF1 permitiendo la unión de snRNP U2, lo que sirve como señal para el reclutamiento del tri-snRNP U4/5/6 y la formación del complejo B (espliceosoma precatalítico). La acción adicional de helicasas de ARN provoca un cambio en la conformación del espliceosoma, lo que conduce a 1) la liberación de snRNPs U1 y U4, 2) la interacción de snRNP U2 y U6, 3) la formación de un lazo en el pre-ARNm y 4) la creación del complejo C.⁵⁰ En este complejo, se llevan a cabo dos reacciones de transesterificación: se elimina el intrón y se unen los extremos de los exones. U2, junto con otras proteínas asociadas son liberadas, dejando el BPS expuesto, lo que permite un ataque nucleofílico por parte del grupo 2'-OH de la adenina correspondiente a BPS al sitio de *splicing* 5' (5' ss). Esto resulta en la liberación del extremo 5' del intrón y la formación de un lazo (lariat) intrónico, creando el complejo C. Luego, Prp8, con U5 y U6, dirige el siguiente paso en el que el 3'-OH del exón 5' ataca al sitio de *splicing* 3' (3' ss). Finalmente, los exones se ligan, el intrón en forma de lazo es liberado y degradado, y los snRNPs restantes junto con otros factores se disocian.⁵⁴

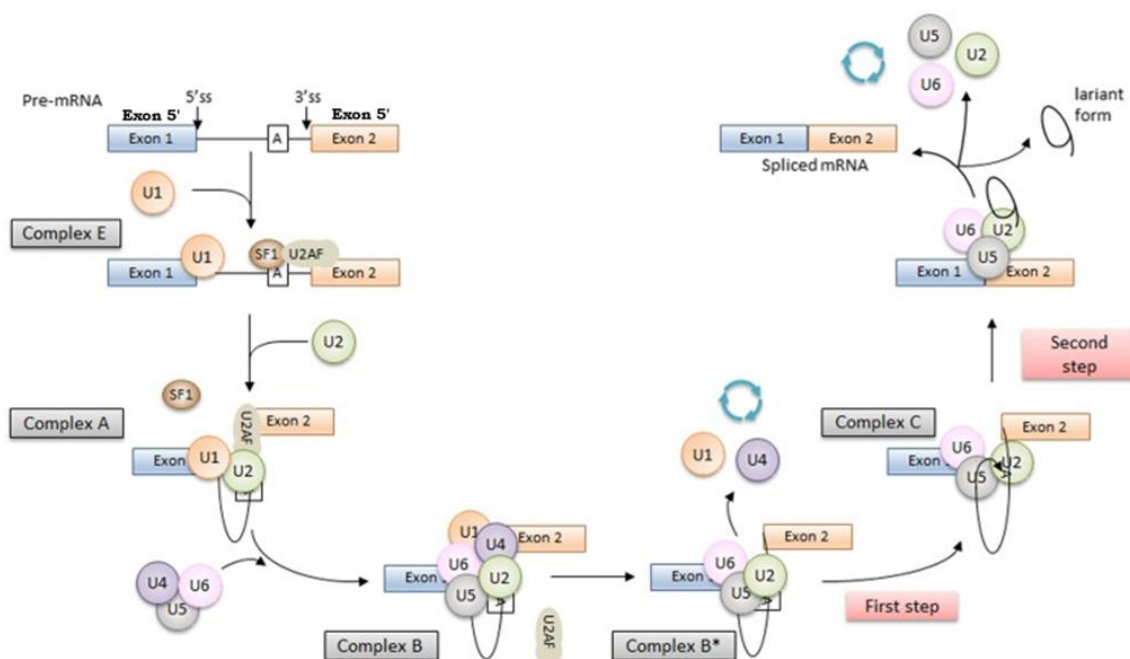


Figura 7. Splicing. Esquema del *splicing* donde se muestran los principales componentes, como las proteínas básicas de reconocimiento de ARN que permiten el reclutamiento de los snRNP del espliceosoma. En rojo se muestran los procesos de transesterificación una vez se forma el complejo catalítico. (Imagen modificada de Anna, A, et al, 2018).

Además del mecanismo básico de *splicing* los elementos reguladores de *splicing* (SREs) funcionan promoviendo o inhibiendo la inclusión del exón en el que residen (cuando son exónicos), o mejoran o inhiben el uso de sitios de empalme adyacentes o la inclusión de exones desde una ubicación intrónica (cuando son intrónicos).⁵⁰ Los SREs funcionan como anclaje para RBPs, siendo las más comunes las proteínas ricas en serina y arginina (SR) y las ribonucleoproteínas nucleares heterogéneas (hnRNP). Anteriormente se había identificado que las proteínas SR se unen a elementos SRE potenciadores, mientras que las hnRNP se unen a elementos SRE silenciadores. Sin embargo, se ha demostrado que ambas pueden tener funciones potenciadoras o silenciadoras dependiendo del sitio en el ARN al que se unan.^{29,51} Funcionan de tal manera que activan o suprimen el reconocimiento de los sitios de *splicing* por parte del espliceosoma.⁴⁷

Los exones pueden ser constitutivamente empalmados, lo que significa que están presentes en cada ARNm proveniente de un pre-ARNm; o bien pueden ser empalmados alternativamente, donde hay variaciones en la forma en cómo se empalman los exones con ayuda de los SREs y las RPBs, generando una mayor diversidad de proteínas a partir de un solo gen.^{53, 54, 55, 56}

1.5 Variantes génicas que modifican el *splicing* de ARN

Una variante génica se puede definir como un cambio permanente en el ADN que difiere de una secuencia de referencia y es potencialmente heredable.⁴ Las variantes génicas por su efecto en la proteína resultante se consideran; sinónimas o silenciosas cuando no alteran la secuencia de aminoácidos, normalmente esto ocurre cuando el cambio es en el tercer nucleótido del codón; mientras que son no sinónimas cuando el cambio es en el primer o segundo nucleótido del codón, lo que resulta en un cambio en el aminoácido producido. Esto incluye las variantes de sentido cambiado o erróneo (*missense*), que sustituyen un aminoácido por otro, y las variantes sin sentido (*nonsense*), que introducen un codón de paro prematuro. También se encuentran las variantes de corrimiento en el marco de lectura (*frameshift*), causadas por variantes de tipo indel (inserción y/o delección) que provocan un desplazamiento en la lectura de los codones, alterando la secuencia de aminoácidos a partir del punto de cambio.^{26,57}

Por otro lado, las variantes que pueden modificar el *splicing* comprenden aquellas que alteran sitios de *splicing* o regiones circundantes. Estas variantes pueden ser sustituciones (generalmente variantes de un solo nucleótido; *Single Nucleotide Variant: SNV*), inserciones o delecciones (indels), inversiones, translocaciones, inserción de elementos transponibles y

variantes en el número de copias (*Copy Number Variants: CNV*), como deleciones o duplicación de segmentos.^{56, 58} A diferencia de las variantes que alteran la proteína, se ha observado que incluso variantes sinónimas (silenciosas) aunque no afecten un aminoácido, pueden alterar el *splicing*, resultando en un ARN aberrante.⁵⁹

De acuerdo con el *American College of Medical Genetics and Genomics (ACMG)*, las variantes génicas se pueden clasificar conforme a su patogenicidad en: patogénicas, probablemente patogénicas, de significado incierto (*Variants of Uncertain Significance: VUS*), probablemente benignas y benignas.⁶⁰

A través de diversos estudios de asociación del genoma completo (*Genome Wide Association Studies: GWAS*), se han identificado más de 200,000 variantes génicas asociadas a rasgos complejos humanos, tales como enfermedades, características físicas o comportamientos.⁶¹ Dentro de las cuáles, el 90% se han identificado en regiones no codificantes del genoma ⁶², donde la mayoría de los mecanismos relacionados a cómo afectan la expresión génica aún no se han caracterizado.⁶¹

La desregulación del *splicing* suele resultar en la omisión de exones (generando una proteína trunca), la retención de intrones (traduciendo regiones no codificantes) o uso de sitios de *splicing* alternativo (alterando la composición de la proteína).⁴⁶ Los sitios alternativos utilizados debido a una variante que evita el reconocimiento de los sitios de *splicing* reciben el nombre de sitios crípticos o no canónicos, los cuales son sitios que se asemejan a una secuencia canónica. Los sitios crípticos suelen pasar desapercibidos la mayor parte del tiempo debido a que no son necesarios, pero si ocurre alguna alteración, como la presencia de una variante en el sitio de *splicing* canónico, el reconocimiento por parte del espliceosoma y proteínas puede confundirse, utilizando este sitio críptico, generando cortes donde no debería (**Figura 8**).⁵⁴ Al no retirar los intrones o al realizarse de manera incorrecta debido a una variante génica que altera este proceso, las secuencias codificadas potencialmente podrían tener consecuencias patológicas, como defectos genéticos, desórdenes neurológicos, cáncer, entre otros.^{56, 58, 59, 61 - 65}

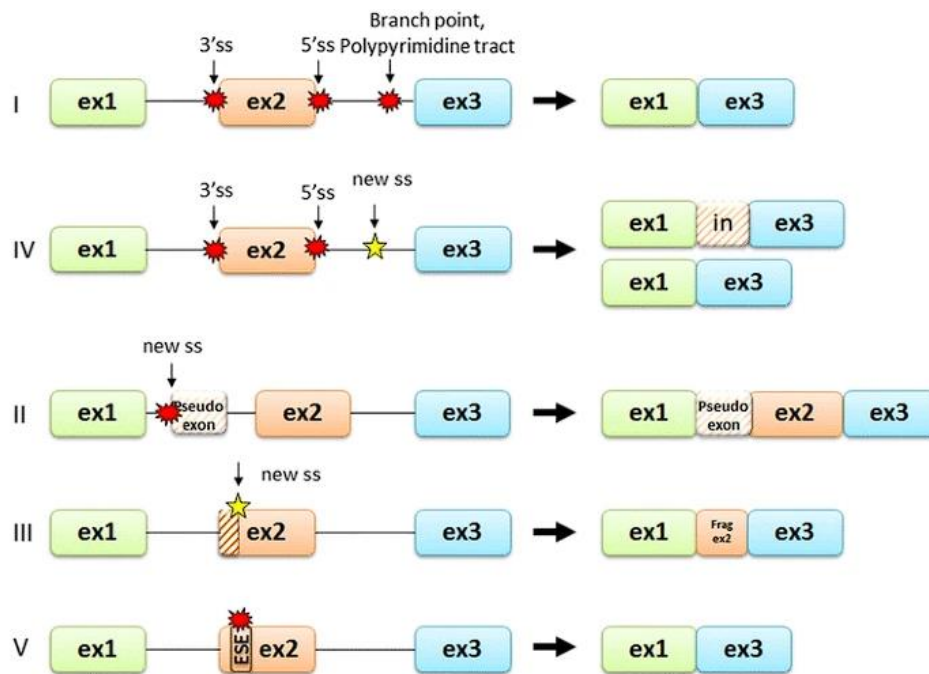


Figura 8. Tipos de *splicing* aberrantes causados por variantes. **I.** Las variantes ocurren en los sitios de *splicing* lo que impide su reconocimiento por la maquinaria de *splicing*, como resultado, el exón 2 es completamente saltado. **II.** Una variante crea un nuevo sitio de *splicing* dentro de un intrón, generando un **pseudoexón** (una secuencia que normalmente no es un exón) erróneamente es reconocido por la maquinaria de *splicing* y es incluido. **III.** Una variante activa un sitio críptico dentro del exón 2. Esto da lugar a que solo parte del exón 2 sea incluido en el ARNm final y se elimine el resto del exón. **IV.** Variantes en los sitios de empalme canónicos que resultan en el uso de un sitio de empalme críptico exónico o intrónico que conduce a la inclusión de un fragmento de intrón o la omisión de un fragmento de exón. **V.** Una variante afecta un ESE (potenciador de *splicing* exónico) en el exón 2, lo que impide el reconocimiento adecuado de este exón por la maquinaria de *splicing*, como consecuencia, el exón 2 es omitido. (Imagen tomada de Anna & Monika, 2018).

Las variantes en los sitios de *splicing* a menudo se identifican como variantes de pérdida de función (*Loss of Function: LoF*) debido al impacto significativo que tienen en la función de la proteína.^{60,66} Estas variantes alteran el *splicing* a través de modificaciones en las señales de reconocimiento en los sitios de *splicing*, que son motivos de unión para proteínas involucradas en el proceso.²⁹ Se ha observado que variantes localizadas a distancias de hasta ± 25 pb de la secuencia señal también pueden tener efectos similares.⁶⁷ Un estudio realizado por Lord et al. (2019) encontró que, en un total de 7833 probandos con trastornos del desarrollo, el 27% de las variantes patogénicas de *nov* de *splicing* se encontraron en sitios no canónicos, mientras que el 73% se encontraron en sitios de *splicing* canónicos. Se resalta que del 35 - 40% de las variantes patogénicas en sitios no canónicos no estaban registradas en bases de datos públicas.⁶⁸

Otro estudio realizado por Blakes et al. (2022) analizó muestras de genoma completo de 38,688 individuos del Proyecto de los 100,000 Genomas, pertenecientes a la rama de enfermedades raras, de los cuales 26,660 eran padres no afectados de personas con enfermedades raras.

Dentro del estudio, se identificaron un total de 258 variantes de *novο* cercanas a sitios de *splicing* o de ramificación. De estas 258 variantes, 83 (32 %) eran de tipo LoF o variantes de *novο* que alteran la proteína, y 129 (50 %) no estaban clasificadas. También se confirmaron cuatro diagnósticos moleculares en los que se detectó omisión de exón por variantes cercanas a sitios de *splicing*.⁶⁶

Diversos estudios han demostrado la importancia de analizar y caracterizar variantes en sitios no canónicos de *splicing* como lo son aquellos que están cerca de las secuencias señal o en secuencias SREs. En un estudio realizado por Chong et al. (2019) en donde se analizaron 27,733 variantes dentro o cerca de exones, donde el 3.8% (1050) se identificaron como variantes que alteran el *splicing* (*Splice-Disrupting Variants*: SDVs), y el 83% de SDVs se localizaron fuera de secuencias canónicas de *splicing*.⁶⁹ Esto se debe a que las posiciones -3 a +6 en el sitio de *splicing* 5' y las posiciones -20 a +3 en el sitio de *splicing* 3' no coinciden siempre con una "secuencia consenso" idealizada y al mismo tiempo son las posiciones en las que se suelen ubicar los elementos *cis* reguladores del *splicing* (SREs), aunque estos no están bien definidos.⁶³

Las variantes en sitios no canónicos son poco reconocidas y se destaca que la interpretación mejorada de variantes de *splicing* es un área de gran promesa para la medicina genómica aún por explorar y descifrar. Estos y otros estudios enfocados al análisis de variantes que alteran el *splicing* coinciden y sugieren que no solo las variantes exónicas sino también las intrónicas cercanas a los límites exón-intrón deben evaluarse cuidadosamente como variantes patógenas, independientemente de la presencia de cambios de aminoácidos. Dado que, en la medicina de precisión, el marco y la lista adquirida de variantes asociadas a *splicing* constituirán recursos útiles para capturar más variantes impulsoras, incluidas aquellas previamente pasadas por alto resaltadas en cada uno de los estudios realizados.^{46, 66-72}

1.6 Vectores de expresión y minigenes

Los vectores de expresión, principalmente plásmidos, son herramientas esenciales en la medicina de precisión y genómica, especialmente en el análisis de variantes génicas ya que permiten la expresión controlada de genes específicos, proceso crucial para estudiar cómo las variantes génicas pueden afectar o no la función de genes y contribuyen a enfermedades.⁷³

Los vectores de expresión contienen elementos como sitios de origen de replicación, promotores fuertes, potenciadores, secuencias de termino y clonación, genes reporteros,

secuencias que aumentan la estabilidad del ARNm o la eficiencia de traducción como secuencias de poli A, entre otros marcadores selectivos (**Figura 9**). Es el conjunto de estos y otros componentes permiten la expresión del vector al introducirlo en una célula, obteniendo el producto de interés. Es debido a esto que son herramientas cruciales en estudios de funcionalidad génica, producción de proteínas para investigación, terapias génicas y desarrollo de vacunas.⁷⁴ Existen diferentes herramientas bioinformáticas como GeneArt, GeneOptimizer, SnapGene, o Benchling para el diseño de vectores, de acuerdo con el propósito se escoge aquel que cumple los requerimientos necesarios para cada estudio.

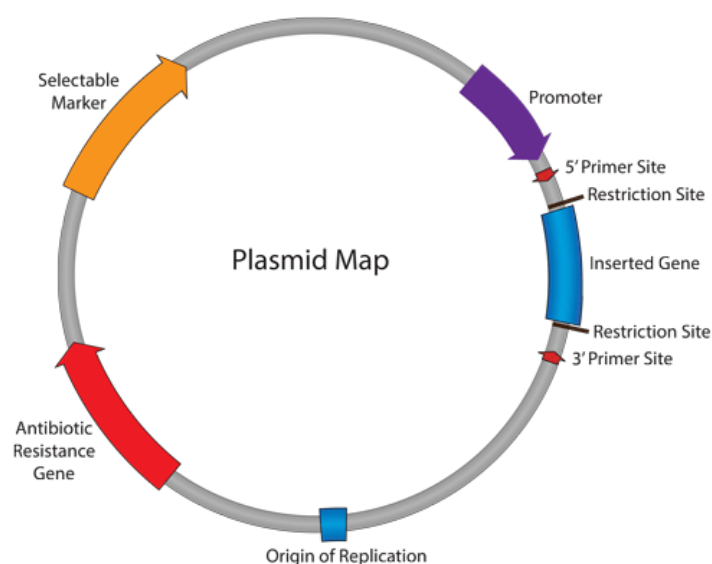


Figura 9. Principales componentes de un plásmido. Mapa de plásmido circular que muestra regiones clave como el origen de replicación (para la duplicación del plásmido), un gen de resistencia a antibióticos (para la selección de bacterias transformadas), un marcador seleccionable, un promotor (para la expresión del gen insertado) y sitios de restricción y *primers* para su manipulación genética. (Imagen tomada de <https://www.addgene.org/mol-bio-reference/>)

Por otro lado, los minigenes son versiones reducidas de genes diseñados para contener solo las secuencias necesarias para la función o expresión del gen de interés. Están diseñados para estudiar funciones específicas del gen, por ello normalmente contienen el exón que se busca estudiar junto con los intrones flanqueantes. Los minigenes se introducen en un plásmido que contiene los elementos necesarios para su expresión, como un promotor, una región de poli A y un terminador para detener la transcripción.^{75, 76}

II. Justificación

Las variantes génicas pueden surgir de *novο* debido a errores en la replicación del ADN o durante la transcripción, por lo que no es necesario que exista un antecedente familiar. Debido al impacto de las variantes génicas no solo a nivel de aminoácido, sino a su efecto en el *splicing*, es imprescindible contar con herramientas accesibles que permitan su evaluación funcional. Aunque una variante génica no cause alteraciones en un aminoácido o sea clasificada como silenciosa o de sentido erróneo puede generar cambios significativos en el *splicing*. Las variantes génicas que se encuentran en sitios no codificantes o que son clasificadas por algoritmos predictivos deben evaluarse para confirmar su efecto, ya que sin un análisis funcional no es posible determinar su patogenicidad de manera concluyente.

Aunado a esto, una variedad de estudios ha identificado y caracterizado variantes patogénicas que no se encontraban registradas en bases de datos públicas como Clinvar, LOVD o gnomAD. Es debido a esto que una herramienta que permita la evaluación de este tipo de variantes sería esencial para determinar su efecto, avanzando hacia una mejor comprensión de los mecanismos de regulación de las secuencias implicadas en el proceso de *splicing*.

Debido a todos estos factores, se destaca la importancia de tener acceso a una herramienta que permita la evaluación funcional de variantes génicas, ya que actualmente no se cuenta con plásmidos comerciales diseñados específicamente para este fin. En caso de requerir evaluar variantes para analizar su efecto en el *splicing*, sería necesario diseñar y adquirir un plásmido que cumpla con las características necesarias, lo cual tiene un alto costo. En este contexto, el presente trabajo se enfocó en el diseño de un minigen utilizando las herramientas y recursos disponibles, el cual permitirá evaluar variantes de *splicing*. Esta iniciativa responde a la necesidad identificada en el Laboratorio de Inmunogénica y Enfermedades Metabólicas del Instituto Nacional de Medicina Genómica que contaba con muestras de pacientes, pero se carecía de una herramienta adecuada que permitiera la evaluación y caracterización de las variantes asociadas a la alteración de *splicing*.

III. Objetivos

Objetivo general:

Construcción de un plásmido con un minigen que permita la evaluación funcional de variantes génicas localizadas en sitios de *splicing* y regiones circundantes.

Objetivos particulares:

1. Diseñar un minigen que contenga los elementos mínimos necesarios para llevar a cabo el proceso de *splicing*.
2. Subclonar el minigen en un vector de expresión adecuado para células de mamífero.
3. Validar la efectividad del vector de expresión para la evaluación de variantes génicas que alteran el *splicing*.
4. Evaluar si el *splicing* puede ocurrir de manera correcta sin la presencia de las secuencias completas de intrones.

IV. Hipótesis

La construcción de un plásmido que contenga un minigen permitirá la evaluación funcional de variantes génicas localizadas en sitios de *splicing*, proporcionando una herramienta eficaz para caracterizar su impacto en el proceso de *splicing*.

V. Antecedentes y marco contextual

Los minigenes clonados en vectores de expresión son herramientas clave que se han utilizado ampliamente para determinar los efectos que puede tener una variante en el *splicing*, resultando especialmente útiles debido a la amplia cantidad de variantes génicas que no se encuentran reportadas como patogénicas, que tienen significado incierto o con sitios crípticos no reportados.^{60, 66-73,75,76,77}

Investigaciones recientes sobre la determinación de variantes que alteran el *splicing* han utilizado minigenes y vectores de expresión para evaluar el efecto que tienen en el ARN y por consiguiente en la proteína. Se han descrito condiciones causadas por variantes asociadas a *splicing* previamente no reportadas, como aniridia, por Tamayo, et al., (2023)⁷⁵, fibrosis quística, por Sharma, et al., (2014)⁷⁸, el complejo de esclerosis tuberosa, por Liu, et al., (2024)⁷⁹, hipomagnesemia primaria con hipocalcemia secundaria, por Song et al., (2022)⁸⁰, solo por mencionar algunos casos. Cada uno de estos trabajos se caracteriza por emplear un método diferente en la construcción del minigen.

El minigen presente en este trabajo se diseñó originalmente para evaluar variantes en el final del intrón 9 y el inicio del exón 10 del gen *HNRNPK*, el cual de acuerdo con el *National Center for Biotechnology Information (NCBI)* codifica para la ribonucleoproteína nuclear heterogénea K, proteína perteneciente a la familia de las proteínas hnRNP reguladoras del *splicing*. Es una proteína de unión a pre-ARNm que ayuda a su procesamiento, así como otros aspectos del metabolismo y transporte del ARNm. La alteración en un regulador clave como *HNRNPK* puede tener efectos de cascada sobre varios genes, afectando procesos biológicos y contribuyendo a la aparición de síndromes y enfermedades de manera indirecta. Un ejemplo de lo que ocurre cuando hay desregulación del gen *HNRNPK* es el síndrome de Au Kline. De acuerdo con la base de datos del NCBI sobre la información del gen, este síndrome tiene un tipo de herencia autosómica dominante causado típicamente por variantes de *novus* (<https://www.ncbi.nlm.nih.gov/books/NBK540283/>). Bases de datos como Clinvar o GnomAD reportan una gran cantidad de variantes génicas en el gen *HNRNPK* que permanecen sin caracterización y son de significado incierto. La construcción aquí realizada permitirá evaluar variantes asociadas al gen *HNRNPK*, sin embargo, su aplicación no se limita exclusivamente a este gen, ya que también podría utilizarse para evaluar variantes de *splicing* asociadas a otros genes.

VI. Material y métodos

El presente trabajo consistió en el diseño de un minigen el cual fue sintetizado en el plásmido pUCDIT-Amp y posteriormente fue subclonado dentro del vector pcDNA3.1/myc-His C para su expresión en células HEK293T. A través de RT-PCR y secuenciación Sanger se comprobó que el *splicing* ocurrió de manera correcta, con lo que se da la posibilidad de introducir variantes génicas en sitios de *splicing* y regiones circundantes con el fin de evaluar si alteran el *splicing*.

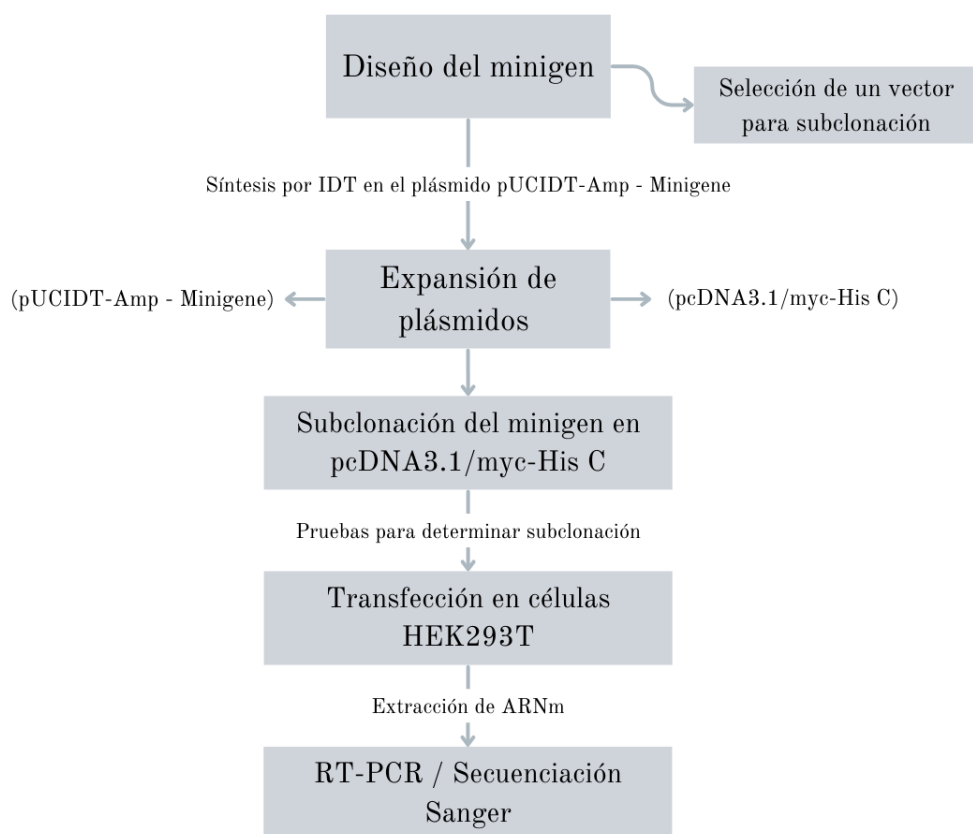


Figura 10. Diagrama de trabajo. El diagrama muestra los puntos básicos del trabajo experimental realizado.

6.1 Diseño del minigen

El minigen fue diseñado para analizar el impacto de variantes que alteran el *splicing*. Para ello se seleccionaron fragmentos de las secuencias correspondientes a tres exones y dos intrones del gen *HNRNPK* (**Figura 11.A**). Se construyó tomando el fragmento final del exón 9, el exón 10 completo y un fragmento inicial del exón 11. Flanqueante al exón 10 se colocó la secuencia inicial y final del intrón 9 y del intrón 10 respectivamente, añadiendo a cada intrón una secuencia de corte EcoR1. Por otro lado, se añadieron secuencias de corte para BamH1 al inicio del minigen y Not1 al final de la construcción (**Figura 11.B**).

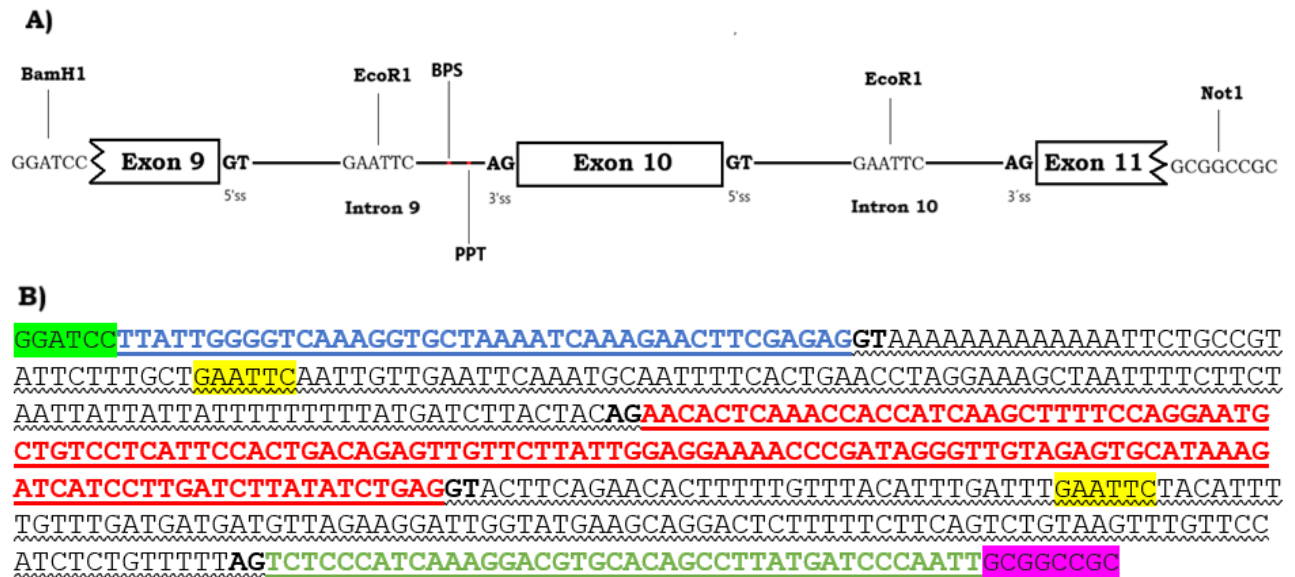


Figura 11. Estructura minigen. **A)** Construcción del minigen con tres exones, dos intrones, sitios de splicing y secuencias de corte para las enzimas marcadas. **B)** Secuencia del minigen construido; en azul se muestra el exón 9, en rojo el exón 10 y en verde el exón 11; en negro las secuencias correspondientes a los intrones 9 y 10. En negritas se resaltan los sitios de splicing 5' y 3' de cada intrón.

Una vez diseñado el minigen (**Figura 11.B**), este fue enviado para su síntesis por la empresa *Integrated DNA Technologies (IDT)*. La secuencia del minigen fue clonada en el vector pUCIDT-Amp (**Figura 12**), que contiene un marcador de resistencia a ampicilina, el cual sirvió para la selección de colonias transformadas correctamente.

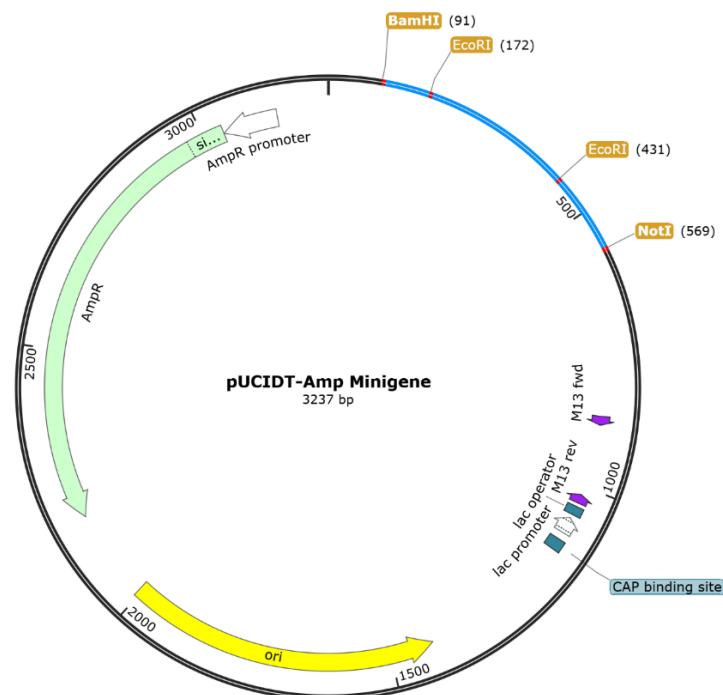


Figura 12. Vector pUCIDT-Amp con minigen. El vector contiene la secuencia de 485 pb del minigen construido representada en azul. Dentro del minigen, en café se encuentran las secuencias de corte para las enzimas de restricción BamHI y NotI para la liberación del inserto. Gen de resistencia a ampicilina (AmpR) junto con su promotor (AmpR promoter). Un sitio de origen de replicación (ori) para su replicación en bacterias. (Imagen diseñada en SnapGene a partir de la secuencia).

6.2 Vector para subclonación

Se seleccionó el vector de expresión **pcDNA3.1 myc / His C** dado que contaba con los componentes necesarios para transfección en células de mamífero y análisis de expresión de genes. Este vector se encontraba disponible en el laboratorio, y para fines de este proyecto fue necesario remover un fragmento (el inserto) que había sido diseñado para otra investigación. En la **Figura 13** se muestra los componentes del vector.

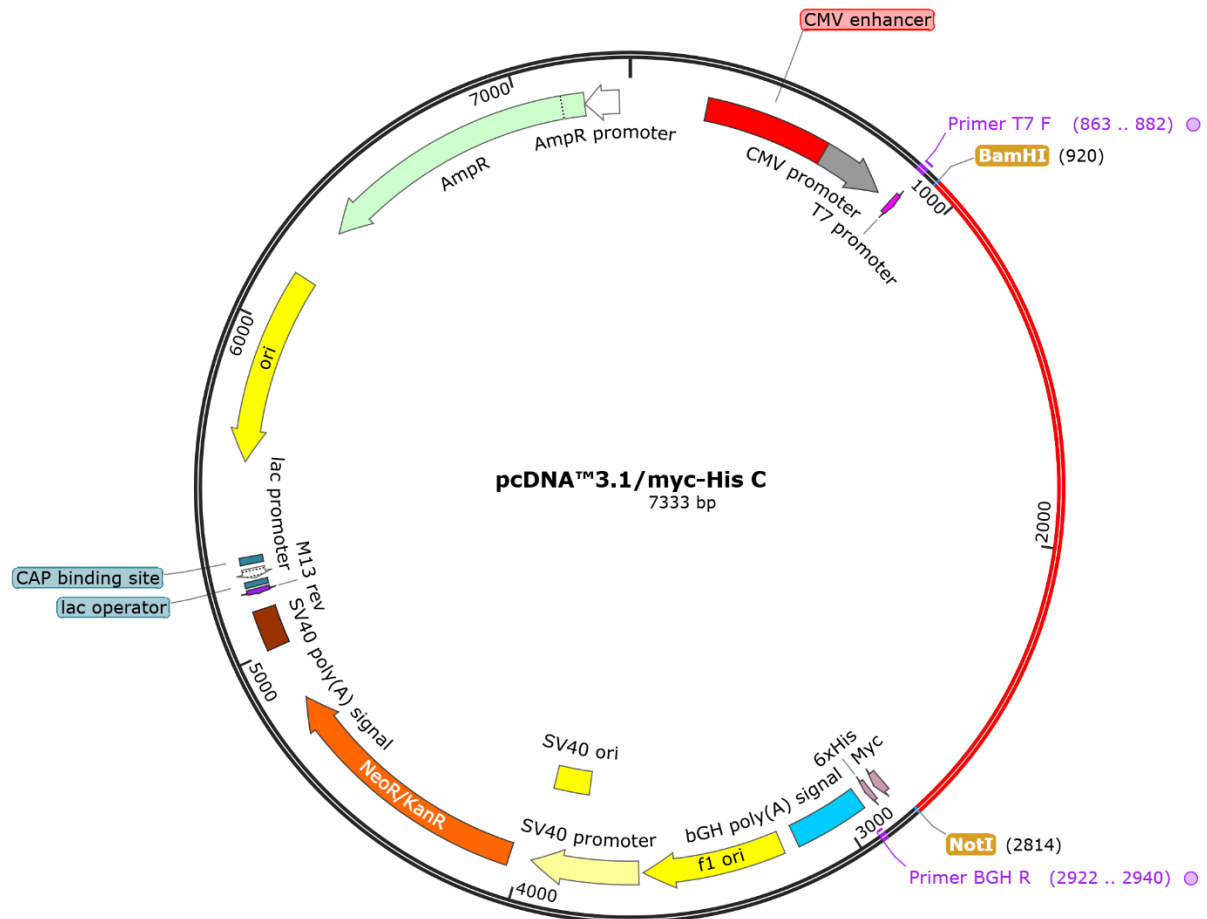


Figura 13. Vector pcDNA3.1/myc-His C. El vector contiene una secuencia de ~1901 pb que se removió (al no ser de interés para esta investigación) representada en rojo. Flanqueante a la secuencia roja se localizan secuencias de corte para las enzimas de restricción BamH1 y Not1 en café. En morado, Río arriba de BamH1 y río abajo de Not1, se localiza una secuencia de unión para el primer T7 Forward y BGH Reverse respectivamente. Gen de resistencia a ampicilina (AmpR) con su promotor (AmpR promoter), proporciona resistencia a la ampicilina permitiendo la selección de bacterias que contienen el vector en presencia de este antibiótico. El promotor (CMV promoter) y potenciador del citomegalovirus (CMV enhancer), altamente eficientes para inducir una expresión génica fuerte en células de mamíferos. His 6x y Myc, secuencia de etiquetas que permite la detección y purificación de proteínas fusionadas usando anticuerpos anti-Myc y la purificación mediante cromatografía de afinidad por níquel (His-tag). Señal de poliadenilación bovina (bGH poly (A) signal), utilizada para asegurar la terminación y estabilidad del ARNm en células eucariotas. Promotor del virus SV40 (SV40 promoter) que dirige la expresión de genes en células de mamíferos. La señal de poliadenilación SV40 (SV40 poli (A) signal) asegura la correcta maduración del ARNm transcrito. Gen de resistencia a la neomicina/kanamicina (NeoR/KanR), que permite la selección de células eucariotas transformadas en presencia de G418. Origen de replicación (ori) para la propagación del vector en células bacterianas. (Imagen diseñada en SnapGene a partir de las secuencias).

6.3 Expansión de plásmidos

6.3.1 Transformación bacteriana

Con células competentes de *Escherichia coli* DH5 α se realizó la transformación bacteriana utilizando el plásmido pcDNA3.1/myc-His C. Se añadieron 2 ng del plásmido pcDNA3.1/myc-His C a 50 μ L de células competentes contenidas en tubos Eppendorf de 1.5 mL y se incubó la mezcla en hielo durante 30 minutos. Posteriormente, se realizó un choque térmico a 42°C durante 60 segundos, seguido de una incubación en hielo por 2 minutos. A continuación, se adicionaron 1000 μ L de medio LB sin antibiótico y se incubó a 37°C durante 1 hora con agitación a 225 revoluciones por minuto (rpm). Después de la incubación, se centrifugó la muestra a 5000 rpm durante 3 minutos para generar un botón celular, se retiraron 400 μ L del sobrenadante para concentrar el volumen restante. Finalmente, las células transformadas se sembraron en placas de agar LB suplementadas con 100 μ g/mL de ampicilina y se incubaron a 37°C durante 24 horas.

6.3.2 Selección de colonias y crecimiento en cultivo líquido

Después de la incubación, se seleccionaron colonias individuales de las placas de agar LB que habían sido sembradas con las células transformadas con los plásmidos pcDNA3.1/myc-His C y pUCIDT-amp. Cada colonia seleccionada se inoculó en 15 mL de medio LB líquido suplementado con 1 μ L de ampicilina por cada mL de medio (15 μ L en total) en tubos de 15 mL. Los cultivos se incubaron a 37°C durante 24 horas con agitación a 150 rpm para permitir el crecimiento bacteriano y la amplificación del plásmido.

6.3.3 Extracción de ácidos nucleicos - Miniprep

Después del cultivo líquido, se procedió a la extracción y purificación de ácidos nucleicos utilizando el kit y protocolo "Plasmid Mini-Prep Kit - Column Kit" de Jena Bioscience. Los cultivos bacterianos de 15 mL se centrifugaron a 3000 rpm durante 10 minutos para recolectar las células en un botón. Se retiró el sobrenadante, dejando solo el botón celular y se añadieron 300 μ L de buffer de lisis a cada uno de los tubos. Se resuspendieron las células con micropipeta. Posteriormente, se les añadió 300 μ L de buffer de neutralización, y la muestra se mezcló por inversión de 4 a 6 veces. Cada muestra se traspasó individualmente a un tubo de 2 mL y se centrifugó a 13,000 rpm durante 5 minutos.

Antes de transferir el sobrenadante, se activaron dos columnas de miniprep con 100 μ L de buffer de activación y se centrifugaron durante 1 minuto. El sobrenadante, que contenía el ADN plasmídico, se transfirió a las columnas de miniprep colocadas en un tubo recolector de 2 mL. Las columnas se centrifugaron a 13,000 rpm durante 1 minuto y el eluido se descartó. Posteriormente, se realizaron dos lavados de 1 minuto con 700 μ L de washing buffer con etanol. Después de los lavados, las membranas se secaron por centrifugación a 13,000 rpm durante 2 minutos.

Finalmente, se reemplazaron los tubos colectores de la columna por tubos Eppendorf nuevos de 1.5 mL. El ADN se eluyó en 30 μ L de agua libre de nucleasas y se centrifugó durante 2 minutos para completar el proceso de purificación. El ADN purificado se cuantificó utilizando un espectrofotómetro NanoDrop ThermoFisher Scientific. La concentración de ADN para el plásmido pUCIDT-amp fue de 29.7 ng/ μ L, mientras que para el plásmido pcDNA3.1/myc-His C fue de 128.7 ng/ μ L. El ADN purificado se almacenó a -20°C para su posterior análisis.

6.4 Subclonación en pcDNA3.1 myc / His C

6.4.1 Restricción BamHI y NotI

Para el plásmido pUCIDT-amp, se preparó una mezcla de 25 μ L que contenía 2.5 μ L de NE Buffer r3 10x, 1 μ L de NotI (New England Biolabs), 1 μ L de BamHI (New England Biolabs), 20 μ L de ADN a una concentración de 243 ng/ μ L y 0.5 μ L de agua libre de nucleasas. Para el plásmido pcDNA3.1/myc-His C, se preparó en un volumen de 15 μ L con 1.5 μ L de buffer 3 10x, 1 μ L de NotI, 1 μ L de BamHI, 3.85 μ L de ADN a una concentración de 1300 ng/ μ L y 7.65 μ L de agua libre de nucleasas. Ambas muestras se incubaron a 37°C durante 65 horas.

6.4.2 Electroforesis en gel de agarosa

Se realizó una electroforesis en gel de agarosa para verificar los productos de la restricción. Para pUCIDT-amp se utilizó agarosa al 1%, y se corrió el gel durante 100 minutos a 90 V. Para pcDNA3.1/myc-His C, un gel de agarosa al 0.6%, el cual se corrió durante 180 minutos a 100 V, manteniendo la cámara de electroforesis sobre hielo para evitar sobrecalentamiento. Las bandas de interés, correspondientes al inserto de 485 pb de pUCIDT-amp y al vector de 5000 pb de pcDNA3.1/myc-His C, fueron cortadas del gel para su purificación.

6.4.3 Extracción y purificación de ácidos nucleicos de gel de agarosa

Seguido a la electroforesis, se procedió a la purificación de las bandas correspondientes al inserto y al vector utilizando el kit “MinElute Gel Extraction Kit de Qiagen”. Las bandas de ADN fueron cortadas cuidadosamente del gel de agarosa y colocadas cada una en un tubo Eppendorf de 1.5 mL. Se añadieron 600 µL de buffer QG a su respectivo tubo con su banda (en una proporción de 3 volúmenes de buffer por cada volumen de gel). Las muestras se incubaron a 50°C durante 10 minutos, agitando ocasionalmente para asegurar la disolución completa del gel. Disuelto el gel, se añadió 1 volumen de isopropanol a cada mezcla para favorecer la unión del ADN a la membrana de la columna MinElute. Tras aplicar la mezcla a las columnas y realizar una breve centrifugación, el eluido fue descartado. Las columnas se lavaron con 500 µL de buffer QG con una centrifugación de 1 minuto. Posteriormente, con 750 µL de buffer PE que reposo por 3 minutos seguido de una centrifugación de 1 minuto. Se repitió el lavado con el buffer PE con el fin de eliminar cualquier contaminante residual. Finalmente, las columnas se secaron 2 minutos por centrifugación y se eluyó el ADN en 30 µL de agua destilada.

6.4.5 Desfosforilación del vector

Para prevenir la recircularización del vector se utilizó la enzima *Shrimp Alkaline Phosphatase* (SAP). Se agregó 1 µL de SAP al volumen total del vector purificado y se incubó a 37°C durante 30 minutos. Una vez completada la reacción, la fosfatasa fue inactivada mediante incubación a 65°C por 10 minutos.

6.5.6 Purificación de ácidos nucleicos

Se llevó a cabo una purificación del ADN para eliminar restos de enzimas y cualquier contaminante que pudiera interferir en la ligación. Se utilizó el kit QIAquick PCR Purification Kit (250) de Qiagen. Se mezclaron 5 volúmenes del buffer PB proporcionado por el kit con 1 volumen de la muestra de PCR. Esta mezcla se transfirió a una columna QIAquick colocada en un tubo recolector. La muestra se centrifugó a 13,000 rpm durante 1 minuto, lo que permitió que los fragmentos de ADN se unieran a la membrana de la columna. Posteriormente, el flujo a través del tubo fue descartado y la columna se lavó añadiendo 750 µL del buffer PE, que contiene etanol, para eliminar impurezas. Se realizó una centrifugación adicional a 13,000 rpm durante 1 minuto. Para asegurar la eliminación completa del buffer de lavado, la columna se centrifugó nuevamente durante 1 minuto más a la misma velocidad. Luego, la columna QIAquick fue transferida a un tubo Eppendorf nuevo de 1.5 mL.

El ADN purificado se eluyó añadiendo 30 μ L de agua libre de nucleasas directamente a la membrana de la columna. Se dejó incubar a temperatura ambiente durante 1 minuto, seguido de una centrifugación a 13,000 rpm durante 2 minutos. Se regresó el centrifugado a la columna y se realizó una segunda centrifugación a las mismas condiciones para optimizar la concentración de ADN. Al finalizar el proceso de purificación se cuantificó el ADN.

6.4.7 Ligación enzimática de fragmentos de ADN

Se utilizó una relación molar de 3:1 inserto a vector, la cual fue calculada con la herramienta NEBcalculator. Se utilizaron 7 μ L del vector a una concentración de 128.7 ng/ μ L, 8 μ L del inserto a una concentración de 29.7 ng/ μ L, 1 μ L de ATP 10 mM, 1 μ L de la enzima T4 DNA ligasa y 4.2 μ L de buffer 5x fast ligation, obteniendo un volumen final de 21.2 μ L. La reacción se incubó a 16°C durante 40 minutos.

6.4.8 Transformación bacteriana con el fragmento ligado

Utilizando el inserto ligado al vector pcDNA3.1/myc-His C se transformaron células competentes de *E. coli* DH5 α siguiendo el protocolo previamente descrito. Se utilizaron 5 μ L de la mezcla de ligación en 50 μ L de células competentes en un tubo Eppendorf de 1.5 mL. Las células transformadas se sembraron en placas de agar LB suplementadas con 100 μ g/mL de ampicilina y se incubaron a 37°C durante 24 horas.

6.4.9 Selección y clonación de colonias en medio LB

Después de la transformación bacteriana se seleccionaron las colonias que mostraron mejor crecimiento. Para facilitar esta selección, previamente se trazaron cuadrículas de 4 x 4 en las placas Petri, dividiendo el área de siembra en 16 secciones, lo que permitió un seguimiento ordenado de las colonias. El proceso de selección e inoculación se llevó a cabo bajo condiciones estériles, utilizando mecheros de Bunsen para mantener un entorno libre de contaminantes. Se seleccionaron ocho colonias que se tomaron con puntas estériles y se inocularon en tubos de 15 mL con 10 mL de medio líquido LB, suplementado con 10 μ L de ampicilina (1 μ L de ampicilina por cada mL de medio). Los cultivos fueron incubados durante 24 horas a 37°C con agitación constante a 150 rpm, permitiendo el crecimiento óptimo de las bacterias para el posterior aislamiento de plásmidos. Después del cultivo bacteriano, se procedió a la extracción de ácidos nucleicos de los medios de cultivo utilizando el Plasmid Mini-Prep Kit - Column Kit de Jena Bioscience como anteriormente se hizo.

6.5 Pruebas para determinar la presencia del inserto en el plásmido

6.5.1 Diseño de *primers*

Se diseñaron dos *primers* flanqueantes a las regiones BamH1 y Not1 en el vector pcDNA3.1/myc.His C. Río arriba del sitio BamH1 se diseñó el *primer* T7 *forward* y río abajo de Not1 se diseñó el *primer* BGH *reverse* (**Figura 13**). Además, los *primers* están contenidos dentro de la región que será transcrita por parte del promotor CMV, lo que permite que el ADNc generado a partir de un ARNm pueda ser amplificado utilizando estos mismos *primers*.

Primer T7 Forward:

5' GAATTTTGTAAATACGACTCACATATAGGG 3'

Primer BGH Reverse:

5' CTAGAAGGCACAGTCGAGG 3'

Figura 14. *Primers* para pruebas. *Primers* diseñados para amplificación de la región del inserto en el vector de expresión pcDNA3.1/myc.His C.

6.5.2 Restricción BamH1 y Not1

Con las enzimas NotI y BamHI se restringieron los plásmidos extraídos de cada una de las colonias para confirmar la presencia del inserto en las construcciones plasmídicas. La reacción se realizó en un volumen de 15 µL que contenía 1.5 µL de buffer 3 10x, 1 µL de NotI, 1 µL de BamHI, 8 µL de ADN plasmídico y 3.5 µL de agua destilada para cada una de las 8 muestras. Se incubaron a 37°C por 24 h.

6.5.3 PCR

Utilizando los *primers* los *primers* T7 *forward* y BGH *reverse* se realizó una PCR a cada uno de los plásmidos extraídos. La reacción se realizó en un volumen de 15 µL que contenía 1.5 µL de buffer Ruby (Jena Bioscience), 0.4 µL de dNTPs, 0.4 µL de *primer* T7 F, 0.4 µL de *primer* BGH R, 1 µL de ADN, 11.2 µL de agua libre de nucleasas y 0.1 µL de taq polimerasa (Jena Bioscience). La reacción se llevó a 94°C durante 2 minutos, seguido de 30 ciclos de desnaturalización a 94°C durante 15 segundos, alineación a 59°C durante 20 segundos y extensión a 72°C durante 45 segundos. Una vez completados los 30 ciclos, una extensión final a 72°C durante 2 minutos. Para concluir, las muestras se mantuvieron a 4°C durante 5 minutos.

6.5.4 Secuenciación Sanger

Una vez que las dos pruebas iniciales confirmaron la correcta subclonación, se procedió a realizar una secuenciación Sanger en dos de las seis muestras de ADN obtenidas de las colonias seleccionadas. Se utilizaron 100 ng de ADN en un volumen final de 15 µL. A esta cantidad de

ADN se añadieron 0.32 μ L de *primer forward* y el volumen restante se completó con agua libre de nucleasas para alcanzar el volumen total de 15 μ L.

La mezcla preparada se envió al *Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán* a través de la *Red de Apoyo a la Investigación (RAi)*. Los datos obtenidos fueron analizados utilizando el software FinchTV para interpretar las secuencias. Posteriormente, las secuencias obtenidas se alinearon con la secuencia del minigen.

6.5.6 Cultivo líquido

Se seleccionó la clona 4 con la cual se preparó un cultivo bacteriano de 500 mL de medio LB en un matraz Erlenmeyer de 1 L, el cual fue sellado con papel aluminio. Posteriormente, el medio se esterilizó por autoclave a 121°C durante 30 minutos. Una vez que el medio se enfrió, se colocó en un ambiente estéril utilizando mecheros para mantener la asepsia. En este ambiente, se añadieron 500 μ L de ampicilina (100 mg/mL), a continuación, el medio se inoculó con una muestra de la clona 4. El cultivo se incubó durante 24h a 37°C con agitación constante a 150 rpm.

6.5.7 Extracción de ácidos nucleicos – Maxiprep

Para la extracción ácidos nucleicos se utilizó el kit “Qiagen Plasmid Maxi Kit (25)”. Después de 24 horas de incubación, las células bacterianas fueron recolectadas por centrifugación a 3500 rpm durante 10 minutos a 4°C. Se descartó el sobrenadante y se dejó secar el tubo por 5 minutos. El pellet bacteriano se resuspendió en 10 mL de buffer P1 (que contenía RNasas) para degradar el ARN presente.

Se realizó la lisis celular añadiendo 10 mL de buffer P2 (solución alcalina con SDS), inmediatamente se mezcló suavemente por inversión para asegurar la completa ruptura de las membranas celulares. La solución se dejó incubar durante 5 minutos a temperatura ambiente. A continuación, se neutralizó la solución añadiendo 10 mL de buffer P3, seguido de centrifugación a 5000 rpm durante 20 minutos a 4°C. Durante la centrifugación se montó una columna provista por el kit sobre un matraz y se equilibró con 10 mL de buffer QBT. Al finalizar la centrifugación el sobrenadante se recuperó utilizando una jeringa de 10 mL y se pasó por un filtro de jeringa colocado sobre la columna. Una vez que todo el sobrenadante pasó por la columna se retiró el filtro y se realizaron dos lavados con buffer QC, donde se agregaron 30 mL sobre la columna, una vez filtrado, se agregaron 30 mL más.

Después de los lavados, se colocó la columna sobre un tubo nuevo de 50 mL, donde se agregaron 15 mL de buffer QF sobre la columna. Al terminar el filtrado, se retiró la columna y se agregaron 10.5 mL de isopropanol, se tapó el tubo y se mezcló por inversión 4 veces. Una vez homogeneizada la muestra, se centrifugo a 5000 rpm durante 30 minutos.

Finalizada la centrifugación, el sobrenadante fue descartado sin tocar el botón de ADN y se añadieron 5 mL de etanol al 70%, se resuspendió la muestra por pipeteo, seguido de una centrifugación a 5000 rpm durante 15 minutos.

Por último, se retiró el sobrenadante cuidadosamente sin tocar el fondo del tubo y se dejó secar de 5 a 10 minutos. Una vez seco, se agregaron 200 μ L de agua libre de nucleasas para resuspender el ADN. Se cuantificó y la concentración de plásmido fue de 1400 ng/ μ L.

6.6 Transfección en células HEK293T

6.6.1 Cultivo celular

Las células HEK293T se mantuvieron en una incubadora a 37°C en un ambiente húmedo con 5% de CO₂. El medio utilizado fue **Advance-MEM** suplementado con piruvato de sodio al 1% v/v (100 mM) y un antifúngico antimicótico 100x. Para la preservación de las líneas celulares. Las células fueron tripsinizadas, luego centrifugadas y lavadas con PBS. Posteriormente, se realizó una segunda centrifugación, y las células se resuspendieron en medio **Advance-MEM** suplementado con suero fetal bovino al 5% v/v y DMSO al 7% v/v para su correcta conservación en congelación.

6.6.2 Transfección

Las células HEK293 fueron sembradas en placas p60 y se mantuvieron en medio Advance-MEM suplementado por 24 horas antes de realizar la transfección. En cada placa, se transfectaron 4 μ g de plásmido utilizando lipofectamina 2000 (Invitrogen), con una proporción de 1.75 μ L de lipofectamina por cada 1 μ g de ADN. Se transfectaron 4 μ g de plásmido, para ello, se prepararon dos diluciones separadas: una de ADN y otra de lipofectamina en Opti-MEM, con un volumen final de 1500 μ L por placa. Las diluciones de ADN y lipofectamina se incubaron por separado durante 5 minutos a temperatura ambiente, y luego se mezclaron, permitiendo que se incubaran juntas durante 25 minutos para la formación de liposomas. Se añadieron 500 μ L de la mezcla a cada placa y las células se incubaron durante 4 horas.

Finalmente, el medio fue reemplazado por Advance-MEM suplementado con suero fetal bovino al 5% para continuar con el cultivo. Este procedimiento garantizó una eficiente transfección y la correcta preservación de las células para los análisis posteriores.

6.6.3 Extracción de ARN con trizol

Para la extracción de ácidos nucleicos de las células HEK293T transfectadas, se utilizó el kit TRIzol Reagent de Invitrogen siguiendo el protocolo del fabricante. Las células se recolectaron y resuspendieron en 1 mL de trizol pipeteando hasta disolver las muestras completamente. Una vez homogeneizada la muestra se incubó a temperatura ambiente durante 5 minutos para permitir la disociación de los complejos de ácidos nucleicos y proteínas.

Posteriormente, se añadieron 200 μ L de cloroformo por cada mL de Trizol y las muestras se agitaron vigorosamente durante 15 segundos. Después, se incubaron a temperatura ambiente durante 2-3 minutos. Las muestras se centrifugaron a 13,000 rpm, durante 15 minutos a 4°C, permitiendo la separación de las fases.

La fase acuosa, que contiene el ARN, fue transferida a un tubo limpio. Se añadieron 500 μ L de isopropanol por cada mL de trizol para precipitar el ARN, seguido de una incubación a temperatura ambiente durante 10 minutos. Se realizó una centrifugación a 13,000 rpm durante 10 minutos a 4°C para recuperar el ARN precipitado.

El pellet resultante se lavó con 1 mL de etanol al 75% v/v, seguido de una centrifugación a 10,000 rpm durante 5 minutos a 4°C. Después del lavado, el pellet de ARN se secó brevemente y se resuspendió en 30 μ L de agua libre de nucleasas para su almacenamiento a -80°C hasta su uso en futuros análisis.

6.7 RT-PCR

Para la síntesis de ADNc se utilizó el protocolo de transcripción inversa "Reverse Transcription protocol for SuperScript IV VILO Master Mix without ezDNase enzyme treatment Invitrogen". La reacción se llevó a cabo en un volumen de 20 μ L que contenía 4 μ L de SuperScript IV VILO Master Mix, 1.3 μ L de ARN a una concentración de 1550 ng/ μ L y 14.7 μ L de agua libre de nucleasas. La mezcla se incubó a 25°C durante 10 minutos, seguido de 50°C por 10 minutos y finalmente a 85°C por 5 minutos.

Seguido de esto se realizó una PCR utilizando 1.5 μ L de buffer Ruby, 0.4 μ L de dNTPs (10 mM), 0.4 μ L de *primer* T7 F, 0.4 μ L de *primer* BGH R y 1.5 μ L del ADNc, 0.1 μ L de Taq polimerasa y 10.7 μ L de agua libre de nucleasas para un volumen final de 15 μ L. La reacción se llevó a las mismas condiciones de temperatura que en la sección 6.5.3. El producto de PCR se visualizó mediante una electroforesis en gel de agarosa al 1% aplicando un voltaje de 80 V durante 50 minutos.

6.8 Purificación de producto PCR

Se utilizó el kit “QIAquick PCR Purification Kit (250) de Qiagen” siguiendo el protocolo del fabricante. Inicialmente se mezclaron 5 volúmenes de buffer PB proporcionado por el kit con 1 volumen de la muestra de PCR. Esta mezcla se transfirió a una columna QIAquick colocada en un tubo recolector. La muestra se centrifugó a 13,000 rpm durante 1 minuto. Posteriormente, el flujo a través del tubo fue descartado y la columna se lavó añadiendo 750 μ L del buffer PE. Se realizó una centrifugación adicional a 13,000 rpm durante 1 minuto. La columna se centrifugó nuevamente durante 1 minuto a la misma velocidad. Luego, la columna QIAquick fue transferida a un tubo nuevo de 1.5 mL. El ADN purificado se eluyó añadiendo 30 μ L de agua libre de nucleasas directamente a la membrana de la columna. Se dejó incubar a temperatura ambiente durante 1 minuto, seguido de una centrifugación a 13,000 rpm durante 2 minutos. Para optimizar la concentración de ADN, se regresó el centrifugado a la columna y se realizó una segunda centrifugación a las mismas condiciones.

Finalmente, se preparó la muestra para secuenciación Sanger con 100 ng de ADN en un volumen total de 15 μ L. A esta cantidad de ADN se añadieron 0.32 μ L de *primer* T7 F y el volumen restante se completó con agua libre de nucleasas para alcanzar el volumen total de 15 μ L.

Las secuencias obtenidas se compararon utilizando la herramienta *BLAST* del *NCBI*. Este análisis permitió identificar las secuencias y verificar la correspondencia con las secuencias obtenidas con la base de datos.

VII. Resultados

El minigen diseñado se encontraba clonado en el plásmido pUCIDT-Amp el cual contenía los sitios BamH1 y NotI. Así mismo, el plásmido pcDNA3.1/myc-His C contenía un fragmento de ~2000pb que no era de interés, sin embargo, el plásmido cumplía con los requerimientos necesarios para la expresión del minigen. Por lo tanto, el *primer* pasó consistió en realizar una restricción enzimática para liberar tanto el inserto (minigen) como el vector de sus respectivos plásmidos.

Restricción de plásmidos pUCIDT-AMP y pcDNA3.1/myc-His C

Los productos de la restricción enzimática (sección 6.4.1 de materiales y métodos) fueron analizados mediante electroforesis en gel de agarosa. Las bandas observadas fueron consistentes con los tamaños esperados. Para el plásmido pUCIDT-AMP se observó una banda de aproximadamente 500 pb correspondiente al inserto (**Figura 15.A**). En el caso del plásmido pcDNA3.1/myc-His C se observó una banda de 5000 pb correspondiente al vector (**Figura 15.B**). Estos resultados confirmaron la correcta restricción de ambos plásmidos. Posteriormente, las bandas fueron cortadas, purificadas y utilizadas para la reacción de ligación para generar la construcción plasmídica.

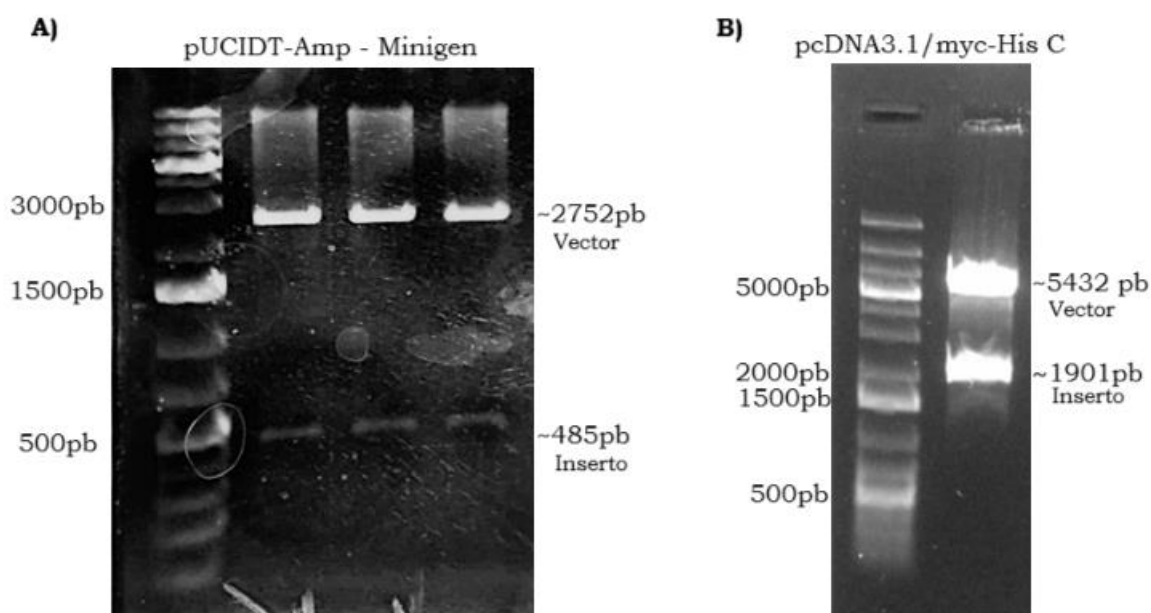


Figura 15. Restricción de plásmidos. A) pUCIDT-AMP restringido con NotI y BamHI, en ~500 pb se observa la banda correspondiente al minigen. B) pcDNA3.1/myc-His C restringido, en ~5000 pb se observa la banda correspondiente al vector liberado de su inserto.

Los fragmentos purificados correspondientes al inserto y al vector se corrieron en un gel de agarosa (**Figura 16**), lo que permitió confirmar la presencia exclusiva de los elementos deseados. Este análisis validó que los fragmentos eran viables para realizar la reacción de ligación.

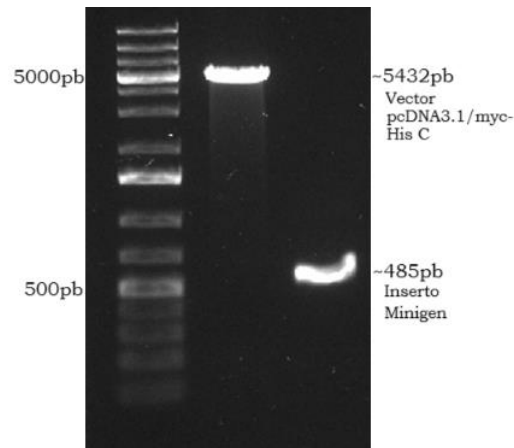
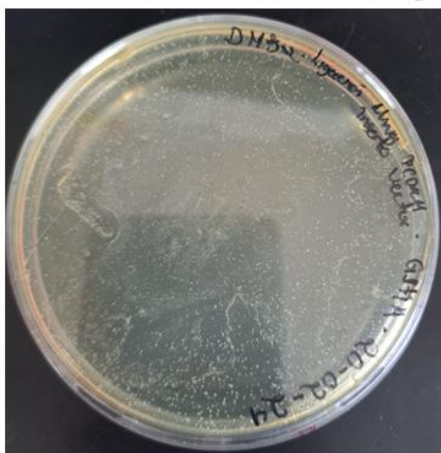


Figura 16. Vector e inserto purificados. En el segundo pozo se observa la banda correspondiente al vector liberado del inserto. En el tercer pozo la banda correspondiente al inserto liberado. Del lado izquierdo se muestran los pesos correspondientes y del lado derecho se muestran los pesos teóricos correspondientes a cada banda.

Transformación bacteriana con plásmido ligado

Tras la reacción de ligación, se transformaron células competentes (sección 6.4.9) *E. Coli* DH5 α y se cultivaron en placas de agar LB suplementadas con ampicilina. El cultivo resultó en el crecimiento de múltiples colonias bacterianas (**Figura 17**). Estas colonias se aislaron en una nueva placa dividida en 16 secciones facilitando su identificación. El proceso se realizó para garantizar un manejo ordenado y la caracterización precisa de los plásmidos.

A)
Células DH5 α transformadas con ligación



B)
Células DH5 α aisladas

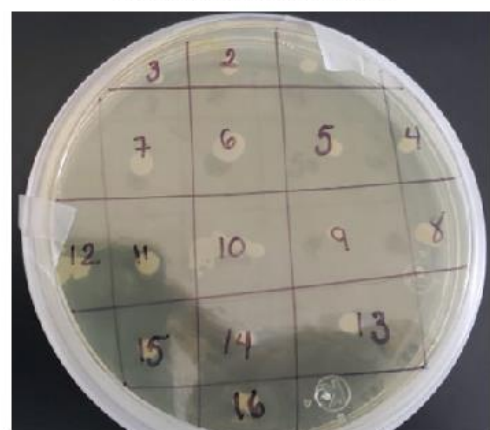


Figura 17. Placas de agar con células DH5 α transformadas. A) Colonias bacterianas obtenidas tras la transformación con la reacción de ligación. B) Colonias aisladas para la caracterizar el plásmido.

Cultivo y extracción de plásmidos

Tras el aislamiento de las colonias transformadas, se seleccionaron las colonias 1, 2, 3, 4, 5, 6, 7 y 8 con las que se realizaron cultivos líquidos miniprep y posteriormente se realizó la extracción de los ácidos nucleicos.

Validación del vector construido

Para confirmar la correcta construcción del vector se realizó la digestión enzimática (sección 6.5) con BamHI y NotI. Los resultados mostraron una separación correcta de los fragmentos en los tamaños esperados (**Figura 18. A y B**). Simultáneamente, se realizó una PCR con los *primers* T7 F y BGH R para amplificar la región del inserto de cada muestra (**Figura 18. A y C**). Ambas pruebas se realizaron en cada una de las muestras de las ocho colonias seleccionadas.

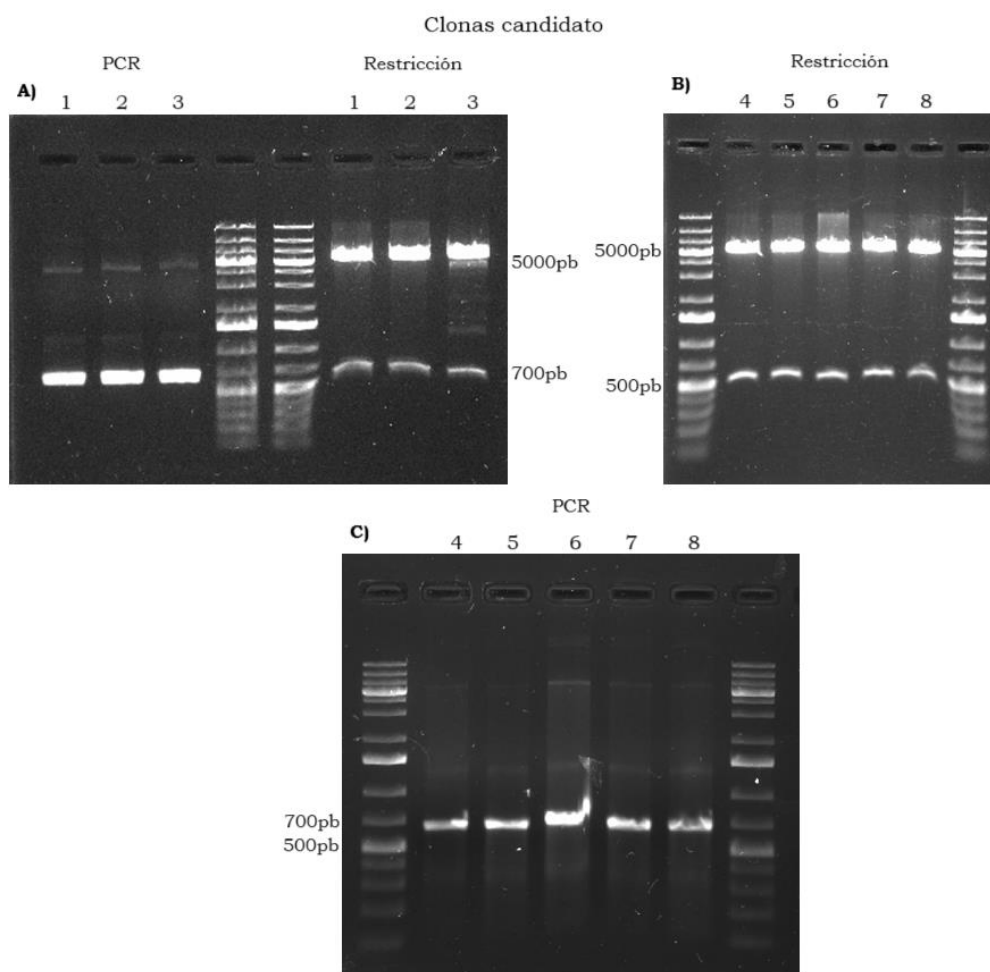


Figura 18. Restricción enzimática y PCR de las clonas candidato. A) Clona 1 – 3, a la izquierda se muestra la amplificación por PCR del inserto con los *primers* T7 F y BGH R en un peso de ~700 pb. A la derecha la restricción enzimática del vector construido con las bandas esperadas. B) Restricción enzimática de las clonas 4 – 8. C) PCR del inserto por los *primers* T7 F y BGH R de la clona 4 – 8. En cada imagen los pesos mostrados son los correspondientes al marcador de peso molecular.

Dado que cada una de las ocho clonas seleccionadas mostró la incorporación exitosa del inserto, se seleccionaron dos de ellas; la clona 4 y 5, las cuales fueron secuenciadas, con lo que se comprobó que la muestra correspondía tanto a una región del vector como al minigen, confirmando la validación del vector construido.

Secuenciación del plásmido clona 4 y 5

Las muestras correspondientes a la clona 4 y 5 fueron analizadas por secuenciación Sanger utilizando el *primer* T7 *forward*. Las secuencias obtenidas se alinearon con la construcción del minigen y mostraron una coincidencia completa con el diseño original (**Figura 19 y 20**).

Minigen_construido	-----GGATCCTTATTGGGGTCAAAGGTGCTAAATCA
Secuenciación_Clona_4	TAGCTTGGTACGAGCTCGGATCCTTATTGGGGTCAAAGGTGCTAAATCA

Minigen_construido	AAGAACTTCGAGAGGTAAAAAAAAAAAAATTGCGGTATTCTTTGCTGA
Secuenciación_Clona_4	AAGAACTTCGAGAGGTAAAAAAAAAAAAATTGCGGTATTCTTTGCTGA

Minigen_construido	ATTCAATTGTTGAATTCAAATGCAATTTTCACTGAACCTAGGAAAGCTAA
Secuenciación_Clona_4	ATTCAATTGTTGAATTCAAATGCAATTTTCACTGAACCTAGGAAAGCTAA

Minigen_construido	TTTTCTTCTAATTATTATTATTTTTTTTATGATCTTACTACAGAACACT
Secuenciación_Clona_4	TTTTCTTCTAATTATTATTATTTTTTTTATGATCTTACTACAGAACACT

Minigen_construido	CAAACCACCATCAAGCTTTTCCAGGAATGCTGTCTCATTCCACTGACAG
Secuenciación_Clona_4	CAAACCACCATCAAGCTTTTCCAGGAATGCTGTCTCATTCCACTGACAG

Minigen_construido	AGTTGTTCTTATTGGAGGAAAACCCGATAGGGTTGTAGAGTGCATAAAGA
Secuenciación_Clona_4	AGTTGTTCTTATTGGAGGAAAACCCGATAGGGTTGTAGAGTGCATAAAGA

Minigen_construido	TCATCCTTGATCTTATATCTGAGGTACTTCAGAACACTTTTTGTTTACAT
Secuenciación_Clona_4	TCATCCTTGATCTTATATCTGAGGTACTTCAGAACACTTTTTGTTTACAT

Minigen_construido	TTGATTGTAATTCTACATTTTGTGTTGATGATGTTAGAAGGATTGGTA
Secuenciación_Clona_4	TTGATTGTAATTCTACATTTTGTGTTGATGATGTTAGAAGGATTGGTA

Minigen_construido	TGAAGCAGGACTCTTTTCTTCAGTCTGTAAGTTTGTTCATCTCTGTTT
Secuenciación_Clona_4	TGAAGCAGGACTCTTTTCTTCAGTCTGTAAGTTTGTTCATCTCTGTTT

Minigen_construido	TTAGTCTCCCATCAAAGGACGTGCACAGCCTTATGATCCCAATTGCGGCC
Secuenciación_Clona_4	TTAGTCTCCCATCAAAGGACGTGCACAGCCTTATGATCCCAATTGCGGCC

Minigen_construido	GC-----
Secuenciación_Clona_4	GCTCGAGGTACCCATTGCAACAAAACATCATCTCAGAAGAGGATCTGAA
	**

Figura 19. Alineamiento del minigen y la clona 4. Alineamiento de la secuencia con la que se diseñó el minigen con el resultado obtenido de la secuenciación Sanger del plásmido extraído de la clona 4 con el *primer* T7 *forward*.

Minigen_construido	-----GGATCCTTATTGGGGTCAAAGGTGCTAAATCA
Secuenciación_Clona_5	TAGCTTGGTACGAGCTCGGATCCTTATTGGGGTCAAAGGTGCTAAATCA

Minigen_construido	AAGAACTTCGAGAGGTAAAAAATAATTCTGCCGTATTCTTTGCTGA
Secuenciación_Clona_5	AAGAACTTCGAGAGGTAAAAAATAATTCTGCCGTATTCTTTGCTGA

Minigen_construido	ATTCAATTGTTGAATTCAAATGCAATTTTCACTGAACCTAGGAAAGCTAA
Secuenciación_Clona_5	ATTCAATTGTTGAATTCAAATGCAATTTTCACTGAACCTAGGAAAGCTAA

Minigen_construido	TTTTCTTCTAATTATTATTATTTTTTTTTATGATCTTACTACAGAACACT
Secuenciación_Clona_5	TTTTCTTCTAATTATTATTATTTTTTTTTATGATCTTACTACAGAACACT

Minigen_construido	CAAACCACCATCAAGCTTTTCCAGGAATGCTGTCCTCATTCCACTGACAG
Secuenciación_Clona_5	CAAACCACCATCAAGCTTTTCCAGGAATGCTGTCCTCATTCCACTGACAG

Minigen_construido	AGTTGTTCTTATTGGAGGAAAACCCGATAGGGTTGTAGAGTGCATAAAGA
Secuenciación_Clona_5	AGTTGTTCTTATTGGAGGAAAACCCGATAGGGTTGTAGAGTGCATAAAGA

Minigen_construido	TCATCCTTGATCTTATATCTGAGGTACTTCAGAACACTTTTTGTTTACAT
Secuenciación_Clona_5	TCATCCTTGATCTTATATCTGAGGTACTTCAGAACACTTTTTGTTTACAT

Minigen_construido	TTGATTTGAATTCTACATTTTGTGTTGATGATGATGTTAGAAGGATTGGTA
Secuenciación_Clona_5	TTGATTTGAATTCTACATTTTGTGTTGATGATGATGTTAGAAGGATTGGTA

Minigen_construido	TGAAGCAGGACTCTTTTCTTCAGTCTGTAAGTTTGTTCATCTCTGTTT
Secuenciación_Clona_5	TGAAGCAGGACTCTTTTCTTCAGTCTGTAAGTTTGTTCATCTCTGTTT

Minigen_construido	TTAGTCTCCCATCAAAGGACGTGCACAGCCTTATGATCCCAATTGCGGCC
Secuenciación_Clona_5	TTAGTCTCCCATCAAAGGACGTGCACAGCCTTATGATCCCAATTGCGGCC

Minigen_construido	GC-----
Secuenciación_Clona_5	GCTCGAGGTCACCCATTTCGAACAAAACATCTCAGAAGAGGATCTGAA
	**

Figura 20. Alineamiento del minigen y la clona 5. Alineamiento de la secuencia con la que se construyó el minigen con el resultado obtenido de la secuenciación Sanger del plásmido extraído de la clona 5 con el *primer T7 forward*.

Debido a que los resultados de la secuenciación mostraron la misma secuencia para la clona 4 como para la clona 5 se decidió utilizar solo una de ellas en la producción de más plásmido para realizar la transfección en la línea celular HEK293T.

Expansión y procesamiento del mensajero del minigen

Se transfectaron células HEK293T (**Figura 21**) con el plásmido producido de la clona 4, con lo que a continuación se realizó la extracción y purificación del ARN. Posteriormente, mediante una RT-PCR se analizó el ARN mensajero derivado del minigen.

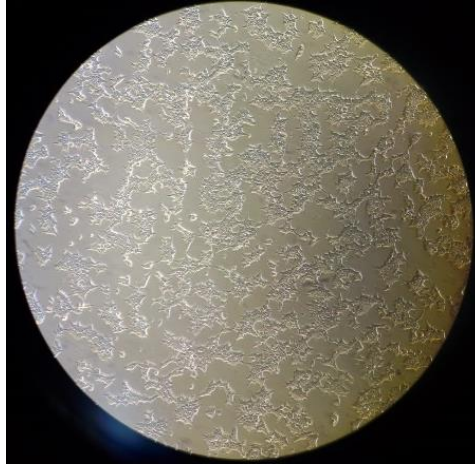


Figura 21. Células HEK293T transfectadas con vector de la clona 4.

Utilizando los *primers* T7 F y BGH R se amplificó la región correspondiente al ARNm dado que el promotor CMV facilita la transcripción del minigen y la amplificación abarca las regiones de unión de los *primers*. En la **Figura 22.A** se muestra el producto de la RT-PCR, mostrando una banda en un peso molecular entre 400 y 500 pb correspondiente al ARNm procesado del minigen.

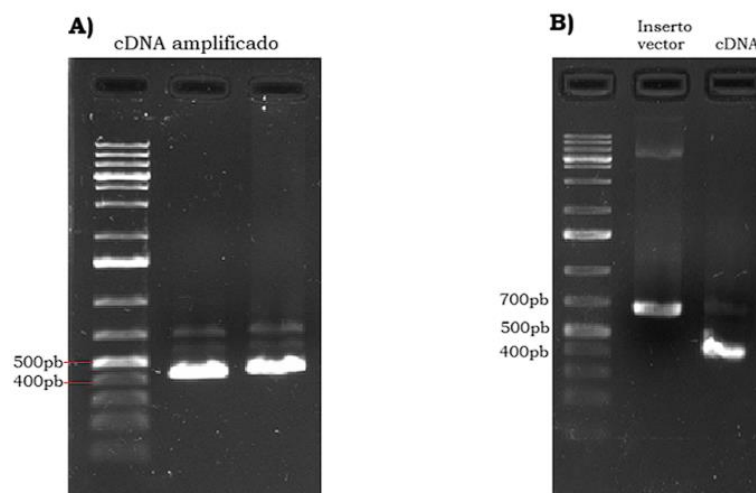


Figura 22. Producto de la RT-PCT. A) Pozos 2 y 3 corresponden al producto de PCR amplificado por los *primers* T7 y BGH del ADNc convertido a partir de la extracción del ARNm de las células transfectadas con el vector construido. B) Comparación entre ADNc amplificado e inserto del vector construido. El pozo 2 corresponde al fragmento amplificado a partir del vector construido por los *primers* T7 y BGH. El pozo 3 corresponde al fragmento amplificado por los mismos *primers* a partir del ADNc generado del ARNm extraído de las células transfectadas con el vector.

Se comparó el ARN procesado con el ADN del minigen, para lo que se realizó una amplificación por PCR en la región del plásmido (**Figura 22.B**) utilizando los mismos *primers* (T7 y BGH). La comparación de los fragmentos amplificados muestra una diferencia en el tamaño de las bandas; el ADNc generado a partir del ARNm presenta un menor tamaño que el correspondiente al minigen. Esto se debe a la eliminación de los intrones como parte del proceso de maduración del pre-ARNm.

Secuenciación de amplicón producido a partir de ADNc

Finalmente, se realizó una secuenciación Sanger del amplicón generado por PCR utilizando los *primers* T7 y BGH con el ADNc generado a partir de ARNm extraído de las células transfectadas. Este análisis permitió una determinación precisa de que el *splicing* había ocurrido de manera correcta y solo había presencia de las secuencias exónicas. En la **Figura 23** se puede observar el electroferograma de la secuencia, destacándose la calidad de los picos que corresponden a cada nucleótido.

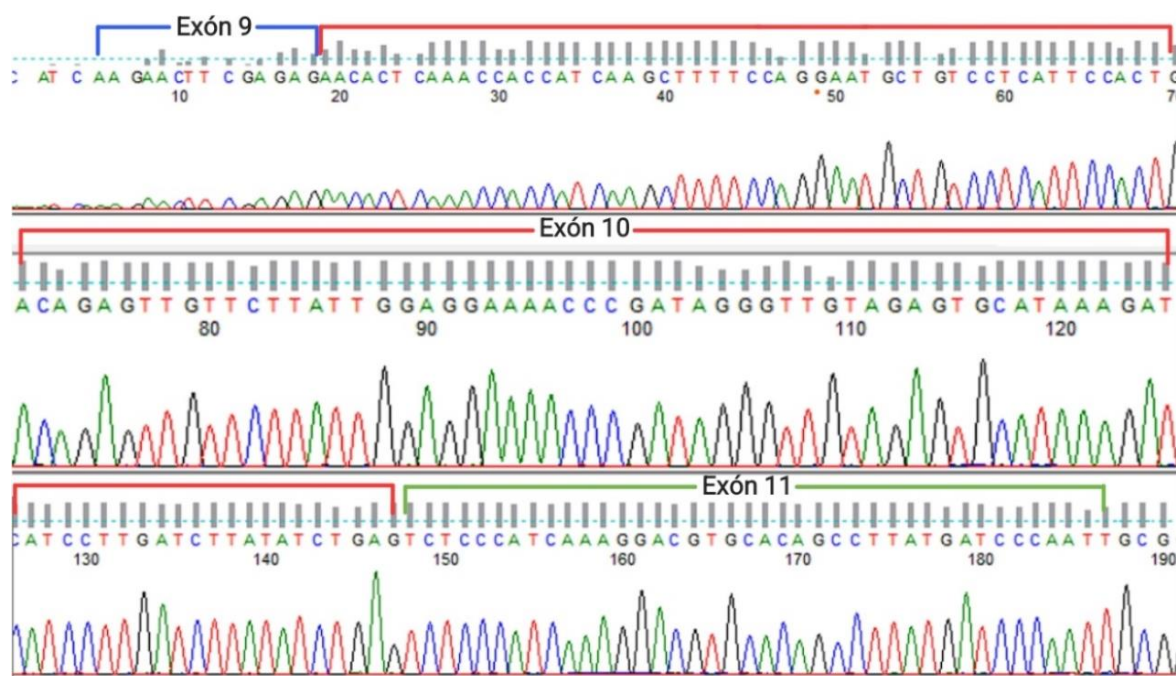
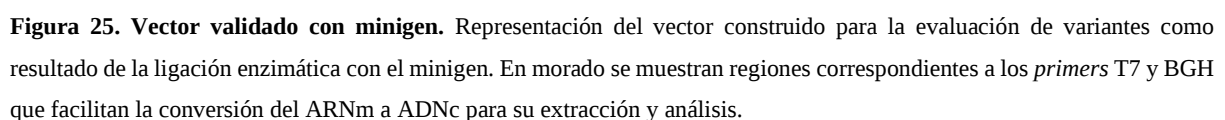


Figura 23. Electroferograma de la secuencia *forward* del ADNc amplificado por PCR. Secuenciación del amplicón producido con los *primers* T7 y BGH del ADNc generado a partir de ARNm.

Los resultados de la secuenciación se alinearon con la secuencia de solo los fragmentos de exones utilizados para la construcción del minigen. Como se puede observar en la **Figura 24**, existe una coincidencia completa con el exón 10, el fragmento final del exón 9 y el fragmento inicial del exón 11. Este alineamiento confirma que el ARN fue correctamente procesado y los intrones fueron eliminados conforme al mecanismo de *splicing* descrito.

Figura 24. Alineamiento exones minigen y secuenciación amplicón ADNc. Alineamiento de la secuencia de exones del minigen con el resultado de la secuenciación Sanger del amplicón producido con los *primers* T7 y BGH del ADNc generado a partir del ARNm. Cada color representa el exón correspondiente visto en la construcción del minigen.

Después de analizar el alineamiento, se comprobó el vector construido cumple con los objetivos propuestos para la evaluación de variantes génicas que alteran el *splicing*. La **Figura 25** resume el diseño final validado del vector, mostrando las regiones clave que lo componen y su funcionalidad demostrada a través de los experimentos realizados.



VIII. Discusión

La construcción del vector realizado en este trabajo permitirá la evaluación funcional de variantes génicas, ya que, dadas las características del minigen diseñado, es posible evaluar variantes en sitios de *splicing*, sitios de ramificación, en el tracto de polipirimidinas, así como en las regiones circundantes a ellos. Además, podrían evaluarse potenciadores y silenciadores exónicos (ESEs, ESSs), ya que, aunque no se cuenta con una identificación completa de estos elementos cis regulatorios^{50, 81}, con que exista una variante en la secuencia del exón podría ser considerada para su evaluación, sin embargo, esto dependerá del contexto de la variante, así como de los antecedentes clínicos y genéticos asociados y de quien se encargue de su análisis. El presente diseño del minigen contiene regiones accesibles para evaluación de variantes dentro de la región del exón central y los sitios circundantes que se encuentren dentro la región que contienen los dos sitios de corte EcoR1. Sin embargo, es importante destacar que la identificación de sitios de ramificación suele ser de las más complicadas ya que al ser solo un nucleótido (A) con secuencias circundantes variables, se dificulta establecer cuál será el sitio en el que se anclará el 5'ss. Es debido a esto que se han desarrollado algoritmos como LaBranchoR o BPP, que permiten su predicción y el efecto de variantes en el *splicing*, sin embargo, se destaca que son herramientas las cuales requieren una interpretación de resultados acompañada de su evaluación funcional.⁸²

Aunque existen protocolos y estudios previos que describen la construcción de vectores con minigenes^{76, 79, 80}, es crucial elegir cuidadosamente las regiones genómicas que se buscan estudiar. En este trabajo, el diseño del minigen incorporó las secuencias necesarias para garantizar un *splicing* adecuado, adaptándolo a las limitaciones de costo, accesibilidad, recursos y objetivos. Existen diseños como el de Liu, et al. (2024) en el que se sintetizó un minigen de 6052 pb con la secuencia completa de intrones y exones (tres exones separados por dos intrones) que directamente se clono en el vector de expresión para evaluar el efecto en el *splicing* de una variante intrónica cercana al 3'ss del intrón 10 (c.1030-2A>T) del gen *TSC1*.⁷⁹ Otro método para el diseño de un minigen es el realizado por Riedmayr, et al., (2018), en el que se utilizaron varios *primers* para la amplificación de la secuencia de interés en el gen *PRPH2* y se clonó el fragmento resultante en un vector seguido de una subclonación.⁷³ Este es un método ya descrito y conocido ⁷⁶, sin embargo es un método poco común el cual tiene un mayor costo y grado de complejidad.

Otro aspecto a destacar es el uso de minigenes para evaluar variantes en regiones que pertenecen a otro gen o contexto genómico, este tipo minigenes reciben el nombre de minigenes híbridos.⁷⁸ Es el caso de Baralle, et al., (2006) en el que se utilizó un minigen para evaluar una variante. En él se concluyó que incluso un cambio sinónimo (donde el aminoácido era el mismo) en la posición c.6792C>T del gen *NF1* produce un salto de exón, lo que demuestra de manera concluyente que el empalme alterado no se debe a mutaciones de sentido erróneo en esta posición para ese gen, sino a una interferencia con una región reguladora presente en el exón 37. Sin embargo, aún más destacable es que analizaron otra variante al inicio del codón donde sí cambió el aminoácido, pero no hubo salto del exón 37.⁵⁹ Esto subraya nuevamente la importancia de evaluar tanto variantes sinónimas como aquellas en regiones intrónicas cercanas a límites exón-intrón.

El diseño del minigen aquí presentado permite liberar un fragmento mediante la enzima de restricción EcoRI. Al evaluar alguna variante deberán ser utilizados *primers* específicos que contengan esta secuencia para insertar posteriormente el amplicón en el vector a través de extremos cohesivos. Los extremos EcoRI están estratégicamente colocados para liberar un exón completo flanqueado por fragmentos intrónicos, permitiendo la evaluación de variantes con homología en la construcción. Aunque el tamaño estándar de un fragmento de PCR es de alrededor de 700 pb, puede variar según el amplicón que se genere por medio de esta técnica.

Una posible limitante del plásmido construido estaría relacionada con su tamaño. Al ser tan grande podría dificultar la transformación de bacterias o la transfección en células eucariontes. Como perspectiva el minigen podría modificarse por mutagénesis dirigida para cambiar el sitio EcoR1 localizado en el intrón 9 por una secuencia de corte diferente, con ello se ampliaría la cantidad de regiones que podrían evaluarse sobreponiendo la limitante que posee el minigen al solo poder evaluar variantes contenidas dentro la región flanqueante por ambos sitios EcoR1.

La disrupción del *splicing* es una de las principales causas de enfermedades genéticas ⁶⁶, por lo que mejorar la comprensión de las variantes génicas en un contexto más amplio que solo en cómo afectan los aminoácidos que serán codificados en la traducción es un factor importante en el desarrollo hacia un mejor entendimiento de la expresión génica y el diagnóstico molecular. Se habían pasado por alto variantes génicas que no se encontraban registradas en bases de datos (como Clinvar y gnomAD). Es debido a esto que se ha buscado replantear la manera en que se interpretan las variantes génicas y en qué proceso de la expresión génica

tendrán impacto. Anteriormente se consideraba de mayor importancia solo el porcentaje del genoma que es codificante (~2%), sin embargo, en la actualidad se destaca que alrededor del 90% tiene una función bioquímica reguladora, por lo que es importante entender la dinámica completa de la regulación génica y cómo las variantes pueden alterar cada uno de los procesos implicados en la expresión de los genes.

Se ha buscado generar un conjunto de reglas y algoritmos con el fin de predecir si una variante génica puede afectar a un gen más allá del cambio que ocurre en el aminoácido. Principalmente se han utilizado herramientas *in silico* con el fin de establecer secuencias consenso y funciones específicas para dichas secuencias, sin embargo, debido a la gran cantidad de elementos reguladores y a su diversidad en secuencia, función, especificidad y a lo poco estudiados que están debido a la complejidad que presentan, no se ha podido establecer completamente un conjunto de reglas claras y acertadas.⁶⁸ Dentro de estas herramientas se encuentran SpliceAI y SAVnet, por mencionar algunos, sin embargo, existen gran cantidad de algoritmos diseñados como los enlistados por Anna, et al. (2018).⁵⁴ Herramientas como SpliceAI ayudan a predecir el impacto de una variante cercana a sitios de *splicing*, calculado si hay pérdida o ganancia de sitios aceptores o donadores de *splicing*, abarcando valores de 0 a 1, donde 1 indica un mayor impacto en el *splicing*, dando puntuaciones que indican la probabilidad de que la variante afecte el *splicing* normal en cada uno de los cuatro escenarios posibles. A pesar de esto, la gran cantidad de elementos involucrados en la regulación del *splicing* y la manera específica del contexto en que operan ^{50, 71, 83}, han hecho que la derivación de dicho código de *splicing* sea un desafío, así mismo y como ya se citó anteriormente, se indica que las predicciones realizadas de manera *in silico* deben de evaluarse funcionalmente (experimentalmente) para una mejor determinación en el efecto de las variantes. En la revisión realizada por Anna, et al., (2018) se probó una variante con SpliceAI (c.2204+6T>C) la cual se predijo como no perjudicial para el *splicing*, sin embargo, estudios experimentales mostraron que esta variante induce la exclusión del exón 20 en el gen *IKBKAP* asociado a disautonomía.⁵⁴

Se ha demostrado que existen variantes que aunque no están localizadas en el sitio consenso, sino circundantes a este, alteran el *splicing*.^{30, 58} Es por ello que no solo las variantes exónicas, sino también las variantes intrónicas cercanas a los límites exón-intrón deberían evaluarse cuidadosamente como posibles variantes patógenas, independientemente de la presencia de cambios de aminoácidos, dado que la manera en que estas variantes alteran el *splicing* es creando un nuevo sitio de *splicing* o interfiriendo con elementos reguladores y sus

interacciones, reduciendo o eliminando el reconocimiento de las secuencias por parte de dichos elementos, alterando por completo la regulación y la expresión génica. Un estudio realizado por Jung et al. (2021), en el que se analizaron 1627 variantes asociadas a cáncer debido a un *splicing* aberrante. Se encontraron variantes a distancias incluso mayores a 20 nucleótidos de los sitios de *splicing* (deep), así como a no más de 20 nucleótidos del sitio (proximal), y aún más destacable es la presencia de variantes de tipo silenciosas, reflejando el papel de las mutaciones sinónimas en el *splicing* del ARNm durante la tumorigénesis.⁸³ Estos hallazgos destacan que no solo las variantes exónicas son importantes, sino aquellas que ocurren en sitios intrónicos también requieren atención prioritaria en el análisis genómico.

La secuenciación genómica de tríos (hijo más ambos padres) ha demostrado que, en promedio, cada individuo presenta alrededor de 74 SNV de *novus*, lo que significa que no están presentes en ninguno de los progenitores.⁴ Esta gran variabilidad entre los genomas humanos individuales puede hacer que sea muy difícil determinar qué variantes son benignas y cuáles podrían estar asociadas con enfermedades. Incluso cuando está presente una variante asociada a una enfermedad, puede encontrarse dentro de un contexto genómico de millones de otras diferencias con respecto a la secuencia de referencia, algunas de las cuales podrían influir en la gravedad de la enfermedad en el individuo.

En este contexto, el vector construido en el presente trabajo ofrece una herramienta valiosa para la evaluación de variantes génicas no solo del gen *HNRNPK*, que es al que pertenecen los fragmentos, sino también a otros genes que cumplan con la misma regla GU-AG para el *splicing*, siendo el caso del 98.7% de los genes humanos.⁵⁴ Este vector tiene el potencial de ser utilizado tanto en aplicaciones prenatales como postnatales para el diagnóstico molecular, ampliando así las posibilidades de caracterización y comprensión de las variantes genéticas en un marco clínico.

IX. Conclusión

Como resultado de este trabajo se puede establecer que el vector construido con un minigen a partir de una subclonación es completamente funcional para el análisis de variantes génicas que alteran el *splicing* ubicadas tanto en sitios de *splicing* como en regiones circundantes. Así mismo, se cumplió de manera satisfactoria con cada uno de los objetivos planteados, por lo que se concluye que no fue necesaria la secuencia completa de los intrones para el *splicing* de este minigen dado que se incluían los elementos mínimos necesarios (SS, BPS, PPT). Con esto se da accesibilidad en el Laboratorio de Inmunogenómica y Enfermedades Metabólicas del Instituto Nacional de Medicina Genómica de una herramienta que no es de uso comercial para el análisis de variantes asociadas a *splicing*.

Simultáneamente se concluye que el análisis de variantes génicas debe de realizarse desde una perspectiva más amplia, considerando no solo el impacto que tienen en la traducción de las proteínas. Como se expuso a lo largo de este trabajo, variantes génicas de tipo silenciosas o de cambio de sentido pueden no afectar directamente la función de la proteína si se analizan únicamente bajo enfoques convencionales como análisis de exoma. Sin embargo, estas variantes podrían tener un impacto significativo en otros puntos críticos de la expresión génica, como el *splicing*, lo que subraya la importancia de adoptar un enfoque integral en su evaluación.

X. Referencias

1. Misteli T. (2020). *The Self-Organizing Genome: Principles of Genome Architecture and Function*. *Cell*. 2020 Oct 1;183(1):28-45. doi: 10.1016/j.cell.2020.09.014. Epub 2020 Sep 24. PMID: 32976797; PMCID: PMC7541718.
2. Xiong Q, Zhang Y, Li J, Zhu Q. Small Non-Coding RNAs in Human Cancer. *Genes (Basel)*. 2022 Nov 9;13(11):2072. doi: 10.3390/genes13112072. PMID: 36360311; PMCID: PMC9690286.3 (18)
3. Alpsoy A, Sood S, Dykhuizen EC. (2021). At the Crossroad of Gene Regulation and Genome Organization: Potential Roles for ATP-Dependent Chromatin Remodelers in the Regulation of CTCF-Mediated 3D Architecture. *Biology (Basel)*. 2021 Mar 27;10(4):272. doi: 10.3390/biology10040272. PMID: 33801596; PMCID: PMC8066914.
4. Jackson M, Marks L, May GHW, Wilson JB. (2018). The genetic basis of disease. *Essays Biochem*. 2018 Dec 2;62(5):643-723. doi: 10.1042/EBC20170053. Erratum in: *Essays Biochem*. 2020 Oct 8;64(4):681. doi: 10.1042/EBC20170053_COR. PMID: 30509934; PMCID: PMC6279436.
5. Hartwell, L. H., Hood, L., Goldberg, M. L., Reynolds, A. E., Silver, L. M., Veres, R. C., & Rindfleisch, T. C. (2016). *From Genes to Genomes* (5th ed.). McGraw-Hill Education.
6. Dixon, P. (2010). Structure and function of the genome. *Basic Science in Obstetrics and Gynaecology*, 1–11. doi:10.1016/b978-0-443-10281-3.00005-1
7. Strachan, T. & Read, A. (2019). *Human Molecular Genetics*. CRC PRESS. pp: 20, 22, 33, 138
8. Furuno M, Kasukawa T, Saito R, Adachi J, Suzuki H, Baldarelli R, Hayashizaki Y, Okazaki Y. (2003). CDS annotation in full-length cDNA sequence. *Genome Res*. 2003 Jun;13(6B):1478-87. doi: 10.1101/gr.1060303. PMID: 12819146; PMCID: PMC403693.
9. ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*. Sep 6;489(7414):57-74. doi: 10.1038/nature11247. PMID: 22955616; PMCID: PMC3439153.
10. Elkon, R., Agami, R. (2017). Characterization of noncoding regulatory DNA in the human genome. *Nat Biotechnol* 35, 732–746. <https://doi.org/10.1038/nbt.3863>
11. Mignone, F., Gissi, C., Liuni, S., & Pesole, G. (2002). Untranslated Regions of mRNAs. *Genome Biology*, 3(3), reviews0004.1. doi:10.1186/gb-2002-3-3-reviews0004
12. Reuveni E, Getselter D, Oron O, Elliott E. (2018). Differential contribution of cis and trans gene transcription regulatory mechanisms in amygdala and prefrontal cortex and modulation by social stress. *Sci Rep*. Apr 20;8(1):6339. doi: 10.1038/s41598-018-24544-3. PMID: 29679052; PMCID: PMC5910421.
13. Gebert, L.F.R., MacRae, I.J. (2019). Regulation of microRNA function in animals. *Nat Rev Mol Cell Biol* 20, 21–37. <https://doi.org/10.1038/s41580-018-0045-7>
14. Schuster SL, Hsieh AC. (2019). The Untranslated Regions of mRNAs in Cancer. *Trends Cancer*. 2019 Apr;5(4):245-262. doi: 10.1016/j.trecan.2019.02.011. Epub 2019 Mar 22. PMID: 30961831; PMCID: PMC6465068.

15. Brown, T. (2018). *Genomes 4* (4th ed.). Garland Science. pp: 1, 2, 4, 247
16. Poltronieri P, Sun B, Mallardo M. (2015). RNA Viruses: RNA Roles in Pathogenesis, Coreplication and Viral Load. *Curr Genomics*. 2015 Oct;16(5):327-35. doi:10.2174/1389202916666150707160613. PMID: 27047253; PMCID: PMC4763971.
17. Patil VS, Zhou R, Rana TM. Gene regulation by non-coding RNAs. *Crit Rev Biochem Mol Biol*. 2014 Jan-Feb;49(1):16-32. doi: 10.3109/10409238.2013.844092. Epub 2013 Oct 28. PMID:24164576; PMCID: PMC4721600.
18. Mignone F, Gissi C, Liuni S, Pesole G. (2002). Untranslated regions of mRNAs. *Genome Biol*. 2002;3(3):REVIEWS0004. doi: 10.1186/gb-2002-3-3-reviews0004. Epub 2002 Feb 28. PMID:11897027; PMCID: PMC139023.
19. Kaikkonen MU, Lam MT, Glass CK. (2011). Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovasc Res*. 2011 Jun 1;90(3):430-40. doi: 10.1093/cvr/cvr097. Epub 2011 May 9. PMID: 21558279; PMCID: PMC3096308.
20. Xiong Q, Zhang Y, Li J, Zhu Q. Small Non-Coding RNAs in Human Cancer. *Genes (Basel)*. 2022 Nov 9;13(11):2072. doi: 10.3390/genes13112072. PMID: 36360311; PMCID: PMC9690286.
21. Cao J. The functional role of long non-coding RNAs and epigenetics. *Biol Proced Online*. 2014 Sep 15;16:11. doi: 10.1186/1480-9222-16-11. PMID: 25276098; PMCID: PMC4177375.
22. Patil VS, Zhou R, Rana TM. Gene regulation by non-coding RNAs. *Crit Rev Biochem Mol Biol*. 2014 Jan-Feb;49(1):16-32. doi: 10.3109/10409238.2013.844092. Epub 2013 Oct 28. PMID: 24164576; PMCID: PMC4721600.
23. Ho PJ, Lloyd SM, Bao X. Unwinding chromatin at the right places: how BAF is targeted to specific genomic locations during development. *Development*. 2019 Sep30;146(19):dev178780. doi: 10.1242/dev.178780. PMID: 31570369; PMCID: PMC6803370.
24. Biegel JA, Busse TM, Weissman BE. SWI/SNF chromatin remodeling complexes and cancer. *Am J Med Genet C Semin Med Genet*. 2014 Sep;166C(3):350-66. doi: 10.1002/ajmg.c.31410. Epub 2014 Aug 28. PMID: 25169151; PMCID: PMC4516040.
25. Tyagi M, Imam N, Verma K, Patel AK. Chromatin remodelers: We are the drivers!! *Nucleus*. 2016 Jul 3;7(4):388-404. doi: 10.1080/19491034.2016.1211217. Epub 2016 Jul 18. PMID: 27429206; PMCID: PMC5039004.
26. Brooker, R. J. (2021). *Genetics: analysis and principles* (7th ed.). McGraw-Hill. pp: 288, 296, 297, 298, 473, 474
27. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2022). *Molecular Biology of the Cell*. (7a ed.). Garland Science. pp: 195, 333, 334, 365,
28. Krebs, J. E., Lewin, B., Kilpatrick, S. T., & Goldstein, E. S. (2017). *Lewin's genes XII / Jocelyn E. Krebs, Elliott S. Goldstein, Stephen T. Kilpatrick. Jones and Bartlett*.
29. Wang, GS., Cooper, T. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 8, 749–761 (2007). <https://doi.org/10.1038/nrg2164>

30. Rogalska, M.E., Vivori, C. & Valcárcel, J. Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects. *Nat Rev Genet* 24, 251–269 (2023). <https://doi.org/10.1038/s41576-022-00556-8>
31. Marques R, Lacerda R, Romão L. Internal Ribosome Entry Site (IRES)-Mediated Translation and Its Potential for Novel mRNA-Based Therapy Development. *Biomedicines*. 2022 Aug 2;10(8):1865. doi: 10.3390/biomedicines10081865. PMID: 36009412; PMCID: PMC9405587.
32. Ryczek N, Łys A, Makalowska I. The Functional Meaning of 5'UTR in Protein-Coding Genes. *Int J Mol Sci*. 2023 Feb 3;24(3):2976. doi: 10.3390/ijms24032976. PMID: 36769304; PMCID: PMC9917990.
33. Sonenberg N, Hinnebusch AG. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*. 2009 Feb 20;136(4):731-45. doi: 10.1016/j.cell.2009.01.042. PMID: 19239892; PMCID: PMC3610329.
34. Prabantu VM, Naveenkumar N, Srinivasan N. (2021). Influence of Disease-Causing Mutations on Protein Structural Networks. *Front Mol Biosci*. 2021 Mar 10;7:620554. doi:10.3389/fmolb.2020.620554. PMID: 33778000; PMCID: PMC7987782.
35. Li M, Goncarenco A, Panchenko AR. (2018). Annotating Mutational Effects on Proteins and Protein Interactions: Designing Novel and Revisiting Existing Protocols. *Methods Mol Biol*. 2017;1550:235-260. doi: 10.1007/978-1-4939-6747-6_17. PMID: 28188534; PMCID: PMC5388446.
36. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res*. 2011 Sep 1;39(17):e118. doi: 10.1093/nar/gkr407. Epub 2011 Jul 3. PMID: 21727090; PMCID: PMC3177186.
37. Delaunay, S., Helm, M. & Frye, M. RNA modifications in physiology and disease: towards clinical applications. *Nat Rev Genet* 25, 104–122 (2024). <https://doi.org/10.1038/s41576-023-00645-2>
38. Wiener, D., Schwartz, S. The epitranscriptome beyond m6A. *Nat Rev Genet* 22, 119–131 (2021). <https://doi.org/10.1038/s41576-020-00295-8>
39. Zhang, X., Zhu, WY., Shen, SY. et al. Biological roles of RNA m7G modification and its implications in cancer. *Biol Direct* 18, 58 (2023). <https://doi.org/10.1186/s13062-023-00414-5>
40. Hocine S, Singer RH, Grünwald D. RNA processing and export. *Cold Spring Harb Perspect Biol*. 2010 Dec;2(12):a000752. doi: 10.1101/cshperspect.a000752. Epub 2010 Oct 20. PMID:20961978; PMCID: PMC2982171.
41. Anand Ramanathan, G. Brett Robb, Siu-Hong Chan, mRNA capping: biological functions and applications, *Nucleic Acids Research*, Volume 44, Issue 16, 19 September 2016, Pages 7511–7526, <https://doi.org/10.1093/nar/gkw551>
42. Proudfoot N. New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol*. 2004 Jun;16(3):272-8. doi: 10.1016/j.ceb.2004.03.007. PMID: 15145351.
43. Ren, F., Zhang, N., Zhang, L. et al. Alternative Polyadenylation: a new frontier in post transcriptional regulation. *Biomark Res* 8, 67 (2020). <https://doi.org/10.1186/s40364-020-00249-6>

44. Girardini KN, Olthof AM, Kanadia RN. Introns: the "dark matter" of the eukaryotic genome. *Front Genet.* 2023 May 16;14:1150212. doi: 10.3389/fgene.2023.1150212. PMID: 37260773; PMCID: PMC10228655.
45. Sickmier EA, Frato KE, Shen H, Paranawithana SR, Green MR, Kielkopf CL. Structural basis for polypyrimidine tract recognition by the essential pre-mRNA splicing factor U2AF65. *Mol Cell.* 2006 Jul 7;23(1):49-59. doi: 10.1016/j.molcel.2006.05.025. PMID: 16818232; PMCID:PMC2043114.
46. Fredericks AM, Cygan KJ, Brown BA, Fairbrother WG. RNA-Binding Proteins: Splicing Factors and Disease. *Biomolecules.* 2015 May 13;5(2):893-909. doi: 10.3390/biom5020893. PMID: 25985083; PMCID: PMC4496701.
47. Will CL, Lührmann R. Spliceosome structure and function. *Cold Spring Harb Perspect Biol.* 2011 Jul 1;3(7):a003707. doi: 10.1101/cshperspect.a003707. PMID: 21441581; PMCID: PMC3119917.
48. Wang Y, Wang Z. Systematical identification of splicing regulatory cis-elements and cognate trans-factors. *Methods.* 2014 Feb;65(3):350-8. doi: 10.1016/j.ymeth.2013.08.019. Epub 2013 Aug 22. PMID: 23974071; PMCID: PMC3932149. Wang, GS., Cooper, T. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 8, 749–761 (2007). <https://doi.org/10.1038/nrg2164>
49. Yamauchi H, Nishimura K, Yoshimi A. Aberrant RNA splicing and therapeutic opportunities in cancers. *Cancer Sci.* 2022 Feb;113(2):373-381. doi: 10.1111/cas.15213. Epub 2021 Nov 30. PMID: 34812550; PMCID: PMC8819303.
50. Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA.* 2008 May;14(5):802-13. doi: 10.1261/rna.876308. Epub 2008 Mar 27. PMID: 18369186; PMCID: PMC2327353.
51. Van Nostrand EL, Freese P, Pratt GA, Wang X, Wei X, Xiao R, Blue SM, Chen JY, Cody NAL, Dominguez D, Olson S, Sundararaman B, Zhan L, Bazile C, Bouvrette LPB, Bergalet J, Duff MO, Garcia KE, Gelboin-Burkhart C, Hochman M, Lambert NJ, Li H, McGurk MP, Nguyen TB, Palden T, Rabano I, Sathe S, Stanton R, Su A, Wang R, Yee BA, Zhou B, Louie AL, Aigner S, Fu XD, Lécuyer E, Burge CB, Graveley BR, Yeo GW. A large-scale binding and functional map of human RNA-binding proteins. *Nature.* 2020 Jul;583(7818):711-719. doi: 10.1038/s41586-020-2077-3. Epub 2020 Jul 29. Erratum in: *Nature.* 2021 Jan;589(7842):E5. doi: 10.1038/s41586-020-03067-w. PMID: 32728246; PMCID: PMC7410833.
52. Jeong S. SR Proteins: Binders, Regulators, and Connectors of RNA. *Mol Cells.* 2017 Jan;40(1):1-9. doi: 10.14348/molcells.2017.2319. Epub 2017 Jan 26. PMID: 28152302; PMCID: PMC5303883.
53. Marasco, L.E., Kornblihtt, A.R. The physiology of alternative splicing. *Nat Rev Mol Cell Biol* 24, 242–254 (2023). <https://doi.org/10.1038/s41580-022-00545-z>
54. Anna, A., Monika, G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genetics* 59, 253–268 (2018). <https://doi.org/10.1007/s13353-018-0444-7>

55. Tao, Y., Zhang, Q., Wang, H. et al. Alternative splicing and related RNA binding proteins in human health and disease. *Sig Transduct Target Ther* 9, 26 (2024). <https://doi.org/10.1038/s41392-024-01734-2>
56. Wang Y, Liu J, Huang BO, Xu YM, Li J, Huang LF, Lin J, Zhang J, Min QH, Yang WM, Wang XZ. Mechanism of alternative splicing and its regulation. *Biomed Rep*. 2015 Mar;3(2):152-158. doi: 10.3892/br.2014.407. Epub 2014 Dec 17. PMID: 25798239; PMCID: PMC4360811.
57. Karp, G. (2020). *Molecular Biology and Cellular Biology* (8th ed.). Wiley. pp: 406, 408, 412, 417, 418, 419, 437,
58. López-Bigas, N., Audit, B., Ouzounis, C., Parra, G., & Guigó, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? *FEBS Letters*, 579(9), 1900–1903. doi:10.1016/j.febslet.2005.02.047
59. Baralle, M., Skoko, N., Knezevich, A., De Conti, L., Motti, D., Bhuvanagiri, M., ... Baralle, F.E. (2006). NF1mRNA biogenesis: Effect of the genomic milieu in splicing regulation of theNF1exon 37 region. *FEBS Letters*, 580(18), 4449–4456. doi:10.1016/j.febslet.2006.07.018
60. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015 May;17(5):405-24. doi: 10.1038/gim.2015.30. Epub 2015 Mar 5. PMID: 25741868; PMCID: PMC4544753.
61. Lappalainen T, MacArthur DG. From variant to function in human disease genetics. *Science*. 2021 Sep 24;373(6562):1464-1468. doi: 10.1126/science.abi8207. Epub 2021 Sep 23. PMID: 34554789.
62. Peña-Martínez EG, Rodríguez-Martínez JA. Decoding Non-coding Variants: Recent Approaches to Studying Their Role in Gene Regulation and Human Diseases. *Front Biosci (Schol Ed)*. 2024 Mar 1;16(1):4. doi: 10.31083/j.fbs1601004. PMID: 38538340; PMCID: PMC11044903.
63. Fredericks AM, Cygan KJ, Brown BA, Fairbrother WG. RNA-Binding Proteins: Splicing Factors and Disease. *Biomolecules*. 2015 May 13;5(2):893-909. doi: 10.3390/biom5020893. PMID: 25985083; PMCID: PMC4496701.
64. Islam MR, Nagar P, Neetole ST, Wan L, Rahman MA. RNA Splicing in Cancer and Targeted Therapies. *Genes (Basel)*. 2023 Oct 29;14(11):2020. doi: 10.3390/genes14112020. PMID: 38002963; PMCID: PMC10671003.
65. Baralle, D., & Buratti, E. (2017). RNA splicing in human disease and in the clinic. *Clinical Science*, 131(5), 355–368. doi:10.1042/cs20160211
66. Blakes, A.J.M., Wai, H.A., Davies, I. et al. A systematic analysis of splicing variants identifies new diagnoses in the 100,000 Genomes Project. *Genome Med* 14, 79 (2022). <https://doi.org/10.1186/s13073-022-01087-x>
67. Rivas, M. A., Pirinen, M., Conrad, D. F., Lek, M., Tsang, E. K., Karczewski, K. J., ... Fromer, M. (2015). Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science*, 348(6235), 666–669. doi:10.1126/science.1261877

68. Lord J, Gallone G, Short PJ, McRae JF, Ironfield H, Wynn EH, Gerety SS, He L, Kerr B, Johnson DS, McCann E, Kinning E, Flinter F, Temple IK, Clayton-Smith J, McEntagart M, Lynch SA, Joss S, Douzgou S, Dabir T, Clowes V, McConnell VPM, Lam W, Wright CF, FitzPatrick DR, Firth HV, Barrett JC, Hurles ME; Deciphering Developmental Disorders study. Pathogenicity and selective constraint on variation near splice sites. *Genome Res.* 2019 Feb;29(2):159-170. doi: 10.1101/gr.238444.118. Epub 2018 Dec 26. PMID: 30587507; PMCID: PMC6360807
69. Chong R, Insigne KD, Yao D, Burghard CP, Wang J, Hsiao YE, Jones EM, Goodman DB, Xiao X, Kosuri S. A Multiplexed Assay for Exon Recognition Reveals that an Unappreciated Fraction of Rare Genetic Variants Cause Large-Effect Splicing Disruptions. *Mol Cell.* 2019 Jan 3;73(1):183-194.e8. doi: 10.1016/j.molcel.2018.10.037. Epub 2018 Nov 29. PMID: 30503770; PMCID: PMC6599603.
70. Yu, Z., Coorens, T.H.H., Uddin, M.M. et al. Genetic variation across and within individuals. *Nat Rev Genet* 25, 548–562 (2024). <https://doi.org/10.1038/s41576-024-00709-x>
71. Shiraishi Y, Kataoka K, Chiba K, Okada A, Kogure Y, Tanaka H, Ogawa S, Miyano S. A comprehensive characterization of cis-acting splicing-associated variants in human cancer. *Genome Res.* 2018 Aug;28(8):1111-1125. doi: 10.1101/gr.231951.117. Epub 2018 Jul 16. PMID: 30012835; PMCID: PMC6071634.
72. Sanders SJ, Schwartz GB, Farh KK. Clinical impact of splicing in neurodevelopmental disorders. *Genome Med.* 2020 Apr 24;12(1):36. doi: 10.1186/s13073-020-00737-2. PMID: 32331533; PMCID: PMC7183108.
73. Riedmayr LM, Böhm S, Michalakakis S, Becirovic E. Construction and Cloning of Minigenes for in vivo Analysis of Potential Splice Mutations. *Bio Protoc.* 2018 Mar 5;8(5):e2760. doi: 10.21769/BioProtoc.2760. PMID: 34179284; PMCID: PMC8203986.
74. Prazeres DMF, Monteiro, GA. (2014). Plasmid Biopharmaceuticals. *Microbiol Spectr* 2:10.1128/microbiolspec.plas-0022-2014. <https://doi.org/10.1128/microbiolspec.plas-0022-2014>
75. Tamayo A, Núñez-Moreno G, Ruiz C, Plaisancie J, Damian A, Moya J, Chassaing N, Calvas P, Ayuso C, Minguez P, Corton M. Minigene Splicing Assays and Long-Read Sequencing to Unravel Pathogenic Deep-Intronic Variants in PAX6 in Congenital Aniridia. *Int J Mol Sci.* 2023 Jan 13;24(2):1562. doi: 10.3390/ijms24021562. PMID: 36675087; PMCID: PMC9863980.
76. Cooper, T. A. (2005). Use of minigene systems to dissect alternative splicing elements. *Methods*, 37(4), 331–340. doi:10.1016/j.ymeth.2005.07.015
77. Marian AJ. Clinical Interpretation and Management of Genetic Variants. *JACC Basic TranslSci.* 2020 Oct 26;5(10):1029-1042. doi: 10.1016/j.jacbts.2020.05.013. PMID: 33145465; PMCID: PMC7591931.
78. Sharma N, Sosnay PR, Ramalho AS, Douville C, Franca A, Gottschalk LB, Park J, Lee M, Vecchio-Pagan B, Raraigh KS, Amaral MD, Karchin R, Cutting GR. Experimental assessment of splicing variants using expression minigenes and comparison with in silico predictions. *Hum Mutat.* 2014 Oct;35(10):1249-59. doi:

79. Liu L, Wang Y, Zhang Z, Yu C, Chen J. Identification of a novel TSC1 gene variant in a patient with atypical vitiligo-like skin lesions: Unveiling the hidden tuberous sclerosis complex. *Mol Genet Genomic Med*. 2024 Mar;12(3):e2403. doi: 10.1002/mgg3.2403. PMID: 38439608; PMCID: PMC10912791.
80. Song J, Lei J, Zhang J, Zhang A, Gan W, Zheng B, Wang C, Gong J. Case Report: A Novel Non-Canonical Splice Site Variant (c.1638+7T>C) in TRPM6 Cause Primary Homagnesemia With Secondary Hocalcemia. *Front Pediatr*. 2022 May 25;10:834241. doi: 10.3389/fped.2022.834241. PMID: 35692977; PMCID: PMC9174589.
81. Barash, Y., Calarco, J., Gao, W. et al. Deciphering the splicing code. *Nature* 465, 53–59 (2010). <https://doi.org/10.1038/nature09000>
82. Leman, R., Tubeuf, H., Raad, S. et al. Assessment of branch point prediction tools to predict physiological branch points and their alteration by variants. *BMC Genomics* 21, 86 (2020). <https://doi.org/10.1186/s12864-020-6484-5>
83. Jung, H., Lee, K.S. & Choi, J.K. Comprehensive characterisation of intronic mis-splicing mutations in human cancers. *Oncogene* 40, 1347–1361 (2021). <https://doi.org/10.1038/s41388-020-01614-3>