

**TESIS DOCTORAL:**

**REGULACIÓN E IMPACTO EPIGENÉTICO DEL  
ELEMENTO MÓVIL HUMANO LINE-1 EN UN GENOMA  
PLURIPOTENTE**

Programa Oficial de Doctorado en Biomedicina

Universidad de Granada

**Doctorando: Juan David Cano Martínez**

**Becario FPU**

Directores de Tesis: José Luís García Pérez y David Landeira Frías

Tesis Doctoral Realizada en:

Genyo, Centro Pfizer-Universidad de Granada-Junta de Andalucía de  
Genómica e Investigación Oncológica

Editor: Universidad de Granada. Tesis Doctorales

Autor: Juan David Cano Martínez

ISBN: 978-84-9163-581-9

URI: <http://hdl.handle.net/10481/48746>

A Lupe, a Elián y a mis padres



“El fracaso no nos sobrecogerá mientras que el deseo para alcanzar el éxito sea lo suficientemente poderoso”,

*El Panadero*



# RESUMEN





## **REGULACIÓN E IMPACTO EPIGENÉTICO DEL ELEMENTO MÓVIL HUMANO LINE-1 EN UN GENOMA PLURIPOTENTE**

LINE-1 o L1 (Long INterspersed Element class 1) es un tipo de ADN móvil, concretamente un retrotransposón autónomo de la clase no LTR (Long Terminal Repeats), que se encuentra actualmente activo en el genoma humano. Existen unas 500.000 copias en el genoma, ocupando un 17% del mismo, y de las cuales se estima que tan solo 80-100 copias son retrocompetentes o activas. Éstas tienen una longitud de 6 kb, presentan una región no traducida con actividad promotora y dos marcos abiertos de lectura que codifican la maquinaria enzimática requerida para su movilización. Los elementos LINE-1 retrocompetentes se movilizan mediante un mecanismo de “copia y pega” denominado retrotransposición, generando una nueva copia del elemento en otra localización genómica, usando un intermediario de ARN. La retrotransposición, que mayoritariamente se produce en etapas tempranas de desarrollo, inserta una nueva secuencia de DNA en el genoma y esto produce una serie de consecuencias para el genoma, pudiendo incluso causar una nueva enfermedad. De hecho, se han identificado más de 100 enfermedades asociadas a la actividad de LINE-1. Por ello, existen diversos mecanismos celulares para el control de la retrotransposición. A nivel post-transcripcional existen evidencias de que distintas especies de ARN (piRNAs, siRNAs y microRNAs) que pueden controlar la retrotransposición. Además, existen datos que sugieren que el complejo proteico DROSHA/DGCR8 implicado en la biogénesis de microRNAs podría regular la retrotransposición. No obstante, se desconoce el efecto real sobre la retrotransposición así como el mecanismo. Por otro lado, existen evidencias de que las copias de LINE-1 presentes en el genoma humano podrían estar reguladas epigenéticamente mediante modificaciones de histonas. Además, se ha demostrado que la mayoría de las nuevas inserciones de L1 en células humanas pluripotentes, producidas desde plásmidos que contienen elementos L1 retrocompetentes recombinantes, son silenciadas epigenéticamente mediante desacetilación de histonas. Sin embargo, se desconocen cuáles son las modificaciones de histonas implicadas en esta represión de las nuevas inserciones de LINE-1 cuando se integran en el genoma humano. Además, se desconoce el impacto epigenético de las nuevas inserciones sobre el genoma adyacente durante el desarrollo embrionario temprano.

En primer lugar, mediante ensayos de retrotransposición se determinó que el microprocesador (DROSHA/DGCR8) afecta negativamente la retrotransposición probablemente a través de su interacción con la región 5' no traducida del ARNm de LINE-1.

Por otro lado, para conocer más acerca de la regulación epigenética de las inserciones *de novo* de LINE-1 y su impacto epimutagenico en un genoma pluripotente, en esta tesis se ha generado una serie de líneas clonales con una nueva inserción de LINE-1 en células pluripotentes humanas (células de carcinoma embrionario y células madre embrionarias). Para ello, se utilizó un nuevo vector de retrotransposición que permite la recuperación de la inserción y, por tanto, conocer su localización genómica. Posteriormente, mediante inmunoprecipitación de cromatina seguida de PCR cuantitativa se ha demostrado que la mayoría de las nuevas inserciones de LINE-1 acumulan modificaciones de histonas asociadas con represión transcripcional como trimetilación de la lisina 9 de la histona 3 (H3K9me3) y probablemente trimetilación de la lisina 27 de la histona 3. Además, se ha demostrado que algunas de las nuevas inserciones de LINE-1 generan cambios en los niveles de enriquecimiento de H3K9me3 en el genoma adyacente llegando estos cambios hasta 1-1,5kb desde el sitio de inserción.

Por otro lado, también se han generado datos que sugieren que las inserciones nuevas endógenas de LINE-1 en células madre embrionarias humanas pueden estar también enriquecidas en H3K9me3. Sorprendentemente, el promotor de una de las inserciones endógenas se encontraba enriquecido en una modificación asociada a actividad transcripcional como es la trimetilación de la lisina 4 de la histona H3 (H3K4me3), lo que supondría que teóricamente pudiera ser transcrita y, por tanto, que pudiera llevar a cabo un nuevo ciclo de retrotransposición.

Estos hallazgos son importantes, pues constituyen, por un lado el primer dato existente de que una nueva inserción de un retrotransposón activo en el genoma humano puede generar “epimutaciones” las cuales podrían tener un gran impacto funcional dependiendo de su localización y, por otro lado, la primera evidencia de que una nueva inserción endógena de LINE-1 podría permanecer activa después de su integración en el genoma humano pudiendo realizar un nuevo ciclo de retrotransposición.





## TABLA DE CONTENIDOS

<b>1. INTRODUCCIÓN</b>	<b>23</b>
<b>1.1. SECUENCIAS REPETIDAS EN EL GENOMA HUMANO</b>	<b>23</b>
<b>1.2. ELEMENTOS MÓVILES EN EL GENOMA HUMANO</b>	<b>24</b>
1.2.1. ADN TRANSPOSONES	24
1.2.2. RETROVIRUS ENDÓGENOS	25
1.2.3. LINE-1	29
1.2.4. SINES	30
1.2.4.1. Elementos Alus	31
1.2.4.2. Elementos SVAs	32
1.2.5. PSEUDOGENES PROCESADOS.	35
<b>1.3. UN RETROTRANPOSÓN AUTÓNOMO ACTIVO: LINE-1.</b>	<b>35</b>
1.3.1. ESTRUCTURA DE LAS COPIAS RETROCOMPETENTES DE LINE-1.	36
1.3.2. CICLO DE RETROTRANSPOSICIÓN	38
1.3.1. ENCUADRE TEMPORAL DE LA RETROTRANSPOSICIÓN	40
1.3.2. CONSECUENCIAS DE LA RETROTRANSPOSICIÓN SOBRE EL GENOMA HOSPEDADOR	42
1.3.2.1. Efectos sobre la expresión génica	42
1.3.2.2. Efectos estructurales sobre el sitio de inserción.	43
1.3.2.3. Otros efectos	45
1.3.3. ENSAYO DE RETROTRANSPOSICIÓN: MIDIENDO LA RETROTRANSPOSICIÓN IN VITRO.	47
1.3.4. MECANISMOS DE CONTROL DE LA RETROTRANSPOSICIÓN	49
1.3.4.1. Control transcripcional	50
1.3.4.2. Control post-transcripcional	50
1.3.4.3. Control de la ribonucleopartícula	52
1.3.4.4. Control pre-integracional	54
1.3.4.5. Control post-integracional.	55
<b>1.4. REGULACIÓN EPIGENÉTICA EN MAMÍFEROS</b>	<b>55</b>
1.4.1. PRINCIPIOS GENERALES DE LA REGULACIÓN EPIGENÉTICA	55
1.4.2. ANÁLISIS DEL EPIGENOMA A ESCALA GENÓMICA	56
1.4.3. MODIFICACIONES DE HISTONAS ASOCIADAS A DISTINTOS ESTADOS DE LA CROMATINA	57
1.4.4. MODIFICACIONES DE HISTONAS MÁS RELEVANTES EN ESTA TESIS	63
1.4.4.1. Modificaciones asociadas a cromatina activa	63
1.4.4.2. Modificaciones asociadas a cromatina inactiva	65
<b>1.5. REGULACIÓN EPIGENÉTICA DE ELEMENTOS MÓVILES EN MAMÍFEROS</b>	<b>66</b>

<b>2. OBJETIVOS DE LA TESIS:</b>	<b>73</b>
<b>2.1. ESTUDIO DEL PAPEL DEL MICROPROCESADOR EN LA REGULACIÓN DEL RETROELEMENTO LINE-1.</b>	<b>73</b>
<b>2.2. ESTUDIO DE LA REGULACIÓN EPIGENÉTICA DE LAS INSERCIONES <i>DE NOVO</i> DE LINE-1 EN UN GENOMA PLURIPOTENTE Y EL IMPACTO EPIGENÉTICO DE LAS MISMAS EN EL GENOMA ADYACENTE.</b>	<b>73</b>
<b>3. MATERIALES Y MÉTODOS</b>	<b>77</b>
<b>3.1. LÍNEAS CELULARES</b>	<b>77</b>
<b>3.2. PLÁSMIDOS</b>	<b>80</b>
<b>3.3. CÉLULAS E.COLI X-GOLD ULTRACOMPETENTES PARA EL RESCATE DE LA INSERCIÓN</b>	<b>81</b>
<b>3.4. ANTIBIÓTICOS</b>	<b>82</b>
<b>3.5. PROTOCOLOS AGRUPADOS SEGÚN EXPERIMENTOS</b>	<b>83</b>
3.5.1. WESTERN BLOT	83
3.5.2. ENSAYO DE RETROTRANSPOSICIÓN EN CÉLULAS HELA	83
3.5.3. GENERACIÓN DE LÍNEAS CLONALES RESISTENTES EN PA-1	84
3.5.4. GENERACIÓN DE LÍNEAS SENSIBLES EN CÉLULAS PA-1	85
3.5.5. GENERACIÓN DE LÍNEAS CLONALES CON INSERCIONES ACTIVAS EN CÉLULAS H9	87
3.5.6. RECUPERACIÓN DE LA INSERCIÓN	89
3.5.7. INMUNOPRECIPITACIÓN DE CROMATINA SEGUIDA DE PCR CUANTITATIVA A TIEMPO REAL (CHIP-qPCR).	92
3.5.8. RECOMBINACIÓN MEDIADA POR FLIPASA	96
3.5.9. CUANTIFICACIÓN POR PCR CUANTITATIVA A TIEMPO REAL DE LAS INSERCIONES ENDÓGENAS	98
3.5.10. GENERACIÓN DE SUBCLONES EN H9-PRAM4	99
3.5.11. PCR CUANTITATIVA A TIEMPO REAL DEL GEN NEO	100
3.5.12. CUANTIFICACIÓN DEL NIVEL DE EXPRESIÓN DE GENES RECEPTORES DE INSERCIONES DE LINE-1	101
3.5.13. ENSAYO DE RESISTENCIA A NEOMICINA EN CÉLULAS PA-1 Y CÉLULAS H9.	102
3.5.14. RELACIÓN ENTRE LOS NIVELES DE EXPRESIÓN DEL GEN NEO Y LOS ENRIQUECIMIENTOS DE H3K9ME3 Y H3K4ME3	102
3.5.15. RELACIÓN ENTRE EL RECLUTAMIENTO DE H3K4ME3 Y H3K9ME3 Y LA EXPRESIÓN DEL GEN RECEPTOR	103
3.5.16. REGRESIÓN LINEAL ENTRE LA LONGITUD DE INSERCIÓN Y EL ENRIQUECIMIENTO DE H3K9ME3.	103
3.5.17. PCRS DE GENOTIPADO	103
3.5.18. ANEXO I. OLIGONUCLEÓTIDOS	- 105 -
3.5.19. ANEXO II. CONTROLES DE CALIDAD: SONICACIONES Y PCRS DE GENOTIPADO	119

3.5.20. ANEXO III: CONTROLES BIOLÓGICOS UTILIZADOS EN LOS EXPERIMENTOS DE CHIP-QPCR	129
---	-----

---

<b>RESULTADOS</b>	<b>137</b>
-------------------	------------

<b>4. RESULTADOS</b>	<b>139</b>
----------------------	------------

---

<b>4.1. ESTUDIAR EL PAPEL DEL MICROPROCESADOR EN LA REGULACIÓN DEL RETROELEMENTO LINE-1</b>	<b>139</b>
---	------------

4.1.1. EFECTO DEL MICROPROCESADOR (DROSHA/DGCR8) SOBRE LOS NIVELES DE ARNM Y PROTEÍNA DE LINE-1	139
---	-----

4.1.2. EFECTO DEL MICROPROCESADOR (DROSHA/DGCR8) SOBRE LA RETROTRANSPOSICIÓN.	140
---	-----

<b>4.2. ESTUDIO DE LA REGULACIÓN EPIGENÉTICA DE LAS INSERCIONES DE NOVO DE LINE-1 Y DEL IMPACTO EPIGENÉTICO DE LAS MISMAS EN EL GENOMA ADYACENTE</b>	<b>146</b>
--	------------

4.2.1. GENERACIÓN DE LÍNEAS CLONALES EN CÉLULAS PA1 Y H9	146
--	-----

4.2.2. REGULACIÓN EPIGENÉTICA DE LAS INSERCIONES ARTIFICIALES DE NOVO DE LINE-1	153
---	-----

4.2.2.1. Modificaciones de histonas asociadas con actividad transcripcional	154
---	-----

4.2.2.2. Modificaciones de histonas asociadas con represión transcripcional	158
---	-----

4.2.3. RELACIÓN ENTRE LA EXPRESIÓN Y EL ESTADO DE LA CROMATINA EN INSERCIONES DE NOVO DE LINE-1	162
---	-----

4.2.4. IMPACTO EPIGENÉTICO DE LAS INSERCIONES DE LINE-1 SOBRE EL GENOMA ADYACENTE	164
---	-----

4.2.4.1. Análisis de H3K4me3	166
------------------------------	-----

4.2.4.2. Análisis de la acetilación de histonas	172
---	-----

4.2.4.3. Análisis de H3K9me3	172
------------------------------	-----

4.2.4.4. Análisis de H3K27me3	178
-------------------------------	-----

4.2.5. ESTABILIDAD EPIGENÉTICA DE LAS LÍNEAS CLONALES DE PA-1	181
---	-----

4.2.6. ELIMINACIÓN DE SECUENCIAS ARTIFICIALES ASOCIADAS A LAS INSERCIONES DE NOVO DE LINE-1	181
---	-----

4.2.6.1. Análisis de modificaciones de histonas sobre las inserciones de LINE-1 flipeadas	187
---	-----

4.2.7. EFECTO DE LA DIFERENCIACIÓN EN LA REGULACIÓN E IMPACTO EPIGENÉTICO DE LAS INSERCIONES ARTIFICIALES DE NOVO DE LINE-1.	189
--	-----

4.2.8. REGULACIÓN EPIGENÉTICA DE INSERCIONES ENDÓGENAS DE NOVO DE LINE-1 EN CÉLULAS H9	193
--	-----

4.2.8.1. Determinación del porcentaje celular con las inserciones endógenas identificadas	193
---	-----

4.2.8.2. Estudio epigenetico de las inserciones endógenas de novo de LINE-1	196
---	-----

4.2.9. ANEXO IV: INSERCIONES ARTIFICIALES Y ENDÓGENAS DE LINE-1	203
---	-----

---

<b>DISCUSIÓN</b>	<b>- 233 -</b>
------------------	----------------

<b>5. DISCUSIÓN</b>	<b>235</b>
<b>5.1. REGULACIÓN DE LA MOVILIDAD DE LINE-1 POR EL MICROPROCESADOR (DROSHA/DGCR8).</b>	<b>235</b>
<b>5.2. REGULACIÓN EPIGENÉTICA DE LAS INSERCIONES <i>DE NOVO</i> DE LINE-1.</b>	<b>237</b>
<b>5.3. CAMBIOS EN EL ENRIQUECIMIENTO DE MODIFICACIONES DE HISTONAS EN LA CROMATINA ADYACENTE.</b>	<b>247</b>
<b>6. CONCLUSIONES</b>	<b>259</b>
<b>7. BIBLIOGRAFÍA</b>	<b>263</b>
<b>8. AGRADECIMIENTOS</b>	<b>287</b>
<b>9. ANEXO V: ARTICULO</b>	<b>289</b>



# INTRODUCCIÓN



## 1. Introducción

### 1.1. Secuencias repetidas en el genoma humano

La secuenciación del genoma humano desveló algo que parecía realmente una contradicción ya que sólo o menos del 5% del genoma humano se corresponde con secuencias codificantes y más o menos el 50% se corresponde con secuencias repetidas [1].

Las secuencias repetidas se pueden clasificar en 5 grandes grupos:

- Transposones
- Pseudogenes
- Repeticiones de secuencias simples  $(A)_n$ ,  $(CA)_n$  o  $(CGG)_n$
- Segmentos de ADN de entre 10-300kb duplicados, es decir, copiados de un sitio a otro.
- Bloque de repeticiones en tándem (centrómero, telómeros, brazos cortos de los cromosomas acrocéntricos y los grupos de genes ribosómicos).

De todas las clases, la más representativa son los elementos transponibles o transposones que conforman el 45% del genoma. Este 45% del genoma humano se divide en cuatro grupos de elementos móviles que se clasifican según el mecanismo de transposición:

- Elementos móviles que se movilizan a través de un intermediario de ADN mediante un mecanismo de “corta y pega”:
  - ADN transposones.
- Elementos móviles que se movilizan a través de un intermediario de ARN mediante un mecanismo de “copia y pega”.
  - Retrovirus endógenos (ERVs del inglés *Endogenous Retroviruses*).
  - *Short Interspersed elements* (SINEs).
  - *Long Interspersed elements* (LINEs).

Los elementos móviles que se movilizan a través de un intermediario de ARN se denominan retrotransposones. A su vez, éstos se clasifican en función de si codifican la maquinaria enzimática requerida para llevar a cabo su propia retrotransposición:

- Retrotransposones Autónomos:
  - Retrovirus endógenos
  - LINEs
- Retrotransposones no Autónomos:
  - SINEs

TIPO ELEMENTO MÓVIL	LONGITUD	Nº COPIAS (aproximado)	% GENOMA	ACTIVOS
<b>ADN TRANSPOSONES</b>	2-3kb	300.000	3%	No
<b>RETROTRANSPOSONES</b>				
<b>AUTÓNOMOS</b>				
<b>Retrotransposones LTR<sup>1</sup></b>				
Retrovirus Endógenos	6-10kb	>450.000	8%	1 Provirus Intacto
<b>Retrotransposones no LTR<sup>1</sup></b>				
LINE-1	6kb	516.000	17%	SI
<b>NO AUTÓNOMOS</b>				
<b>SINEs</b>				
Alu	300bp	1.090.000	11%	SI
SVA	2kb	2.700	0,2%	SI
<b>Pseudogenes Procesados</b>	variable	8.000-17.000	8000-17000 copias	SI

**Tabla 1:** Elementos móviles en el genoma humano. Adaptada de [1]–[4]. <sup>1</sup>LTR, del inglés Long Terminal Repeats.

## 1.2. Elementos móviles en el genoma humano

### 1.2.1. ADN transposones

Este tipo de ADN móvil representa un 3% del genoma humano. Estos elementos se encuentran repartidos en cuatro grandes familias que engloban unas 300.000 copias aproximadamente, siendo los más abundantes los denominados MER1-Charlie y MER2-Tigger. La estructura de la secuencia de estos elementos es la siguiente: dos repeticiones invertidas en tándem denominadas TIR (del inglés *Tandem Inverted Repeats*) que flanquean un marco abierto de lectura que codifica para una transposasa (Figura 1). El ciclo de transposición comienza con la transcripción y la posterior síntesis proteica de la transposasa. Una vez, la transposasa está de vuelta en el núcleo, ésta se une a las secuencias TIR escindiendo la copia del ADN transposón e insertándolo en una localización genómica diferente. No obstante, los ADN transposones nos son activos en el genoma humano a día de hoy. Todos las copias de ADN transposones del genoma humano pueden ser considerados fósiles. Por otro lado, la transposasa no es totalmente específica ya que reconoce las últimas 26 pares de bases de la secuencia de los ADN transposones de tal

forma que puede reconocer y movilizar tanto una copia activa como una copia que ya se encuentre inactiva. Por ello, los ADN transposones, necesitan la transmisión horizontal para perpetuarse a lo largo de la evolución de los genomas. De esta forma, y en general, su expansión en un mismo genoma está condenada a la extinción. Sin embargo, existen casos demostrados de transmisión horizontal, de una especie a otra, como es el caso del ADN transposón de mamíferos *Mariner* que engloba unas 14.000 copias en el genoma humano. En este caso, existen evidencias de que este ADN transposón ha llegado al genoma humano por transferencia horizontal desde insectos debido a su gran parecido de secuencia. Además para otros ADN transposones del genoma humano, existe también un elevada similitud con otros ADN transposones del genoma de especies muy alejadas filogenéticamente [1], [5].

### 1.2.2. Retrovirus endógenos

Los retrovirus endógenos son muy similares a retrovirus exógenos como pueden ser gamma-retrovirus, beta-retrovirus o spuma-retrovirus. Representan el 8% del genoma en humano y el 10% del genoma en ratón. Según los elementos que componen su secuencia se distinguen hasta cuatro tipo diferentes de retrovirus endógenos:

- *Retrovirus endógenos completos*: presentan largas repeticiones terminales denominadas LTRs (del inglés *Long Terminal Repeats*) que varían en longitud y que pueden ir desde 300 pares de bases hasta 1000 pares de bases acomodando éstas, las señales que regulan la transcripción. Las LTRs delimitan un espacio de 6-9 kb donde se encuentran los genes *GAG*, *POL* y *ENV*, además de las señales para la poliadenilación del ARNm, señales para el empaquetamiento de viriones y señales para la replicación (PBS, del inglés *Primer Binding Site*) así como un tracto de poli purinas (PPT, del inglés *Poly Purine Tract*). Por último, suelen estar flanqueados por duplicaciones del sitio de inserción cortas de 4 a 6 pares de bases (Figura 1). Atendiendo al grado de similitud de las regiones conservadas del gen *POL* respecto a distintas clases de retrovirus exógenos, estos elementos completos se pueden clasificar a grandes rasgos en tres clases: Clase I similares a Gamma retrovirus, Clase II similares a Beta retrovirus, Clase III similares a spuma retrovirus. En ratón y en humanos se pueden encontrar estas tres clases. Ejemplos de retrovirus endógenos pertenecientes a esta categoría son el MVL (del inglés *Murine Leukemia Virus*) en ratón y el HERV-K (del inglés *Human Endogenous Retrovirus with K-tRNA PBS*) en humano.
- *Retrovirus endógenos deficientes*. Los retrovirus que se encuadran dentro de esta categoría presentan la falta de alguna de las regiones codificantes, generalmente el gen

ENV. Dos ejemplos de estos retrovirus son el MusD (del inglés *Mus type-D related retrovirus*) e IAP (del inglés *Intacisternal A- Particle*) en ratón.

- *Retrovirus endógenos que presentan únicamente las LTRs con la señales PBS y PPT bien posicionadas junto con señales de empaquetamiento.* Entre las LTRs existe generalmente ADN no codificante. Son seguramente retrotranspuestos por la maquinaria de los retrovirus endógenos completos. Algunos ejemplos de esta categoría son el ETn (del inglés *Early Transposon*) en ratón y MaLR (del inglés *Mammalian apparent LTR Retrotransposon*) en ratón y humanos.
- *Retrovirus “Solo LTRs”.* Sólo presentan una LTR. Difíciles de identificar, sólo si se comparan con las LTRs de otros retrovirus endógenos. Su aparición se debe a la recombinación entre la dos LTRs de un retrovirus endógeno quedando únicamente una LTR. Son bastante abundantes en el genoma humano de hecho, representan un 85% del total de retrovirus endógenos y frecuentemente actúan como promotores.

En el genoma de humano el nº de retrovirus endógenos activos queda restringido a la presencia de alguna variante rara de la familia HERV-K ya que existen locis en individuos que tienen una inserción de este elemento completa o bien presentan sólo el “Solo LTR” y el alelo vacío. De hecho, recientemente, analizando 2500 genomas humanos se ha identificado un provirus de la familia HERV-K intacto en el cromosoma X de algunos individuos [6]. Por otro lado, en el genoma de ratón existen varios miembros de distintas clases que se pueden retrotransponer. Además, algunos son incluso infecciosos y, por tanto, tienen un ciclo extracelular.

En general, para simplificar solo se habla de tres clases y se obvia el hecho de la falta de algunos componentes en su secuencia y todos los ERVs se clasifican dentro de las tres clases determinadas según la conservación del gen *POL* (Clase I, II, II).

A continuación se muestra una tabla con las distintas familias de los retrovirus endógenos de ratón y humano según la clase a que pertenecen y según si son todavía activos o se sospecha que pueden serlo. Existen distintas formas de nombrar a los retrovirus endógenos basadas en cuando se descubrieron. En primer lugar, se nombraba en función del virus exógeno más cercano filogenéticamente mientras que en la actualidad se nombran en función del organismo donde fue primero descubierto y la letra que representa el ARN transferente que utilizan como cebador para su reversotranscripción tal como HERV-K que significa, del inglés, *Human Endogenous Retrovirus with K-tRNA PBS*.

Retrovirus endógenos						
Ratón					Humano	
Clase	Autónomo	Componentes	Ciclo <sup>1</sup>	Movilidad	Clase	Movilidad
<b>Clase I</b>					<b>Clase I</b>	
<b>Subclase I</b>					HERV-H	No
MuRRs	No	Mutados Gag, Pol, Env		No	HERV-W	No
MuERV-C	No	Mutados Gag, Pol, Env		No	HERV-E	No
GLN	Sí	Gag, Pol, Env	Extra	Sí		
MmERV	Sí	Gag, Pol, Env		1 provirus intacto		
VL30	No	Proviene de deleciones de MmERV				
McERV	Sí	Gag, Pol, Env		3 provirus mutados		
<b>Subclase II</b>						
MuLV	Sí	Gag, Pol, Env	Extra	Sí		
<b>Clase II</b>					<b>Clase II</b>	
MMTV	Sí	Gag, Pol, Env		2 o 3 alelos completos activos	HERV-K	1 provirus intacto
MusD	Sí	Gag, Pol	Intra	Sí		
ETn	No	Derivados de MusD. ADN no codificante	Intra	Sí		
IAP	Sí	Gag, Pol	Intra	Sí		
<b>Clase III</b>					<b>Clase III</b>	
MuERV-L	Sí	Gag, Pol	Intra	Sí	HERV-L	No
MaLR	No	ADN no codificante	Intra	Sí	MaLR	No

**Tabla 2:** Clases y familias de retrovirus endógenos en los genomas de ratón y humano con algunas de sus características.<sup>1</sup>Extra o intra es la abreviación para extracelular e intracelular respectivamente. Los nombres de las familias son los siguientes: MuRR, murine retroviral-related sequences; MuERV-C, Murine Endogenous Retrovirus with C-tRNA PBS; MmERV, Mus musculus Endogenous Retrovirus; VL30, Virus Like 30; McERV, Mus caroli Endogenous Retrovirus; MuLV, Murine Leukemia Virus; MMTV, Mouse Mammary Tumor Virus; MusD, Mus type-D related retrovirus; ETn, Early Transposon; IAP, Intracisternal A-type Particle; MuERV-L, Murine Endogenous Retrovirus with L-tRNA PBS, MaLR, Mammalian apparent LTR Retrotransposon. HERV-H, Human Endogenous Retrovirus with H-tRNA PBS; HERV-W, Human Endogenous Retrovirus with W-tRNA PBS, HERV-E, Human Endogenous Retrovirus with E-tRNA PBS; HERV-K, Human Endogenous Retrovirus with K-tRNA PBS; HERV-L, Human Endogenous Retrovirus with L-tRNA PBS. Adaptada de [3], [7].

○ *Adquisición y amplificación de los retrovirus endógenos en el genoma murino y humano.*

Se postula que la adquisición de estos elementos móviles en el genoma humano tuvo lugar a través de la infección de células germinales por retrovirus exógenos hace aproximadamente 20-40 millones de años. Por tanto, una vez que existe un provirus en el genoma hospedador, éste puede desaparecer, fijarse o incluso aumentar su número de copias en la población en función de su impacto sobre la capacidad reproductiva del individuo. Este tipo de adquisición de retrovirus endógenos en el genoma fue demostrado con experimentos en ratones en la década de los 70 y de los 80. El trasplante de oocitos no infectados en hembras infectadas, reveló la presencia de retrovirus endógenos en el genoma de la descendencia lo que demostraba una infección de dichas células [3].

De este modo, el proceso de adquisición de un retrovirus endógeno comienza por la endogenización de un virus exógeno convirtiéndose en un retrovirus endógeno. El número de copias de este elemento en el genoma hospedador puede aumentar por una nueva infección de las células germinales por un virus exógeno del mismo grupo, por reinfección de células germinales por el propio retrovirus endógeno o bien por retrotransposición mediante la adquisición de un ciclo intracelular. Los retrovirus endógenos activos que tengan los genes, *GAG*, *POL* y *ENV* intactos pueden reinfectar células germinales mediante la formación de partículas virales exógenas. De hecho, para algunos retrovirus endógenos de ratón se han detectado tales partículas virales y hoy día, es conocido, que estos elementos siguen amplificándose en el genoma de ratón vía reinfección de células germinales debido a la actividad de aquellos ya insertados en el genoma. Por otro lado, en humano se podría especular con que este fenómeno todavía podría tener lugar debido a la existencia de un provirus intacto de la familia HERV-K [6]. Por otro lado, existen algunos tipos de retrovirus endógenos de ratón que han aumentado su fitness replicativo bien adquiriendo un ciclo de replicación intracelular (IAP) evitando la respuesta del hospedador a la infección o bien movilizándose en *trans*, es decir, utilizando la maquinaria de otros retrovirus endógenos autónomos endógenos o bien la maquinaria de retrovirus exógenos (VL-30, MaLR).

Otro fenómeno asociado a los retrovirus endógenos es la recombinación a nivel de ARN. Retrovirus endógenos no activos pueden convertirse en activos bien mediante la recombinación de su ARN con el ARN de un retrovirus endógeno activo o bien mediante la recombinación con un retrovirus exógeno. Asimismo, la recombinación entre dos retrovirus endógenos parcialmente inactivos puede llevar a la formación de un retrovirus endógeno activo que podría ser incluso infeccioso. El fenómeno de la recombinación está



muy bien documentado para retrovirus endógenos de Clase I de ratón. Además, en el laboratorio se ha conseguido resucitar un retrovirus endógeno de humano de la familia HERV-K con muy baja capacidad infectiva mediante recombinación de ARN.

Revisado en [1], [3], [7]–[9].

### 1.2.3. LINE-1

Los elementos tipo LINE-1 son retrotransposones autónomos. Es el único tipo de ADN móvil que se moviliza actualmente en el genoma humano de forma autónoma. Constituye el 17% del genoma humano. De hecho, existen aproximadamente 500.000 copias en el genoma. No obstante, su actividad también ha contribuido a la formación de otro 11% del genoma humano ya que ha movilizó en *trans* otros retrotransposones no LTR como es el caso de los retrotransposones no autónomos conocidos como SINEs (descritos en apartado 1.2.4). Existen datos que muestran que los elementos LINE-1 son activos a día de hoy en el genoma humano ya que con frecuencia desencadenan la aparición de una enfermedad genética por una nueva inserción que desactiva un gen. De hecho, en los últimos 30 años se han identificado numerosas inserciones de LINE-1 causantes de enfermedad tales como la hemofilia A, la distrofia muscular y la granulomatosis crónica [10]. El dato que confirma la movilización actual de este elemento es la ausencia de la inserción causante de la enfermedad en los padres del paciente como se demostró para un paciente de granulomatosis crónica [11]. Además, las inserciones progenitoras de aquellas inserciones responsables de la generación de enfermedades son activas en un ensayo de retrotransposición *in vitro* [11], [12]. Por otra parte, se ha conseguido aislar inserciones presentes en individuos de un determinado origen étnico que no están presentes en otros individuos de un origen étnico diferente [13]. Asimismo, estas inserciones todavía mantienen su capacidad para retrotransponerse en un ensayo de movilidad *in vitro* [12], [13]. A día de hoy, también se han identificado inserciones de *novó* de LINE-1 en tumores de diversa índole las cuales estaban ausentes en el tejido sano [14]–[16]. Por tanto, todas estas evidencias conjuntamente indican que los elementos LINE-1 son actualmente activos en el genoma humano.

Una copia activa de LINE-1 tiene una longitud de 6000 pares de bases, donde se distinguen tres grandes fracciones: una primera parte de esta secuencia que corresponde con una región 5' no traducida denominada 5'UTR (del inglés 5' Untranslated Region) con actividad promotora de unas 900 pares de bases, una segunda parte llamada marco abierto de lectura 1 denominado ORF1 (del inglés *Open Reading Frame 1*) que codifica para una proteína de unos 40kd de nombre ORF1p (del inglés *Open Reading Frame 1*

*protein*) y una tercera parte llamada marco abierto de lectura 2 denominado ORF2 (del inglés *Open Reading Frame 2*) que codifica para una proteína de unos 150kd de nombre ORF2p (del inglés *Open Reading Frame 2 protein*), revisado en [2]. ORF1 codifica para una proteína que tiene una actividad demostrada de chaperona de unión a ácidos nucleicos [17] mientras que ORF2 codifica para una proteína que tiene dos actividades enzimáticas tales como endonucleasa [18] y reversotranscriptasa [19]. ORF1p y ORF2p son indispensables para la retrotransposición [12]. Asimismo, en el extremo 3' de la secuencia de LINE-1, existe una corta región sin traducir denominada 3'UTR (del inglés *3' Untranslated Region*) que es prescindible para la retrotransposición [12]. Esta región sin traducir en 3' termina en una serie de adeninas que conforman la denominada cola de poli A la cual es necesaria para la retrotransposición [20] (Figura 1). (La estructura de LINE-1 se detalla más adelante en el apartado 1.3.1.)

○ *Ciclo de retrotransposición.*

El ciclo de retrotransposición comienza con la transcripción desde el promotor de la región 5' UTR que da lugar a un ARN mensajero bicistrónico. Este ARNm sale al núcleo donde es traducido, generándose en el citoplasma ORF1p y ORF2p que se unen a su propio ARNm en *cis* y da lugar a la formación de una ribonucleopartícula que es un intermediario de la retrotransposición. La ribonucleopartícula, una vez formada, penetra en el núcleo donde ORF2p reconoce una secuencia consenso y mediante un mecanismo denominado TPRT (del inglés *Target Primed Reverse Transcription*) genera una nueva inserción de LINE-1 en el genoma, en la mayoría de los casos flanqueada por duplicaciones del sitio de inserción de una longitud variable (2-20nt). Además, la mayoría de las nuevas inserciones de LINE-1 se resuelven con una truncación en 5', perdiendo parte de su secuencia [2]. (El ciclo de retrotransposición se detalla más adelante en el apartado 1.3.2.)

#### 1.2.4. SINEs

Los SINEs son un tipo de retrotransposón no autónomo específico de primates. Para su movilización necesitan la maquinaria enzimática codificada por el retrotransposón autónomo LINE-1. Junto con LINE-1 constituyen los dos elementos transponibles que son activos a día de hoy en el genoma humano. Dentro de los SINEs se distinguen dos tipos de elementos: los elementos Alu y los elementos SVA. En primer lugar, los elementos Alu son los más abundantes dentro de este tipo y además es el elemento móvil con un mayor número de copias en el genoma ya que existen 1.000.000 de copias del mismo. Estos elementos Alu derivan ancestralmente del ARN 7SL que forma parte del ribosoma. En segundo lugar, se encuentran los elementos SVA, mucho menos abundantes en el genoma

ya que sólo existen 2.700 copias. Éstos se caracterizan por ser una composición de porciones de otros retrotransposones y son exclusivos de primates superiores.

#### 1.2.4.1. Elementos Alu

La longitud de un elemento Alu es 280 pares de bases que se reparten de la siguiente forma: dos monómeros derivados del ARN 7SL separados por una región central corta pero rica en adeninas. El monómero de la izquierda contiene secuencias cajas A y B conservadas para que el Alu sea transcrito por la polimerasa III, probablemente al principio del mismo. Por otro lado, al final del monómero de la derecha también se encuentra una zona rica en adeninas. Este elemento, no contiene secuencias terminadoras de la transcripción, con lo que suele usar terminadores (regiones ricas en timinas, normalmente más de 4 consecutivas) situados aguas abajo en el genoma adyacente generando transducciones en 3' las cuales consisten en arrastrar secuencias genómicas situadas aguas abajo de la inserción e insertarlas en una nueva localización genómica (Figura 1) (detallado en apartado 1.3.4.4.) [21].

##### o *Ciclo de retrotransposición*

Los elementos Alu dependen íntegramente de los elementos LINE-1 para su movilización en el genoma humano, pues su movilidad es dependiente de ORF2p. Probablemente, el hecho de utilizar únicamente una de las proteínas codificadas por el elemento LINE-1, ha favorecido una mayor efectividad en la colonización del genoma humano llegando hasta el millón de copias. De esta forma, en tejidos donde los elementos LINE-1 no son capaces de retrotransponerse, quizás los Alus puedan hacerlo y además pueden nutrirse de las ORF2p sintetizadas desde el ARNm de LINEs mutados en ORF1p que no sean activos.

El ciclo de retrotransposición de los elementos ALU comienza con la transcripción desde su promotor por la polimerasa III generándose un ARN el cual presumiblemente se ensambla en un complejo con tres proteínas: el heterodímero SRP9/14 (del inglés Signal Recognition Particle 9/14) y la proteína unidora de Poli A (PABP del inglés Poli A Binding Protein). Además existe una tercera proteína que reconoce la estructura de ARN. Este complejo parece formarse en el nucléolo y favorece la interacción con el ribosoma donde se está produciendo la síntesis de las proteínas provenientes del elemento LINE-1. Es allí donde supuestamente, el elemento Alu secuestra ORF2p. Este complejo ARN/proteínas migra al núcleo para generar una nueva copia de Alu mediante TPRT (más detallado en el apartado 1.3.2.) mediado por ORF2p de LINE-1. Por ello, las inserciones canónicas de ALU se encuentran flanqueadas por duplicaciones del sitio de inserción, se integran en sitios del

genoma con secuencias consenso reconocidas por el dominio endonucleasa de ORF2p y, aunque son raros, los elementos Alu también se pueden encontrar 5' truncados.

○ *Particularidades de los elementos Alu*

Al igual que ocurre con los elementos LINE-1, no todas las copias de Alu son capaces de retrotransponerse, de hecho, solo algunas inserciones tienen todavía hoy día la capacidad de llevar a cabo este proceso dentro de una subfamilia. Esto se debe a una serie de características de las inserciones de Alu, que se detallan a continuación:

- El promotor de los elementos Alu es bastante débil, y está bastante influenciado por el contexto genómico/epigenético donde la nueva inserción ha aterrizado.
- En cada ronda de movilización, los elementos Alu pueden acumular mutaciones en la cola de poli A generando una secuencia que parece un microsatélite y que afectan a su movilidad negativamente.
- Además, en cada ronda de movilización, la cola de poli A se acorta lo cual también afecta negativamente a su movilización.
- Por otro lado, la secuencia del elemento Alu acumula mutaciones lo cual hace que se afecten partes de la secuencia del ARN que pueden ser importantes para la adquisición de la estructura secundaria del ARN o bien para la interacción con proteínas importantes para la movilización.
- Por último, también la longitud de la secuencia adyacente genómica que arrastra en su extremo 3' (dependiente de la localización del terminador de la transcripción en el genoma adyacente) puede afectar negativamente a la retrotransposición.

Existen evidencias de que los elementos Alu activos hoy día en el genoma humano pertenecen a la subfamilia Y [22]. De hecho, existe un elevado número de inserciones en la población humana de esta subfamilia que son polimórficos, lo que quiere decir, que se están presentes en el genoma de un individuo y no en el genoma de otros. Incluso algunos de éstos son dimórficos, es decir, que están en el genoma de algunas razas y no están presentes en el genoma de otras. De hecho, todas las inserciones causantes de enfermedad, excepto una, donde se pudo identificar el origen filogenético del elemento Alu, pertenecen a la subfamilia Y.

(Revisado en [10], [23]–[25].

#### *1.2.4.2. Elementos SVAs*

El otro tipo de elemento SINE, también activo en el genoma humano, es un retrotransposón formado por porciones de otros retrotransposones. La estructura típica de

un elemento SVA (del inglés *SINE-VNTR-Alu*) es la siguiente, presenta en su extremo 5' una repetición del hexámero CCCTCT seguido de dos elementos ALU antisentido, seguido de un número variable de secuencias repetidas en tándem ricas en GC, y por último, una secuencia que tiene homología con el gen *ENV* y una secuencia LTR procedentes ambas de un retrovirus endógeno ancestral, seguido todo ello de una señal de poliadenilación canónica AATAAA (Figura 1). Las inserciones de SVA tienen más o menos una longitud de 2.000 pares de bases aunque pueden oscilar entre 700 y 4.000 pares de bases.

#### ○ *Ciclo de retrotransposición*

Existen varias líneas de evidencia que indican que es la polimerasa II probablemente la encargada de transcribir los elementos SVA. Dicha transcripción puede comenzar aguas arriba de la inserción o bien en un promotor interno de la misma. El análisis de algunas inserciones retrocompetentes indica que el inicio de la transcripción ha tenido lugar aguas arriba de la inserción usando un promotor vecino, constituyendo este tipo de inserción alrededor del 8% de las inserciones de SVA. La transcripción puede terminar en varios lugares de la inserción, bien en la señal de poliadenilación canónica, en la señal interna en la zona de homología con retrovirus endógeno ancestral y, por otro lado, puede saltarse estas señales y terminar la transcripción en una señal de poliadenilación situada en el genoma adyacente dando lugar a transducciones en 3' (detallado en apartado 1.3.4.4). Cualquiera de las formas de terminación de la transcripción da lugar a un ARN retrocompetente. Posteriormente, el ARN tiene que entrar en contacto con la ribonucleopartícula de LINE-1 lo cual se desconoce si produce a nivel del ribosoma o bien a nivel del nucléolo. Una vez incorporado a la ribonucleopartícula, el ARN de SVA se reverso transcribe por ORF2p y se integra en su nueva localización genómica [26]. Esta nueva inserción de SVA presenta las características típicas de una inserción de LINE-1 ya que suele presentar duplicaciones del sitio de inserción, una cola de poli A, integración en la secuencia consenso reconocida por ORF2p y además presenta truncaciones en 5' aunque son menos abundantes que las truncaciones en 5' del retroelemento LINE-1 [26].

Los SVA son todavía activos en el genoma humano, de hecho, algunas de las inserciones nuevas que se generan dan lugar a enfermedades como por ejemplo una inserción *de novo* en el gen *BTK* que es responsable de la agammaglobulinemia [27], [28] o bien la inserción *de novo* en el gen de la  $\alpha$ -espectrina que genera la eliptocitosis hereditaria [29]. Además existen otras 10 inserciones asociadas a enfermedad en humanos.

(Revisado en [10], [30]).

**ADN TRANSPOSONES**



**RETROTRANSPOSONES**

**AUTÓNOMOS**

**Retrotransposones no LTR**

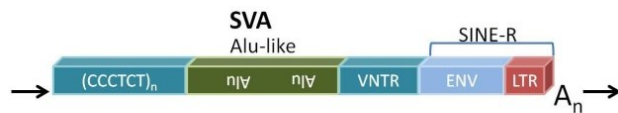
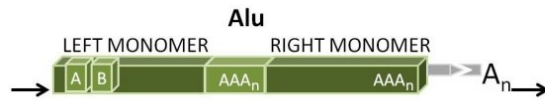


**Retrotransposones LTR**



**NO AUTÓNOMOS**

**SINES**



**Pseudogenes**

**Procesados**



<sup>1</sup>**Figura 1: Tipos de ADN móvil en el genoma humano.** Se muestra la estructura canónica de cada uno de los tipos de ADN móvil presentes en el genoma humano. Las abreviaciones usadas son las siguientes: TIR, Tandem Inverted Repeats; ERVs (endogenous retroviruses); HERV-K (Human Endogenous Retrovirus-K); LTR (Long Terminal Repeats); GAG, group specific antigen; POL, polymerase; ENV, envelope protein; ORF1, open reading frame 1; ORF2, open reading frame 2; 5' UTR (5' Untranslated Region); CC, coiled coil; RRM, RNA Recognition Motif; CTC, Carboxyl-Terminal Domain; EN, Endonuclease; RT, Reverse Transcriptase; C, Cysteine-rich domain; 3', región 3' sin traducir; A y B, caja A y B del promotor de Alu reconocidas por la Polimerasa III; AAA<sub>n</sub>, zona rica en adeninas VNTR, Variable Number of Tandem Repeats; SINE-R, dominio derivado de un HERV-K. A<sub>n</sub> significa cola de poli A; 5',3' hacen referencia a regiones sin traducir; Ex, Exón. Las flechas negras hacen referencia a las duplicaciones del sitio de inserción propias del mecanismo de integración mediado por ORF2p. Adaptada de [2], [4].

<sup>1</sup> Para evitar confusiones con la traducción, en esta figura se ha utilizado la nomenclatura en inglés para cada una de las partes de los distintos elementos móviles.

### 1.2.5. Pseudogenes procesados.

Se puede considerar otro tipo de retrotransposón no autónomo a los pseudogenes procesados, es decir, ARNm celulares procesados carentes de secuencias reguladoras que pueden ser retrotranspuestos por la maquinaria enzimática del retroelemento LINE-1 dando lugar a una nueva “inserción” que presenta las características propias de una inserción de LINE-1 (Figura 1). Existen entre unos 8.000-17.000 pseudogenes procesados en el genoma humano, siendo los más comunes ARNm que codifican proteínas del ribosoma o ARNm de genes de expresión constitutiva [10]. Además, recientemente, se ha demostrado que la aparición de un caso granulomatosis crónica está ligada a la presencia de una inserción mutagénica de un pseudogen procesado procedente del gen *TMF1* dentro del gen *CYBB* [31]. Remarcar, que la inserción que propició la enfermedad estaba presente en el paciente, en la madre pero no lo en los abuelos, por lo que la inserción tuvo lugar recientemente y además muestra como la retrotransposición de ARNm celulares es un proceso que actualmente sigue ocurriendo en el genoma humano [31].

### 1.3. Un retrotransposón autónomo activo: LINE-1.

Como se mencionó más arriba constituye un 17% del genoma repartido en 500.000 copias. Se estima que el número de copias de LINE-1 que retienen la capacidad de movilizarse se encuentra entre 80 y 150 copias [13] ya que la mayoría se encuentra inactivas debido a truncaciones en 5', inversiones y mutaciones puntuales en los dos marcos abiertos de lectura [2]. Es de destacar, que la capacidad de retrotransposición de éstas recae sobre unas pocas inserciones de LINE-1 [32] las cuales pertenecen a la subfamilia L1Hs (del inglés *Human specific*), de la cual tan solo un pequeño grupo denominado Ta (del inglés *Transcribed Active*) engloba a la mayoría de todas las copias con capacidad para retrotransponerse. Miembros de este grupo se diferencian por un trinucleótido AGA en la posición 5930-5932 y una G en la posición 6015 [33]. Estas copias que son capaces de retrotransponerse se denominan retrocompetentes.

Por otro lado, los elementos LINE también han contribuido a la formación de una gran parte del genoma de ratón ya que ocupa un 18%. Además, es importante resaltar que existen unas 3000 copias activas en el genoma de ratón [4]. La gran diferencia con los LINEs humano radica en la estructura de la 5'UTR ya que presenta una serie de repeticiones y además presenta una secuencia espaciadora entre la región 5'UTR y ORF1. Los elementos activos se engloban en tres subfamilias (A, T<sub>f</sub>, G<sub>f</sub>) las cuales se clasifican en función de la secuencia de la 5'UTR [4]. Por lo demás, los elementos LINE-1 humanos y de ratón son muy iguales y, por tanto, todo lo que se explica a continuación es aplicable también a los LINEs de ratón.

### 1.3.1. Estructura de las copias retrocompetentes de LINE-1.

#### ○ *Región 5' no traducida.*

La región 5' no traducida (5'UTR) es una región rica en GC [34]. De hecho, existen islas CpG a lo largo de dicha secuencia que hacen posible que los retroelementos LINE-1 puedan ser controlados vía metilación del promotor [35], [36]. Esta región de los elementos LINE-1 presenta actividad promotora sentido ya que es capaz de dirigir la expresión de un gen reportero fusionado en su extremo 3' [37]. Por otro lado, los primeros 100-150 nucleótidos son importantes para la actividad transcripcional pues se demostró que la delección de estos provocaba una pérdida importante de su actividad transcripcional [37], [38]. Además esta secuencia presenta motivos de unión a factores de transcripción tales como YY-1, Runx3, Sp1 y proteínas relacionadas con SRY [37], [39]–[42] que regulan su transcripción. Por otro lado, la región 5' no traducida presenta también un promotor antisentido que se inicia entre los nucleótidos 400 y 600 [43]. Este promotor es capaz de dirigir la transcripción en sentido contrario generando transcritos que contienen una porción de la 5'UTR y una segunda porción procedente de la secuencia del ADN genómico inmediatamente situado aguas arriba [44], [45]. Según ensayos mediante plásmidos en células en cultivo, ambos promotores son activos en células pluripotentes aunque el promotor sentido es más potente que el promotor antisentido [44]. La presencia de ambos promotores en la región 5' UTR da lugar a la existencia de la teoría de que el funcionamiento de ambos favorecería la aparición de especies de ARN de doble cadena con una región común desde donde ser podrían originar siRNAs (del inglés small interfering RNAs) vía DICER que podrían estar relacionados con el control de la movilidad del propio retroelemento [46], [47].

Sorprendentemente, y muy recientemente, se ha demostrado que la región 5'UTR también alberga un marco abierto de lectura denominado ORF0 (del inglés, Open Reading Frame 0). Este ORF0 coincide con la región 452-236 de la región 5'UTR en la orientación antisentido y da lugar a transcritos de fusión mayoritarios en células madre inducidas pluripotentes respecto a células diferenciadas. Además, estos ARNs se localizan en el citoplasma y se asocian con el ribosoma. Este ORF0 también puede dar lugar a la formación de proteínas de fusión con exones vecinos debido a la presencia de dos sitios donadores de splicing. Cabe destacar que la sobreexpresión de ORF0 incrementa ligeramente la retrotransposición, aunque el mecanismo de acción no se conoce [48].



○ *Marco abierto de lectura 1*

Esta región se encuentra justo a continuación de la región 5' no traducida. Dicha región codifica para ORF1p la cual es una proteína de aproximadamente 40kd con actividad chaperona de ácidos nucleicos y de unión a ARN [17]. En la región N-terminal, necesaria para la interacción de la proteína con otras proteínas, existe un dominio de cremallera de leucina (Leucine Zipper o LZ) [49], [50]. Seguidamente, se presenta un motivo de unión a RNA denominado RRM (del inglés RNA Recognition Motif) [51], el cual va seguido de un dominio C-terminal (CTD, del inglés Carboxyl-Terminal Domain). Es necesaria la actividad conjunta de ambos para su unión al RNA[51]. En la traducción del mensajero de LINE-1 se producen múltiples copias de ORF1, que forman homotrímeros mediante la interacción cola con cola (*coiled-coiled* o C-C) en su parte N-terminal; estos trímeros se unen al RNA mediante sus dominios RRM y CTD, formando así parte de la ribonucleopartícula completa de LINE-1 [52]. Por otra parte, esta proteína es totalmente imprescindible para la retrotransposición ya que mutaciones en aminoácidos muy conservados en los dominios RRM y CTD llevan a una pérdida total de la movilidad del retroelemento LINE-1 [12], [51].

○ *Marco Abierto de Lectura 2*

El segundo marco abierto de lectura codifica para ORF2p que es una proteína multidominio de aproximadamente 150kd. En su extremo más amino terminal presenta un dominio endonucleasa (EN) [18] mientras que aguas abajo de éste, se encuentra un dominio reverso transcriptasa (RT) [19]. Ambas actividades son cruciales para la retrotransposición [12]. La actividad endonucleasa de ORF2p es la encargada de hacer una mella en el ADN en una secuencia consenso 5'-TTTT/AA-3' liberando un extremo 3'OH [18]. ORF2p es capaz de llevar a cabo la reversotranscripción independiente del estado de ORF1p y del dominio EN, es decir, mutaciones en ambos dominios no afectan a la capacidad de reversotranscripción [53]. A diferencia de otras reversotranscriptasas como MMLV RT (del inglés Moloney Murine Leukemia Virus Reverse Transcriptase), ORF2p no necesita una homología perfecta del ADN molde con el ADN cebador [53], [54]. Además, ORF2 presenta otro dominio situado en su extremo más carboxilo, el dominio C, que consiste en un dominio rico en cisteínas cuya función es desconocida, sin embargo, su existencia es indispensable para la formación de la ribonucleopartícula así como para que exista retrotransposición [12], [55]

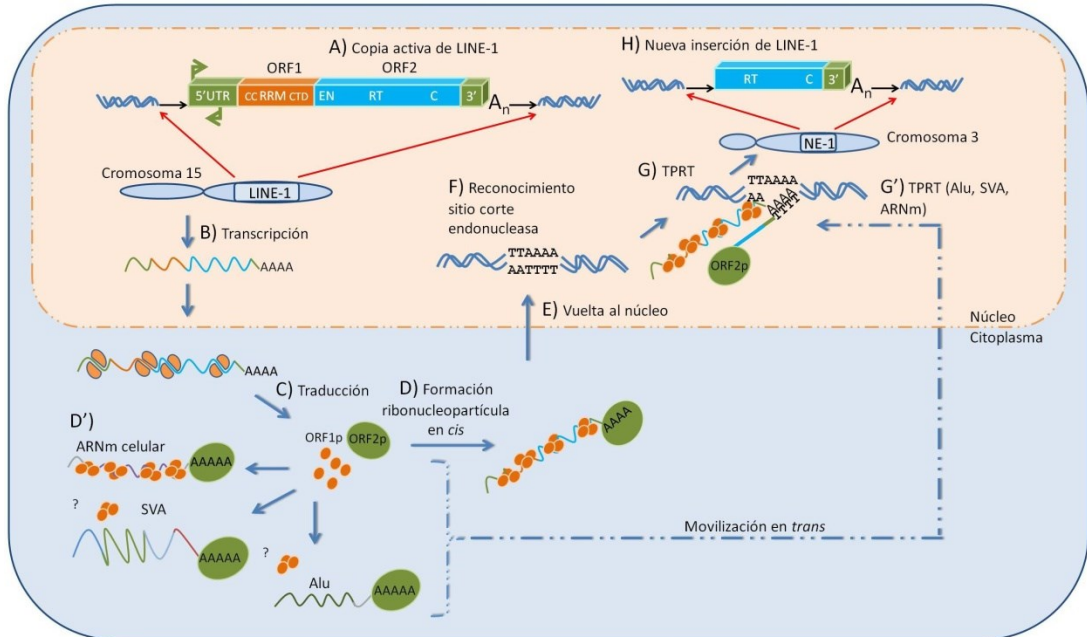
○ *Región 3' no traducida*

Esta región tiene una longitud de 206 nucleótidos y contiene una zona rica en purinas que podría formar cuádruplex de guanina [56]. Sin embargo, esta región no parece tener importancia para la retrotransposición porque la delección de 145 pares de bases de esta región no afecta a la retrotransposición [12]. Esta región contiene una señal de poliadenilación relativamente débil que alguna vez es saltada por la ARN polimerasa II, utilizando señales de poliadenilación canónicas que se encuentran aguas abajo de la inserción en el genoma contiguo. De esta forma, se produce la movilización de secuencias situadas aguas abajo de la inserción en un fenómeno denominado transducción en 3' (detallado en apartado 1.3.4.4.) [21]. En cualquier caso, aparecen una serie de adeninas que conforman la denominada cola de poli A acompañando a la inserción de LINE-1 en su extremo 3' terminal. Esta cola de poli A es totalmente crucial para la retrotransposición del retroelemento LINE-1 [20].

1.3.2. Ciclo de retrotransposición

El proceso de generación de una nueva copia del retroelemento LINE-1 se conoce como ciclo de retrotransposición. Este ciclo se puede dividir en varias fases para que su estudio sea más sencillo: una primera fase nuclear seguida de una fase citoplasmática y finalmente una segunda fase nuclear.

- *Primera fase nuclear.* El ciclo de retrotransposición comienza con la transcripción desde el promotor sentido ubicado en la región 5' UTR de una copia retrocompetente de LINE-1 situada en algún lugar del genoma. El ARN mensajero resultante es un ARNm bicistrónico que viaja fuera del núcleo para ser traducido (Figura 2A, B)
- *Fase citoplasmática.* El ARN mensajero de LINE-1 se traduce en el citoplasma por un mecanismo no convencional de terminación y reiniciación de la traducción (Figura 2C) [57]. Una vez traducido, este ARNm es unido por las moléculas de ORF1p y ORF2p generadas desde su propia traducción, en un fenómeno conocido como preferencia en *cis* [58], formando la ribonucleopartícula (Figura 2D). Ésta es un intermediario de la retrotransposición ya que presenta actividad positiva en un ensayo denominado LEAP) (del inglés *L1 Element Amplification Protocol*) [53], [55]. Estas ribonucleopartículas se han aislado exitosamente en células de carcinoma embrionario de ratón, células madre embrionarias humanas y células de carcinoma embrionario humano [59]–[61]. Por otro lado, no es descartable la unión de otras proteínas celulares así como de otros ARNs celulares [62]–[64]. Además, estas ribonucleopartículas parecen localizarse en foci citoplasmáticos ya que ORF1p, ORF2p y el ARN mensajero de LINE-1 son detectados en el citoplasma formando agregados [55].



**Figura 2: Ciclo de retrotransposición del retroelemento LINE-1.** A),B) El promotor de una copia completa de LINE-1 situada en algún lugar del genoma, dirige la transcripción del retroelemento LINE-1. El ARNm resultante sale del citoplasma. C) El ARNm de LINE-1 es traducido en el citoplasma. D) Los proteínas resultantes de la traducción (ORF1p y ORF2p) se ensamblan en torno al mismo ARNm del que fueron sintetizadas en un fenómeno conocido como preferencia en cis dando lugar a la ribonucleopartícula D'. Una vez las proteínas han sido sintetizadas, éstas pueden ser secuestradas por los retrotransposones no autónomos para ser retrotranspuestos. Por parte de Alu y SVA no está clara la necesidad de ORF1p para completar su ciclo de retrotransposición. E) Una vez ensamblada, la ribonucleopartícula vuelve al núcleo. F) La ribonucleopartícula reconoce una secuencia consenso a través del dominio endonucleasa de ORF2p. G). Generación de una nueva copia del retroelemento LINE-1 mediante el proceso de TPRT (del inglés Target Primed Reverse Transcription) comenzando con el corte en la secuencia consenso en la hebra de abajo liberando un extremo 3'OH que sirve como cebador para la reverso transcripción. G'). Siguiendo este mismo mecanismo de TPRT se generan las inserciones de los retrotransposones no autónomos. H). La consecuencia final de este proceso es una nueva inserción de LINE-1 en un lugar del genoma diferente. En la mayoría de los casos, estas inserciones están truncadas en 5'.

- Segunda fase nuclear: Una vez que la ribonucleopartícula se ha formado, entonces ésta migra al núcleo, sin embargo, se desconocen los detalles de la vuelta al núcleo (Figura 2E). De hecho, existe cierta controversia en torno a la necesidad de que exista

división celular, es decir, que se desintegre la membrana nuclear para que tenga lugar la entrada de la ribonucleopartícula al núcleo [4]. Una vez en el núcleo, se produce una nueva inserción de LINE-1 a través del mecanismo TPRT. Basado en ensayos *in vitro* con un retrotransposón de tipo no-LTR (R2 en *Bombix nori*) similar a LINE-1 [65] así como con ensayos *in vitro* con LINE-1 [66] se propone el siguiente modelo. En primer lugar, la actividad endonucleasa de ORF2p reconoce un sitio de corte consenso (5'-TTTT/AA-3') en la hebra de abajo (Figura 2F)[18]. El corte en dicho lugar libera un extremo 3'OH que sirve de cebador para la actividad reversotranscriptasa de ORF2p que utiliza el ARNm del retroelemento como molde para sintetizar la hebra antisentido del mismo (Figura 2G). Posteriormente, se produce el corte en la hebra de arriba aguas abajo del corte que se produjo en la hebra de abajo (3'-5'). Una vez producido el segundo corte, se debe producir la síntesis de la segunda hebra del retroelemento LINE-1, que da lugar a la aparición de una nueva inserción de LINE-1. Gran parte de los pasos que llevan a la síntesis de la segunda hebra son desconocidos. El resultado final es una inserción nueva de LINE-1 flanqueada por duplicaciones del sitio de inserción [67], es decir, secuencias de 2-20 nucleótidos iguales a ambos lados de la inserción que son consecuencia del mecanismo de inserción. La nueva inserción de LINE-1 puede estar truncada en 5' como ocurre en la mayoría de los casos o bien puede presentar la secuencia completa del mismo (Figura 2H) [4], [34].

La síntesis de la segunda hebra podría estar mediada por factores celulares implicados en la reparación del ADN [65]. Esto podría explicar por qué un gran porcentaje de las inserciones de LINE-1 tienen un extremo 3' intacto mientras que se encuentran truncadas en 5'. De hecho, se especula con que la truncación en 5' se puede producir porque existe una retrotransposición abortiva donde la reversotranscriptasa se desliga de la hebra antisentido de LINE-1 durante el corte de la hebra de arriba. Este proceso puede ser impulsado por los mecanismos de reparación de ADN que pueden reconocer el intermediario ADN:ARN generado durante la síntesis de la cadena antisentido del retroelemento LINE-1 [4]. En línea con esta hipótesis, existen proteínas pertenecientes a vías de reparación del ADN que regulan la retrotransposición [68]–[70].

### 1.3.1. Encuadre temporal de la retrotransposición

#### ○ *Retrotransposición durante el desarrollo embrionario temprano*

Para asegurarse el éxito evolutivo, el retroelemento LINE-1 deber generar copias de sí mismo que puedan ser transmitidas a la siguiente generación. Por tanto, está claro que

para que esto pueda ocurrir la retrotransposición tiene que producirse bien en células germinales o bien en etapas tempranas del desarrollo. Células humanas embrionarias, células de carcinoma embrionario humano así como células madre pluripotentes inducidas sobreexpresan los elementos LINE-1 ya que se detectan altos niveles del ARNm de LINE-1 y de ORF1p comparado con células diferenciadas. Además las ribonucleopartículas de LINE-1 pueden ser aisladas en estos tres tipos celulares [36], [44], [59], [60], [71]. Según estos datos, la expresión de estos elementos LINE-1 parece estar ligada a la pluripotencia mientras que desaparece con la diferenciación. De hecho, los niveles de ARNm de LINE-1 aumentan a medida que se produce la reprogramación de células humanas diferenciadas de linaje eritroide hacia células madre pluripotentes inducidas [72]. Por otro lado, dichos tipos celulares como son células madre embrionarias humanas, células madre pluripotentes inducidas y células de carcinoma humano embrionario también son capaces de albergar retrotransposición de LINE-1 usando el ensayo reportero *in vitro* de retrotransposición descrito por Moran et al., [12], [36], [59], [60]. Por tanto, todos estos datos de forma conjunta, sugieren que la retrotransposición podría ocurrir en estadios tempranos del desarrollo. Además, ha sido demostrado recientemente que células madre pluripotentes inducidas son capaces de acumular nuevas inserciones endógenas de LINE-1. [73]. No obstante, células germinales mamíferas sobreexpresan el ARN mensajero de LINE-1 y presentan niveles elevados de ORF1p [74]–[76]. De cualquier forma, un estudio realizado en 2009 usando ratones transgénicos demostró que la mayoría de las nuevas inserciones se acumulan durante el desarrollo embrionario siendo raras las inserciones en células germinales [77]. Por otra parte, la búsqueda de la copia progenitora de una inserción de L1 que había producido la enfermedad de la coroderemia (enfermedad degenerativa del ojo ligada al cromosoma X) en un paciente holandés demostró que la inserción que produjo la enfermedad en el paciente había ocurrido durante el desarrollo embrionario temprano de la madre ya que ésta era un mosaico somático y un mosaico en su línea germinal [78]. Por tanto, todos estos datos apoyan el hecho de que la mayoría de las inserciones se acumulan durante etapas tempranas del desarrollo embrionario.

#### ○ *Retrotransposición en tejidos adultos*

Por otro lado, también se ha descrito que la retrotransposición puede tener lugar en células diferenciadas lo cual no favorece la herencia de estas inserciones de LINE-1. De hecho, células progenitoras neuronales, presentan niveles moderados de ARNm de LINE-1 y ORF1p respecto a células madre embrionarias [79], [80]. Sin embargo, este parece un fenómeno restringido a este tipo de célula pluripotente somática ya que en otras células pluripotentes adultas no se observa este fenómeno y los niveles de ARNm de LINE-1 y ORF1p son muy similares a células diferenciadas [79]. Además, estos estudios han

demostrado que la retrotransposición *in vitro* puede tener lugar en células progenitoras neuronales murinas y humanas respectivamente [79], [80]. Asimismo, esta retrotransposición *in vitro* también puede ser albergada por neuronas adultas totalmente diferenciadas que no se dividen derivadas en cultivo desde células progenitoras neuronales [79]. Finalmente, varios estudios usando tecnologías como RC-seq (del inglés *Retrotransposon Capture Sequencing*) o WGS (del inglés *Whole Genome Sequencing*) confirmaron que en el cerebro adulto existe retrotransposición endógena de LINE-1 [81]–[83].

La retrotransposición en células somáticas no parece estar restringida a neuronas sino que además se produce también en células tumorales. Son múltiples los estudios que demuestran la presencia de nuevas inserciones de LINE-1 en diversos tipos de tumores epiteliales [14]–[16], [84], [85]. Además, estas inserciones pueden contribuir a la malignidad del tumor como se demostró con inserciones endógenas en genes supresores de tumores en un caso de carcinoma hepatocelular [16].

### 1.3.2. Consecuencias de la retrotransposición sobre el genoma hospedador

La movilización del retroelemento LINE-1 conlleva una serie efectos sobre el genoma hospedador, de los cuales hay algunos que pueden llegar a causar enfermedad o se especula con que podría hacerlo:

#### 1.3.2.1. Efectos sobre la expresión génica

##### ○ *Mutagénesis insercional*

Se puede producir una alteración de la expresión génica por interrupción directa de un exón debido a la integración de una nueva copia de LINE-1 dentro de dicho exón. Ejemplos de este mecanismo de mutagénesis son las inserciones en el exón 14 del gen del factor VIII que dieron lugar a la ausencia de actividad de dicho factor resultando en la aparición de hemofilia tipo A en dos individuos sin parentesco [86]. Por otro lado, un caso de granulomatosis crónica ligada al cromosoma X en un paciente holandés es generada por una inserción de LINE-1 que se insertó dentro del exón 4 del gen *CYBB* dando lugar a la ausencia de Citocromo b, que es el complejo del que forma parte el producto proteico del gen mutado [11].

##### ○ *Alteración del splicing*

La inserción de LINE-1 puede alterar el patrón del splicing de un gen debido a la presencia de sitios de splicing en su secuencia [87]. Dichos sitios de splicing se pueden

utilizar junto con los sitios de splicing situados en el gen dando lugar a nuevas variantes génicas no funcionales. De hecho, en el genoma de referencia existen eventos de splicing entre el elemento LINE-1 y genes humanos [87]. Por otro lado, el patrón de splicing también puede ser alterado ya que la inserción puede interrumpir la secuencia de un sitio donador de splicing así como también la secuencia de un sitio de aceptor de splicing. De hecho, existe un caso recién publicado de retinoblastoma familiar donde intervienen ambos supuestos relacionados con el splicing como causa de la enfermedad. En primer lugar, la inserción de LINE-1 se había producido entre el intrón 14 y el exón 15 del gen *RB1* justo en la secuencia necesaria para que se produzca un splicing correcto. De esta forma se propició un splicing aberrante que dio lugar a la aparición de variantes de ARNm a las que le faltaba el exón 15 debido a un fenómeno conocido como salto del exón. En segundo lugar, esta inserción provocó la aparición de nuevas variantes de ARNm que incluían exones crípticos debido al uso del sitio donador para el splicing situado en la posición +92 de la secuencia de LINE-1 [87] junto con otros sitios aceptores localizados aguas abajo de la inserción [88]. El fenómeno de salto de exón ha sido descrito en otras enfermedades generadas por nuevas inserciones de LINE-1 [89], [90].

○ *Pausa transcripcional*

Existe una correlación invertida entre el nivel de expresión del gen y la presencia de un elemento LINE-1 dentro del mismo. Esto se debe a que las inserciones de LINE-1 inducen pausa transcripcional ya que su secuencia es rica en adeninas y timinas. Para ello, la inserción se debe encontrar en el mismo sentido transcripcional que el gen. Por otro lado también pueden inducir terminación prematura debido a presencia de señales de poliadenilación [2], [91].

*1.3.2.2. Efectos estructurales sobre el sitio de inserción.*

Según el modelo propuesto por Gilbert et al., [92], el lugar donde se produce el corte en la hebra de arriba durante el mecanismo de TPRT determina que se produzcan duplicaciones del sitio de inserción o bien deleciones del sitio de inserción. Si el corte se produce aguas abajo del corte de la hebra de abajo, entonces se genera una duplicación del sitio de inserción, sin embargo, cuando este corte se produce aguas arriba se puede generar una deleción del sitio de inserción.

○ *Duplicaciones del sitio de inserción.*

Las duplicaciones del sitio de inserción son una característica propia del mecanismo de inserción del retroelemento LINE-1 [67] y suelen oscilar entre 2 y 20 nucleótidos. Sin embargo, en células en cultivo, utilizando un vector modificado respecto al empleado en

Moran et al., [12] que permite la recuperación de las inserciones, se han observado inserciones de LINE-1 acompañadas de largas duplicaciones del sitio de inserción llegando hasta 6.000 pares de bases [92], [93] provocando la duplicación de un exón del gen donde se había insertado el retroelemento LINE-1. No obstante, en el genoma de referencia, las duplicaciones del sitio de inserción mayores de 50 nucleótidos son escasas [4].

- *Deleciones del sitio de inserción.*

La integración de una nueva inserción de LINE-1 puede ir acompañada de deleciones del lugar de inserción. De hecho, en los mismos estudios citados en el apartado de arriba [92], [93], en células en cultivo se demostró la generación de deleciones mediadas por la integración de una nueva copia de LINE-1 las cuales podían alcanzar hasta una longitud de más de 71 kilobases [92]. En un estudio similar, LINE-1 generó 3 deleciones independientes que englobaban varios exones de distintos genes cada una [94]. Además, las deleciones también ocurren *in vivo*, ya que un paciente con deficiencia del complejo de la piruvato deshidrogenasa presentaba una inserción de copia completa de LINE-1 en el gen *PDHX* (componente X del gen de la piruvato deshidrogenasa) ocasionando una deleción de 46 kb [95].

- *Deleciones del sitio de inserción asociadas a la vía independiente de la endonucleasa.*

Existe una vía alternativa de retrotransposición donde el retroelemento LINE-1 puede aprovechar lesiones existentes en el genoma para insertarse sin necesidad de tener que hacer la mella en el ADN con su propia actividad endonucleasa albergada en ORF2p. Esto fue demostrado por primera vez en Morrish et al., [96] utilizando diferentes líneas celulares de ovario de hámster deficientes para componentes de la vía de reparación del ADN NHEJ (del inglés *Non-Homologous End Joining*). Estas inserciones presentan características algo diferentes a las típicas de una inserción canónica de LINE-1. No tienen duplicaciones del sitio de inserción, presentan deleciones del sitio de inserción que pueden variar en longitud (hasta 1,5kb en el caso de la inserción más larga descrita), incorporación de ADN no procedente de la secuencia del LINE-1 en el extremo 5' y en el extremo 3', además de no tener cola de poli A y estar truncadas bien sobre la señal de poliadenilación o aguas arriba de la misma. Cabe destacar que estas inserciones ocurren de forma natural ya que se han encontrado en el genoma humano [97]. Un caso de síndrome braquio oto renal es generado por una deleción de 17kb del gen *EYA1* debido a la integración de una nueva inserción de LINE-1. Esta inserción podría estar truncada en 3' ya que carece de cola de poli A. Por tanto, este es un caso de enfermedad que podría estar generado por este tipo de retrotransposición independiente de endonucleasa [98].



○ *Deleciones asociadas a la formación de LINE-1 quiméricos*

En células en cultivo, se ha observado como en algunos casos las nuevas inserciones de LINE-1 que se integran en regiones del genoma ricas en secuencias LINE-1, pueden dar lugar a la formación de L1 quiméricos que contienen una porción de un elemento L1 que se encuentra aguas arriba de la inserción. Estas quimeras pueden ir asociadas a cambios estructurales del sitio de inserción como deleciones. De hecho, se ha documentado deleciones de hasta 3 kb asociadas a este tipo de fenómeno [92], [93].

○ *Barajamiento de exones.*

En algunas ocasiones (10-30%), la ARN polimerasa II encargada de la transcripción del retroelemento LINE-1 ignora la débil señal de poliadenilación situada en su región 3' no traducida, y ésta prosigue adentrándose en el genoma colindante hasta encontrar otra señal de terminación de la transcripción. Esto es lo que se conoce como transducción en 3', y de esta forma el retroelemento LINE-1 puede arrastrar fracciones de genoma y depositarlas en un sitio diferente del genoma dando lugar a nuevos genes como se demostró con el ensayo de retrotransposición [21]. Es un fenómeno bastante común, ya que aproximadamente el 20% de los LINE-1 del genoma están flanqueados por una fracción del genoma procedente de otro lugar [99], [100]. Además, las transducciones en 3' han contribuido a que se puedan aislar inserciones progenitoras de algunas copias que han generado enfermedades ya que estas regiones en 3' constituyen una etiqueta genómica única que permite la identificación de la copia progenitora [101]. Por otro lado, este fenómeno puede tener unos beneficios para el genoma hospedador debido a que puede haber una evolución del genoma pues se favorece el transporte de elementos reguladores como promotores así como se pueden movilizar fracciones de genes de un lugar a otro generándose nuevos genes. Además, este fenómeno puede también ocurrir en la región 5' dando lugar a una transducción en 5', es decir, en algunas ocasiones se produce la transcripción del retroelemento LINE-1 desde algún promotor situado aguas arriba de tal modo que la nueva inserción de LINE-1 lleva asociada un fragmento del genoma en 5' procedente del lugar donde se encontraba la copia madre [1]. Es posible que transportar fracciones de genoma de un sitio a otro pueda producir alteraciones, como ocurre con el gen *Nr2* donde una inserción de LINE-1 de copia completa con una transducción en 5' de 28 nucleótidos dio lugar a un splicing erróneo de tal forma que se generó una enfermedad asociada a degeneración de la retina en un modelo murino [102].

### 1.3.2.3. *Otros efectos*

○ *Efectos epigenéticos.*

Inserciones de *novο* artificiales son silenciadas epigeneticamente rápidamente después de la integración, en células de teratocarcinoma humano embrionario mediante desacetilación de histonas [60]. Si estos cambios generados sobre la inserción de LINE-1 pueden afectar al genoma colindante y tener un efecto sobre la expresión génica es algo desconocido hasta la fecha. De esta forma, es posible especular con que cambios epigenéticos en el genoma adyacente podrían tener un impacto sobre elementos reguladores como promotores o enhancers alterando la expresión génica.

○ *Posibles efectos derivados del promotor antisentido.*

Los transcritos producidos desde este promotor contienen una parte de la secuencia del retroelemento LINE-1 así como una parte de la secuencia del genoma situada aguas arriba del mismo [44], [45]. Las funciones celulares de estos transcritos no son conocidas, y, por tanto, una nueva inserción puede dar lugar a la formación de nuevo transcrito quimérico que podría tener un impacto sobre la homeostasis celular. Un porcentaje de los ARNs no codificantes de cadena larga en mamíferos contienen secuencias de elementos móviles [103]. Las funciones de estos transcritos no codificantes de cadena larga son múltiples e intervienen regulando multitud de procesos celulares [104], [105]. Por tanto, estos ARNs que se han demostrado ser de cadena larga y poliadenilados [44], [45] podrían actuar como ARNs de cadena larga no codificantes ejerciendo algún tipo de función. De hecho, transcritos antisentido que contienen secuencias de retrovirus endógenos de ratón así como transcritos derivados del retrovirus endógeno HERV-H (del inglés *Human Endogenous Retrovirus H*) intervienen en el mantenimiento de la pluripotencia en células madre embrionarias [106], [107]. Por tanto, se puede especular con que una nueva inserción de copia completa podría dar lugar a la generación de un nuevo ARN de cadena larga en la nueva localización genómica alterando el funcionamiento celular.

Otro posible efecto derivado de la existencia de dos promotores, es el fenómeno conocido como “división del gen” (del inglés *Gene Breaking*). Aunque puede ser un fenómeno bastante raro, sería un efecto que podría resultar en la aparición de una enfermedad. Este consiste en que una inserción génica de copia completa antisentido respecto a la orientación del gen podría dividir la unidad transcripcional en dos transcritos debido a la acción de los dos promotores en dos direcciones diferentes [2], [108].

○ *Efectos derivados de la retrotransposición de ARNs celulares*

Las proteínas de LINE-1 pueden producir variaciones en el genoma humano a través de la movilización en *trans* de ARNs celulares a un sitio diferente generando nuevas proteínas de fusión, lo cual en principio podría eventualmente causar enfermedad en

función del tipo de proteína de fusión que se pudiera originar. Hasta la fecha, tan solo se conocen ejemplos beneficiosos como la formación de la proteína de fusión entre TRIM5 $\alpha$ -locus y el cDNA de la ciclofilina lo cual ocasionó que monos que eran sensibles al virus del SIDA, pasaran a ser resistentes [109]. En humanos existe un ejemplo de un gen *PIPSL* que proviene de la fusión de los ARNs de *PIP5K1A* y la subunidad 26 del proteosoma que experimentó retrotransposición y que parece haber sido seleccionado positivamente [110], [111].

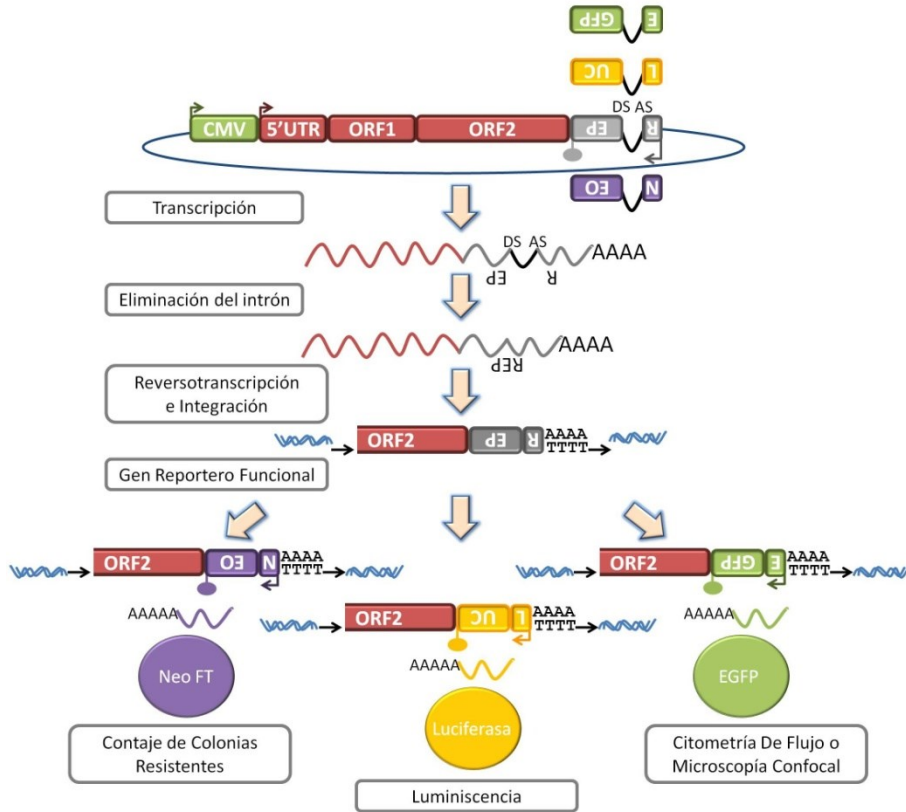
○ *Reordenaciones cromosómicas mediadas por recombinación entre alelos de LINE-1.*

La masiva presencia de elemento LINE-1 a lo largo de genoma favorece que se pueda producir recombinación entre copias del mismo generándose translocaciones, deleciones o duplicaciones cromosómicas. De hecho, existen varios ejemplos de NAHR (del inglés, *Non Allelic Homologous Recombination*) entre dos copias de LINE-1 que dan lugar a una serie de síndromes como la deficiencia de fosoforilasa kinasa, síndrome de Alport y el síndrome de Ellis van Creveld [2].

**1.3.3. Ensayo de retrotransposición: midiendo la retrotransposición *in vitro*.**

El ensayo de retrotransposición es una técnica que permite medir la movilidad del elemento LINE-1 en células en cultivo. Esta técnica fue desarrollada por John Moran en 1996, permitiendo demostrar que copias progenitoras de inserciones de LINE-1 causantes de enfermedad [101], [112] eran efectivamente activas. Además esta técnica ha permitido conocer cuáles son las partes de la secuencia del retroelemento LINE-1 necesarias e indispensables para la retrotransposición como aquellas que pueden ser prescindibles [12], [20], [57]. Asimismo, una evolución de este ensayo permitió en su momento caracterizar inserciones de LINE-1 mediante la recuperación de las mismas desde el genoma hospedador descubriendo una gran gama de alteraciones asociadas a la inserciones de LINE-1 en distintos contextos genéticos [92], [93], [96]. Igualmente, se ha utilizado para demostrar el efecto de muchos reguladores de la retrotransposición [63], [113]. En definitiva, se trata de un ensayo que contribuyó y sigue contribuyendo a la expansión del conocimiento en el campo de la biología de LINE-1. De hecho, incluso permitió desarrollar un modelo de retrotransposición en ratón demostrando que el cerebro *in vivo* podía acomodar retrotransposición de LINE-1 [114].

Este ensayo se basa en la transfección celular con un plásmido que lleva una copia activa de LINE-1 fusionada a un casete reportero que permite medir la retrotransposición. En el primer tipo de ensayo desarrollado por Moran *et al.*, [12], se etiquetó una copia supuestamente activa, aquella progenitora de la inserción de LINE-1 causante de la



**Figura 3. Esquema del ensayo de retrotransposición.** Brevemente, una copia activa de LINE-1 fusionada a un gen reportero (*NEO*, *EGFP*, *LUC*) que solo puede ser funcional después de una ronda de retrotransposición. Una vez el gen reportero es funcional, la retrotransposición puede ser medida en función de las características de éste. Abreviaciones utilizadas: REP, reportero; EGFP, gen de la proteína verde fluorescente; LUC, gen de la luciferasa; NEO, gen de resistencia a neomicina; CMV, promotor intermedio-temprano del citomegalovirus; DS, donador de splicing; AS, aceptor de splicing; Neo FT; Neomicina fosfotransferasa.

hemofilia, con un gen de resistencia a neomicina. La copia de LINE-1 fusionada al gen de resistencia a neomicina fue clonada en el vector de expresión episomal pCEP4, donde la expresión está controlada por el promotor intermedio-temprano del citomegalovirus (CMV). El gen de resistencia tiene la información para la síntesis de la enzima neomicina fosfotransferasa que degrada el antibiótico neomicina. Este gen de resistencia a neomicina fue clonado en antisentido respecto al sentido transcripcional del LINE-1. Además, este gen se encuentra interrumpido por el intrón IVS-2 de la  $\gamma$ -globina que a su vez se encuentra clonado en sentido respecto al sentido transcripcional de LINE-1. Asimismo, el gen de resistencia a neomicina lleva su propio promotor heterólogo (SV40) y su propia señal de poliadenilación. Sin embargo, el gen no puede dar lugar a una proteína funcional desde su

promotor debido a la orientación del intrón que impide su eliminación mediante splicing. Por tanto, el diseño del constructo favorece que únicamente el intrón se pueda eliminar mediante splicing cuando la transcripción tiene lugar desde el promotor de LINE-1. De esta forma, el ARNm resultante carecerá de intrón y una vez sea reverso transcrito e integrado en el genoma, entonces, el gen de resistencia a neomicina será funcional. Esto hace que el gen de resistencia a neomicina pueda ser expresado desde su promotor provocando que las células, donde haya tenido lugar la retrotransposición, sean resistentes a neomicina y puedan crecer en presencia de este antibiótico. Por tanto, las células serán resistentes a neomicina solo después de una ronda de retrotransposición. De esta forma, después de un periodo de selección con neomicina, las células donde hubo retrotransposición habrán formado colonias que se podrán fijar y teñir para posteriormente ser contadas y estimar la tasa de retrotransposición en cuantas condiciones experimentales se requiera (Figura 3).

Posteriormente, se han generado otros constructos que llevan otros genes reporteros que permiten medir la retrotransposición de una forma diferente. De hecho, existe una variante del ensayo con el gen de la proteína verde fluorescente [115] como gen reportero que permite la cuantificación de la retrotransposición mediante citometría de flujo o microscopía confocal. Más tarde también se desarrolló un nuevo tipo de ensayo de retrotransposición que usaba como gen reportero el gen de la luciferasa con lo que la retrotransposición puede ser cuantificada midiendo la emisión de luminiscencia [116]. Estos nuevos ensayos han permitido expandir el nº de aplicaciones del ensayo de retrotransposición (Figura 3).

#### 1.3.4. Mecanismos de control de la retrotransposición

Teniendo en cuenta todas las posibles formas mediante las cuales LINE-1 puede alterar el genoma (apartado 1.3.2), es lógico pensar que deben existir mecanismos por parte de la célula para controlar la retrotransposición de tal forma que se mantenga a unos niveles donde pueda ser más beneficiosa que perjudicial, pues como se puede deducir de algunas de las posibles alteraciones asociadas a la misma, ésta también puede tener efectos beneficiosos para el genoma hospedador. Además, a todo esto hay que añadir, que algunos de los fenómenos descritos también se asocian a la movilización en *trans* de elementos Alu, SVA y de ARNs celulares, con lo cual, las posibilidades de afectar el genoma hospedador se multiplican. Es por ello que es crucial una buena regulación de la actividad de LINE-1. Existen evidencias de que la retrotransposición puede ser controlada a distintos niveles a lo largo del ciclo de retrotransposición. Por tanto, se van a analizar los diferentes mecanismos de control según la etapa del ciclo donde dicho mecanismo ejerza su control.

#### 1.3.4.1. Control transcripcional

Este control se ejerce mediante la metilación del promotor del retroelemento LINE-1. El promotor del retroelemento LINE-1 presenta 20 islas CpG que pueden ser metiladas por la DNMT3L (del inglés *de novo* DNA methyltransferase 3-like) inhibiendo la transcripción del promotor del LINE-1 en células germinales. De hecho, un modelo de ratón knock-out para *DNMT3L*<sup>-/-</sup> presenta un incremento del ARNm de LINE-1 en las células progenitoras germinales masculinas [35]. La metilación es un mecanismo que opera sobre todo en células diferenciadas ya que el grado de metilcitosinas en el promotor de LINE-1, y, por tanto, el grado de metilación del promotor es superior en células diferenciadas comparadas con células pluripotentes [36]. De esta forma, el grado de metilación del promotor de LINE-1 correlaciona inversamente con la cantidad de ARNm de LINE-1 (Figura 4A)[36].

Por otro lado, los elementos LINE-1 presentes en el genoma de ratón se encuentran decorados con H3K9me3 en células madre embrionarias. La ausencia de la enzima responsable del depósito de esta modificación conlleva la reactivación transcripcional de estos elementos [117] (Figura 4B).

#### 1.3.4.2. Control post-transcripcional

Existen distintos posibles mecanismos a este nivel para controlar la actividad de LINE-1.

En primer lugar, en la línea germinal de mamíferos y *Drosophila* existe una clase de ARNs pequeños denominados piRNAs (del inglés *PIWI Interacting RNAs*) que participan en el control de los elementos transponibles. Brevemente, estos ARNs pequeños se producen desde la transcripción de clusters de elementos transponibles. Posteriormente, a través del mecanismo ping-pong se produce la aparición de piRNAs maduros en el citoplasma que junto con las proteínas PIWI median la destrucción endonucleolítica de los ARNs de elementos transponibles. Los piRNAs actúan como guías ya que son complementarios a los ARNs de los elementos transponibles (Figura 4G, H, K, L). En ratones mutantes para una serie de proteínas PIWI, se produce una desrepresión de los elementos móviles LINE-1 durante la espermatogénesis [118], [119].

Otras especies de ARNs que podrían operar a nivel post-transcripcional son los microRNAs. Células embrionarias de ratón *DICER*<sup>-/-</sup> (proteína implicada en la biogénesis de microRNAs) presentan un mayor nivel de ARN mensajero de LINE-1 y de ORF1p. Consistentemente, células *DICER*<sup>-/-</sup> albergan mayor tasa de retrotransposición *in vitro* [47].

Esta proteína forma parte de la ruta de biogénesis de microRNAs, con lo que, este incremento del ARNm de LINE-1 se podría deber a la ausencia de algún microRNA que pudiera ser un regulador negativo del retroelemento LINE-1. En efecto, se ha descrito que mir128 regula negativamente la retrotransposición tanto en células pluripotentes como en células diferenciadas. De hecho, este microRNA produce un descenso de los niveles de ARNm de LINE-1, ORF1p y tiene un efecto negativo sobre la retrotransposición *in vitro*. El efecto de este microRNA es directo pues se demostró la presencia del ARNm de LINE-1 y mir128 formando un complejo con AGO2 (proteína efectora de la ruta de biogénesis de microRNAs) (Figura 4E,F,K,L) [120].

Por otro lado, el efecto que puede tener la ausencia de DICER en la regulación del retroelemento LINE-1 también se puede deber a la ausencia de siRNAs (del inglés *small interfering RNAs*) que se podrían generar desde dos promotores solapantes en sentidos opuestos, como son los contenidos en la región 5' no traducida de LINE-1 [37], [43]. Se especula con que quizás la expresión de ambos promotores, que además se puede producir de manera simultánea [44] en células pluripotentes humanas, podría dar lugar a la aparición de especies de ARN de doble cadena que podrían ser precursores de siRNAs. Estos precursores serían procesados por DICER y los siRNAs podrían actuar sobre el propio ARNm del retroelemento LINE-1, lo cual ya fue sugerido en un estudio en células humanas [46]. A día de hoy no está claramente demostrado que existan estas especies de ARNs que regulan la actividad de LINE-1 en las células. Sin embargo, se sabe que el ARN procedente de ambos promotores puede formar un ARN de cadena doble y que puede ser procesado *in vitro* por DICER generando siRNAs. Esto siRNAs tienen un efecto negativo sobre la retrotransposición *in vitro* lo cual podría estar mediado por la degradación del ARNm del retroelemento LINE-1 [121]. Además, en ratón se encontraron ARNs pequeños de 18-21 nucleótidos que mapean exactamente en el promotor de LINE-1 de las subfamilias T<sub>F</sub> y G<sub>F</sub> (Figura 4I,J,K,L) [122].

Por último, también es de destacar la posible implicación de la proteína DGCR8 que forma parte del complejo DROSHA/DGCR8 denominado microprocesador, el cual participa en los primeros pasos de la biogénesis de microRNAs. Los microRNAs se transcriben por la ARN polimerasa III inicialmente como un ARN largo de varias kilobases. Este ARN adquiere una estructura secundaria que consiste en una serie de estructuras tipo horquilla (stem loops) a lo largo del mismo, flanqueadas por una horquilla final en un extremo y por ARN de cadena simple en el otro. DGCR8 es la proteína del complejo microprocesador que reconoce el ARN de cadena simple y se une al microRNA facilitando el corte del mismo por parte de su compañera DROSHA [123]. El papel de DGCR8 puede ser importante en la regulación de los elementos LINE-1 ya esta proteína parece que se une físicamente al ARN

del retroelemento LINE-1 [124]. De hecho, podría ocurrir como en el caso de ARNs codificantes de proteína que son directamente regulados por el microprocesador de una forma independiente de DICER [124]. Sin embargo, no existen datos sobre si esta unión de DGCR8 y el ARNm del retroelemento LINE-1 se manifiesta en algún tipo de regulación sobre el mismo. De hecho, no existen estudios que aborden el efecto que puede llegar a tener el complejo microprocesador sobre la movilidad del retroelemento LINE-1.

Además, existen proteínas que directamente podrían mediar la destrucción del ARNm de LINE-1 como es la proteína RNASEL (del inglés *Ribonuclease L*) que inhibe la retrotransposición *in vitro* de LINE-1. Se piensa que el efecto se podría deber al reconocimiento de las especies de ARN de doble cadena complementarios originados desde ambos promotores del retroelemento LINE-1 (Figura 4M-N) [118].

#### 1.3.4.3. Control de la ribonucleopartícula

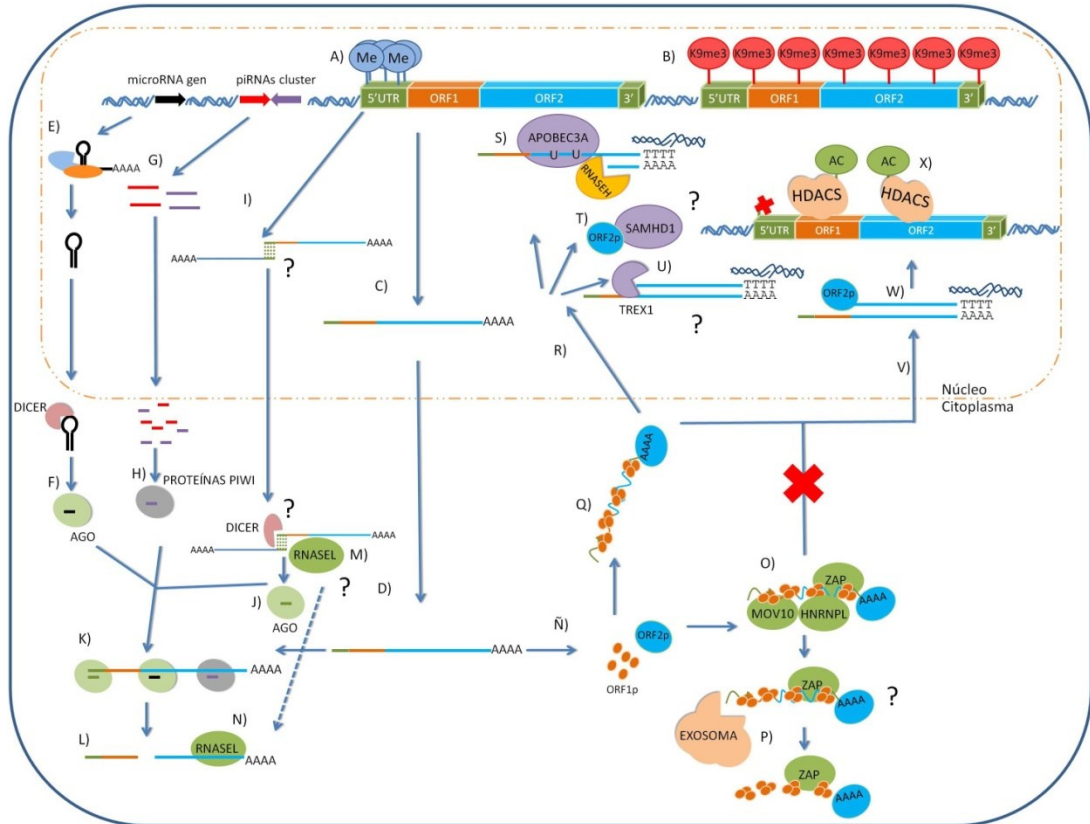
Como se ha mencionado más arriba, las proteínas ORF1p y ORF2p se une en *cis* a su ARNm formando el intermediario de la retrotransposición que se denomina ribonucleopartícula. Distintos elementos reguladores parece que podrían ejercer su papel regulador a este nivel (Figura 4O, P).

MOV 10 es una ARN helicasa de la superfamilia tipo UPF1 dependientes de ATP. Esta proteína colocaliza con ORF1p en gránulos citoplasmáticos, co-immunoprecipita con la ribonucleopartícula de LINE-1 además de unirse al ARNm de LINE-1. Inhibe la retrotransposición *in vitro* y su presencia correlaciona inversamente con la cantidad de ARNm de LINE-1. Sin embargo, su mecanismo de acción se encuentra todavía en debate [118].

ZAP es una proteína de la familia de las PARP (del inglés Poly ADP-Ribose Polimerase). Esta proteína es fundamentalmente citoplasmática y su modo de acción se basa en promover la degradación del ARN de virus mediante el reclutamiento de proteínas del exosoma y su transporte a gránulos citoplasmáticos. Inhibe la retrotransposición *in vitro* de LINE-1 y se une a la ribonucleopartícula así como colocaliza con ORF1p y el ARNm de LINE-1 en gránulos de estrés. Otro tipo de mecanismos como inhibición de la traducción no puede ser excluido [118].

Por otro lado, se ha descrito en ratón que la proteína HNRNPL, que se une a la señal de entrada al ribosoma (IRES, del inglés Internal Ribosome Entry Site) en el ARNm de LINE-1, es un factor restrictivo de la retrotransposición ya que inhibe la retrotransposición *in vitro*. Además, esta proteína co-immunoprecipita con la ribonucleopartícula de LINE-1 [118].





**Figura 4. Mecanismos de regulación de los retroelementos LINE-1 durante el ciclo de retrotransposición.** La transcripción desde una copia activa de LINE-1 puede ser inhibida mediante metilación del promotor de LINE-1 (A) o mediante el depósito de H3K9me3 (B). Si se eluden estos mecanismos, se produce transcripción (C) y el ARNm migra al citoplasma (D) donde puede ser traducido (Ñ) o puede ser degradado (K-L) como resultado final de varios mecanismos de control que actúan a nivel de ARNm como son microRNAs (E-F, procesamiento de microRNAs hasta cargar el microRNA maduro en la proteína AGO) piRNAs (G-H, procesamiento de los piRNAs hasta cargar el piRNA en las proteína PIWI) y siRNAs (I-J, procesamiento de ARNs de interferencia hasta que el ARN pequeño de interferencia es cargado en AGO). Igualmente, es posible que una proteína (RNASEL) pudiera mediar directamente su degradación reconociendo los ARNs de cadena doble procedentes de LINE-1 (M-N). Si el ARNm consigue escapar a estos mecanismos de control, entonces, es traducido (Ñ) y puede formar la ribonucleopartícula con ORF1p y ORF2p (Q). Sin embargo, la ribonucleopartícula, puede ser también objeto de control por varias proteínas citoplasmáticas que inhiben la retrotransposición. Esta ribonucleopartícula podría ser dirigida a gránulos citoplasmáticos y el ARNm ser degradado vía exosoma como ocurre con algunos virus (O, P). Cuando se elude este control, la ribonucleopartícula puede llegar al núcleo (R) donde mediará la generación de una nueva inserción. Sin embargo, a nivel preintegracional también existen otros mecanismos para inhibir la retrotransposición mediante desaminación de citosinas (APOBEC3A) (S), mediante secuestro de

ORF2p (SAMHD1) (**T**) o mediante degradación del ADN de cadena simple (TREX1) (**U**). No obstante, si se eluden estos sistemas de control, se puede llegar a generar una nueva inserción de LINE-1 (**V-W**) aunque puede ser rápidamente silenciada epigenéticamente mediante la intervención de histonas desacetilasas (**X**). Datos más concretos, en el texto.

Estudios de co-inmunoprecipitación de ribonucleopartículas de LINE-1 y de proteómica de afinidad revelan la existencia de una gran cantidad de proteínas, cuyo papel en la defensa del genoma frente a virus es conocido, que co-inmunoprecipitan con la ribonucleopartícula y que además inhiben la retrotransposición *in vitro* de LINE-1 [118].

#### 1.3.4.4. Control pre-integracional

Cuando la retrotransposición se encuentra en sus últimos pasos, una vez que la ribonucleopartícula ha conseguido migrar al núcleo para proceder con el mecanismo de TPRT y terminar generando una nueva inserción en algún lugar del genoma, también existen una serie de reguladores de la retrotransposición.

La familia APOBEC son citidinas desaminasas que actúan convirtiendo citosinas en uracilos. Su función es conocida frente a retrovirus. La familia consta de 8 miembros (A-H), de los cuales algunos son capaces de inhibir la retrotransposición en distintos grados. A, B, C y F inhiben la retrotransposición siendo APOBEC3A el miembro que ejerce una mayor restricción sobre la retrotransposición *in vitro*. Esta diferencia de potencial restrictivo tiene que ver con su localización subcelular siendo las más efectivas aquellas que se localizan bien sólo en el núcleo o bien en ambos compartimentos [125]. El modelo de acción propuesto es que durante el proceso de TPRT, cuando el híbrido de ADN:ARN es resuelto por la actividad de la ARNasa H para comenzar con la síntesis de la segunda hebra, entonces se produce la desaminación de las citosinas expuestas en la cadena simple de ADN de LINE-1 lo cual podría reclutar UNG (uracilo ADN glicosilasa) y APE (endonucleasa apurínica/apirimidínica) para la reparación o eliminación del ADN de cadena simple desaminado de LINE-1 (Figura 4S) [126]. Este mecanismo podría contribuir a la restricción de la retrotransposición de LINE-1 en células pluripotentes ya que APOBEC3B se expresa en células madre embrionarias humanas y es capaz de disminuir la tasa de retrotransposición *in vitro* de LINE-1 [127]. Esto concuerda con el hecho de que la retrotransposición se produce en etapas tempranas del desarrollo [59], [77], [78]. Sin embargo, no se descartan otros posibles papeles en la regulación de los elementos LINE-1 para estas proteínas tales como dirigir la ribonucleopartícula hacia gránulos de estrés para la degradación del ARN vía siRNA [118].

Otras proteínas que podrían inhibir la retrotransposición a nivel preintegracional son TREX-1 y SAMHD1 las cuales están relacionadas con el síndrome Aicardi-Goutières.

SAMHD1 es una nucleótido trifostato nucleasa cuya acción conlleva la disminución del nivel de dNTPs. [118]. Sin embargo, se ha demostrado que la inhibición sobre la retrotransposición es independiente de su función trifostato nucleasa ya que podría actuar secuestrando ORF2p impidiendo la reversotranscripción. (Figura 4T) [128]. Sin embargo, no se descarta otro tipo de inhibición, relacionada con el secuestro de la ribonucleopartícula hacia gránulos de estrés [129]. Otra proteína cuya acción podría tener lugar también a nivel preintegracional, sería TREX-1 (ADN exonucleasa). Esta proteína tiene capacidad exonucleolítica 3'>5'. La sobreexpresión de esta proteína lleva asociado el aumento de ADN de cadena simple de LINE-1 y una disminución de la retrotransposición *in vitro* (Figura 4U) [130].

#### 1.3.4.5. Control post-integracional.

En células pluripotentes, otra forma de control es el silenciamiento epigenético de las nuevas inserciones artificiales de LINE-1 mediante desacetilación de histonas del locus donde se ha insertado el retroelemento. Esto ocurre rápidamente después de la integración de una nueva inserción de LINE-1, y parece ser un mecanismo propio de células pluripotentes ya que en células diferenciadas el silenciamiento mediante desacetilación de histonas se encuentra atenuado [60]. Una vez que el retroelemento LINE-1 ha conseguido escapar al resto de mecanismos de control, el silenciamiento mediante desacetilación de histonas aseguraría que las nuevas inserciones de copia completa no se pudieran retrotransponer de nuevo manteniéndolas transcripcionalmente inactivas (Figura 4X).

Como se observa en la figura 4 existen una gran variedad de mecanismos para que la retrotransposición esté controlada y no sea perjudicial para la integridad del genoma. Hay que tener en cuenta que, algunas de las proteínas implicadas son redundantes y posiblemente participen inhibiendo la retrotransposición a distintos niveles. En este caso se han referido las funciones más probables para cada uno de los elementos reguladores. Además, no todos los mecanismos operan al mismo tiempo en un tipo celular concreto y de la misma forma no todos actúan en los diferentes estados celulares.

### 1.4. Regulación epigenética en mamíferos

#### 1.4.1. Principios generales de la regulación epigenética

La epigenética se puede definir como el estudio de la información heredable presente en el genoma, no contenida en la secuencia del ADN, que regula la expresión génica de una forma dependiente del tiempo y del tejido celular. Esta información permite que las células de un organismo multicelular puedan adquirir diferentes fenotipos a lo largo de la

diferenciación celular alcanzando un fenotipo con unas funciones totalmente diferentes respecto a otras células diferenciadas pertenecientes a otros tejidos.

La epigenética, entre otras cosas, estudia los distintos estados de la cromatina. Se puede definir cromatina como la forma en la cual el ADN está empaquetado en la célula. La estructura básica de la cromatina es el nucleosoma. El nucleosoma está formado por un octámero de cuatro proteínas denominadas histonas (H3, H4, H2A, H2B) y 147 pares de bases enrolladas alrededor del mismo. Los residuos de las colas amino terminal de las histonas pueden ser covalentemente modificados por una serie de enzimas que pueden añadir distintos grupos químicos a estos residuos provocando que el estado de la cromatina cambie. Este es un proceso dinámico ya que también existen enzimas que eliminan estos grupos químicos añadidos a las histonas. Las modificaciones químicas añadidas pueden ser acetilación, metilación, fosforilación, ADP ribosilación, sumoylación, ubiquitinación, deiminización e isomerización de prolina. Estas modificaciones de histonas funcionan estableciendo dominios globales de cromatina o bien orquestando procesos como la reparación del ADN, la replicación o la transcripción génica. Para ello, estas modificaciones tienen un efecto sobre el estado de la cromatina provocando que éste fluctúe entre una forma más relajada, que permite que tengan lugar funciones asociadas al ADN, y una forma más compacta que evita este tipo de funciones. De hecho, la acetilación y la fosforilación reducen las cargas negativas de las histonas disminuyendo la interacción electrostática entre las histonas y el ADN propiciando posiblemente una relajación de la cromatina. Por otro lado, las modificaciones también pueden reclutar enzimas o proteínas con dominios específicos de reconocimiento de una determinada modificación que pueden tener un efecto sobre el estado de la cromatina. De hecho, la acetilación de lisinas puede reclutar complejos remodeladores de la cromatina que favorecen la apertura de la cromatina. Así como también una modificación asociada con represión como H3K9me3 (trimetilación de la lisina 9 de la histona 3) puede reclutar a la proteína HP1 para reforzar la compactación de la cromatina. Incluso otras modificaciones evitan la unión de complejos remodeladores con actividad antagónica [131]–[133].

#### 1.4.2. Análisis del epigenoma a escala genómica

El ChIP (del inglés *Chromatin Immunoprecipitation*) se basa en la utilización de un anticuerpo específico para una determinada modificación de histona y utilizarlo para inmunoprecipitar y enriquecer la cromatina donde abunda dicha marca. Esta técnica seguida de PCR cuantitativa (ChIP-qPCR del inglés *Chromatin Immunoprecipitation followed by quantitative PCR*) permite un análisis locus específico. Por otro lado, el ChIP se puede acoplar a un CHIP de ADN (ChIP-on-CHIP) diseñado según el propósito del experimento,

donde se encuentran representadas una serie de regiones del genoma. De esta forma el ADN inmunoprecipitado se hace hibridar con el ADN del CHIP para encontrar las regiones que presentan una determinada modificación de histona. Evidentemente, para análisis globales, el ChIP-on-CHIP mejoraba bastante al ChIP-qPCR ya que da una visión más global de la localización de una determinada modificación, y de hecho, gracias al ChIP-on-CHIP se hicieron grandes avances en el campo de la epigenética [134]–[136]. Después de la revolución de la secuenciación masiva llegó el ChIP-seq (del inglés *Chromatin Immunoprecipitation followed by deep Sequencing*), por primera vez descrito en Barski et al., [137] y los estudios de epigenética comenzaron a revelar verdaderos mapas de modificaciones de histonas que permiten asociar una combinación de histonas a un determinado estado de la cromatina, a un determinado elemento de ADN o a una región del genoma y, por tanto, definen dominios dentro del genoma [137]–[143] confirmando la existencia de un código de histonas. Tanto es así, que este código de histonas se puede utilizar para encontrar nuevos elementos reguladores como promotores, enhancers así como regiones codificantes que no estaban anotadas [144]. De hecho, la combinación de histonas presente sobre un elemento regulador correlaciona con su función y su actividad en un determinado tipo celular [138], [145]. Por tanto, el conjunto de combinaciones de histonas sobre distintos elementos reguladores permite incluso la identificación de un determinado tipo celular [145]. De hecho, se ha llegado a establecer una huella epigenética a nivel de una sola célula [146]. Por tanto, cuanto más preciso sea el código de histonas será más sencillo asociar sin ambigüedad una determinada función biológica a una determinada combinación de modificaciones de histonas.

#### 1.4.3. Modificaciones de histonas asociadas a distintos estados de la cromatina

En este apartado se analizan las modificaciones de histonas asociadas a distintos estados de la cromatina utilizando datos existentes en células pluripotentes y diferenciadas tanto de humano como de ratón. En este caso, se considerarán tres escenarios posibles, un primer escenario donde la cromatina se encuentra más “abierta o relajada” y que se muestra accesible para una serie de funciones como la transcripción génica o la reparación del ADN. Este tipo de cromatina se ha denominado genéricamente como eucromatina [131]. En este caso, se revisarán las modificaciones de histonas asociadas a una de las funciones que están asociadas a este tipo de cromatina como es la transcripción génica. Posteriormente, se revisarán las modificaciones de histonas asociadas a un estado de la cromatina en el que la misma se encuentra más “cerrada o compacta” con una baja tasa de transcripción y se denomina heterocromatina. En este caso, se revisarán las modificaciones de histonas que participan en el proceso de silenciamiento génico asociado a la heterocromatina facultativa. Ésta se caracteriza porque es un tipo de

heterocromatina dinámica influenciada por el proceso de diferenciación celular pues depende de cada linaje celular [147]. Además, se revisarán las modificaciones de histonas presentes en otro tipo de heterocromatina como es la heterocromatina constitutiva propia de zonas ricas en ADN repetido como son los centrómeros y las regiones próximas a los telómeros. La localización y la composición de este tipo de heterocromatina es universal en todos los tipos celulares [131], [147].

○ *Eucromatina: Transcripción génica*

Antes de comenzar, aclarar que los nombres abreviados de las modificaciones de histonas que aparecerán en el texto se interpretan de la siguiente forma: La letra H seguida de un nº hace referencia a la histona en concreto donde se produce la modificación. Por otro lado, la letra K seguida de un nº hace referencia específicamente al residuo de lisina que es modificado. Las abreviaciones me o ac hacen referencia a metilación o acetilación respectivamente. Por último, el nº que sigue a me o ac se refiere al grado de metilación o acetilación (mono, di, tri).

La transcripción de un gen está asociada con el reclutamiento de una serie de modificaciones sobre los elementos reguladores del mismo. Estos elementos reguladores son los promotores y los enhancers. Los promotores son elementos de ADN que regulan la expresión génica en *cis*. Es difícil determinar los límites de un promotor así que se considera a la región +/-2kb desde el punto de inicio de la transcripción como el promotor [143]. Los enhancers son secuencias de ADN que contienen secuencias de unión a proteínas que modulan la expresión génica de una forma independiente de la distancia y de su orientación [148]. La transcripción génica se asocia con el enriquecimiento de H3K4me3 sobre el promotor del gen en cuestión. Esta modificación presenta una correlación positiva con la expresión génica [136], [137], [140], [141], [149]. Asimismo, la transcripción génica también se asocia con el reclutamiento de H3K9ac sobre el promotor pues la presencia de esta marca y H3K4me3 está relacionada con el reclutamiento del complejo de preiniciación de la ARN Polimerasa II [150]. Al igual que H3K4me3, H3K9ac también correlaciona con la expresión génica [151]. Por último, genes activos presentan una serie acetilaciones en una serie de residuos de lisina de las histonas H2A, H2B, H3 y H4 en su promotor [140], [142].

Además de las modificaciones presentes sobre el promotor, la transcripción génica también se asocia con el enriquecimiento de una serie de modificaciones sobre el enhancer que puede controlar la actividad en *trans* del gen en cuestión. Un enhancer activo se caracteriza por la presencia sobre todo de dos modificaciones de histonas H3K4me1y H3K27ac [135], [152], [153]. Además se reclutan acetilaciones de la histona 3

tales como H3K9ac y H3K14ac así como también la acetilación de la histona H4 [135], [151].

Aquellos genes que se expresan presentan enriquecimientos de una serie de modificaciones sobre el cuerpo del gen. Fundamentalmente, la modificación que aparece enriquecida sobre el cuerpo del gen de genes transcripcionalmente activos es H3K36me<sub>3</sub>, la cual correlaciona con la actividad del gen [137]. Además, ésta se puede presentar acompañada de otras modificaciones como H4K20me<sub>1</sub>, H2BK5me<sub>1</sub> [154] y H3K79me<sub>1</sub> [155]. Por otro lado, H3K36me<sub>3</sub> se encuentra enriquecida en los exones respecto a los intrones [154]–[156]. Parece ser que H3K36me<sub>3</sub> puede ser importante para un correcto splicing del ARNm naciente. De hecho, H3K36me<sub>3</sub> se encuentra más enriquecida sobre los exones que no forman parte de isoformas resultantes de splicing alternativo [154]–[156]. La conservación de este proceso entre especies confirma la importancia de la presencia de esta modificación para el correcto splicing [156].

De esta forma, genes que presentan esta combinación de modificaciones de histonas sobre los distintos elementos reguladores tales como promotores y enhancers así como sobre el propio marco abierto de lectura suelen ser genes transcripcionalmente activos. El reclutamiento de estas modificaciones para la expresión génica es universal tiene lugar tanto en células embrionarias como en otros tipos de células pluripotentes así como en células diferenciadas [140], [157].

○ *Heterocromatina facultativa: Represión génica*

En células embrionarias, los promotores de genes silenciados, que codifican proteínas clave reguladoras del proceso de desarrollo, presentan el doble marcaje H3K4me<sub>3</sub>/H3K27me<sub>3</sub>. Estos promotores son conocidos como bivalentes [134], [158]. Los genes controlados por promotores bivalentes se expresan a niveles muy bajos o no se expresan, por tanto, existe una correlación negativa con la expresión génica [134], [159], [160]. Estos genes se resuelven, a medida que avanza el proceso de desarrollo, en una dirección u otra, perdiendo H3K4me<sub>3</sub> y permaneciendo reprimidos o bien perdiendo H3K27me<sub>3</sub> y activándose [134]. Igualmente, estos genes bivalentes pueden permanecer en estado bivalente en función del fenotipo celular final [141]. Por otro lado, estos genes bivalentes, en algunos casos, también pueden incorporar H3K9me<sub>3</sub> para asegurar el silenciamiento génico y mantener el estado de pluripotencia estableciendo una nueva capa de silenciamiento génico [161]. De hecho, la pérdida de H3K9me<sub>3</sub> sobre estos promotores trivalentes conlleva un aumento de la expresión génica [161]. Por otro lado, también existe un porcentaje bajo de promotores que presentan únicamente H3K27me<sub>3</sub> en células embrionarias [159], [160]. Cabe destacar, que la forma en que se produce el silenciamiento

mediado por H3K27me3 en células embrionarias, se basa en la generación de dominios de H3K27me3 de aproximadamente 5Kb [141], que se pueden expandir hasta 100 kb como en el caso de los genes *HOX* [160].

En células embrionarias de ratón, también existen una serie de enhancers que permanecen inactivos y que, por tanto, no dirigen la expresión de sus genes objeto de control. Estos enhancers se caracterizan por presentar H3K4me1 como única modificación de histona. De hecho, la presencia únicamente de esta marca correlaciona negativamente con la expresión génica de aquellos genes que supuestamente están bajo su control [152]. Además, los genes bajo el control de estos enhancers son genes específicos de etapas más tardías del proceso de desarrollo o bien específicos de linaje celular [152]. En células embrionarias humanas, los enhancers inactivos también presentan H3K27me3 además de H3K4me1. De la misma forma, estos enhancers dirigen la expresión de genes importantes para la siguientes etapas del proceso de desarrollo como gastrulación, formación de capas germinales etc [153]. Tanto enhancers enriquecidos en H3K4me1 como enhancers enriquecidos en H3K4me1 y H3K27me3 son activos en etapas posteriores del desarrollo embrionario cuando se produce la especificación de linaje celular o bien en células más diferenciadas. Por otro lado, estos enhancers bivalentes H3K4me1-H3K27me3 se han encontrado también en células diferenciadas como IMR90 donde, seguramente, su función sea la de inhibir la expresión de genes que no son propios de tal linaje [162]. Por tanto, la represión en células embrionarias está asociada a la presencia de genes bivalentes, trivalentes, así como genes reprimidos por H3K27me3 y mediante enhancers bivalentes o quiescentes que controlan la actividad de genes importantes en etapas posteriores del desarrollo.

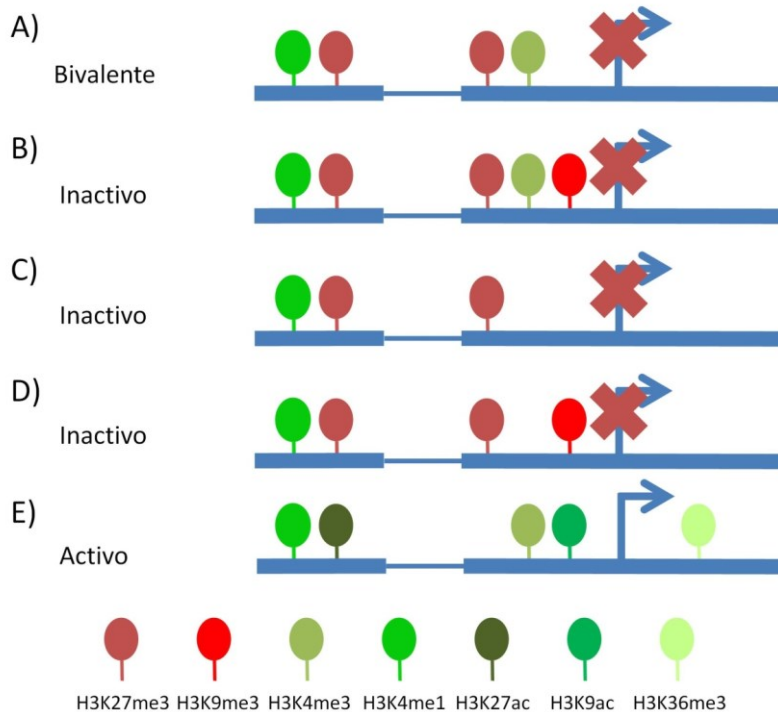
Cuando se produce la diferenciación celular, los promotores bivalentes se pueden resolver de distintas formas. En lo que se refiere a represión, estos promotores bivalentes en células embrionarias de ratón pueden evolucionar hacia un estado en el que presenten únicamente H3K27me3 o bien también pueden permanecer en estado bivalente aunque el número de promotores bivalentes en células diferenciadas es menor [134], [137], [141]. Por otro lado, algunos genes bivalentes también se pueden resolver hacia el doble marcaje H3K27me3/H3K9me3 añadiendo una nueva capa de silenciamiento génico sobre genes reguladores del desarrollo asegurando así el correcto mantenimiento del linaje celular [163]. Además, en aquellos promotores bivalentes que se resuelven hacia el enriquecimiento únicamente de H3K27me3, se produce una expansión del dominio de H3K27me3 que cubre el promotor de los genes respecto a las células embrionarias [163]. Además estos dominios de expansión de H3K27me3 en MEFs (del inglés *Mouse Embryonic*



*Fibroblasts*) se correlacionan con represión génica [164]. Estos dominios de H3K27me<sub>3</sub>, como se menciona en el apartado anterior, no son específicos de células diferenciadas sino que también se encuentran en células embrionarias [159], [160]. Sin embargo, de media estos dominios son más extensos en células diferenciadas y ocupan un mayor porcentaje del genoma [163]. De hecho, la fracción de genoma ocupado por H3K27me<sub>3</sub> se triplica en células IMR90 así como en otros tipos celulares diferenciados como son células T CD4+, NHEK, K562 y HUVEC respecto a células madre embrionarias humanas. Cabe destacar que en cada tipo celular los genes englobados por estos dominios de H3K27me<sub>3</sub> son aquellos que no son activos en el linaje celular al que pertenece dicho tipo celular [163]. Estos dominios de expansión de H3K27me<sub>3</sub> se observan también cuando se diferencian células embrionarias humanas a células madres mesenquimales [162]. También fue apreciado por Zhu et al., [165] observando un incremento del enriquecimiento del H3K27me<sub>3</sub> en el genoma comparando células embrionarias humanas con diversos tipos de células diferenciadas y tejidos humanos. Este fenómeno no ocurre sólo cuando existe silenciamiento a través de H3K27me<sub>3</sub> sino que también es un fenómeno propio del silenciamiento por H3K9me<sub>3</sub>, ya que se observa también un incremento de la extensión de los dominios de H3K9me<sub>3</sub> en células IMR90 comparados con células embrionarias humanas [163].

Parece ser que en células diferenciadas existe un aumento de la cromatina inactiva o heterocromatina para reprimir genes específicos de programas celulares de pluripotencia [163]. En células diferenciadas de ratón (cerebro e hígado) se generan dominios de H3K9me<sub>2</sub> cuyo porcentaje de ocupación del genoma es mayor comparado con el porcentaje mínimo de ocupación en células embrionarias. Son dominios que suelen contener genes reprimidos importantes para programas de diferenciación celular diferentes [166]. Sin embargo, este hallazgo se encuentra en debate [167]. En el proceso de diferenciación de células madre hematopoyéticas humanas se observa como existe un aumento de la heterocromatina y una redistribución de la misma en las células más diferenciadas hacia la periferia nuclear. Además, existe una colocalización entre la heterocromatina marcada con H3K9me<sub>3</sub> y la lámina nuclear [168]. Esto también había sido observado por Zhu et al., [165] ya que en células diferenciadas los dominios de cromatina ricos en H3K9me<sub>3</sub> y H3K27me<sub>3</sub> se encontraban en contacto con la lámina nuclear sugiriendo que estas regiones se podrían corresponder con LADs (del inglés Lamina Associated Domains). Según Guelen et al., [169], estos dominios se caracterizan por ser pobres en genes, presentar baja expresión génica, marcas como H3K27me<sub>3</sub> y H3K9me<sub>2</sub> enriquecidas y además contienen elementos repetidos como LINEs en el interior y SINEs en los bordes. En resumen, en células diferenciadas existe un aumento de la heterocromatina

bien a través de dominios de H3K27me3 o bien a través de dominios de H3K9me3 que delimitan regiones de represión transcripcional.

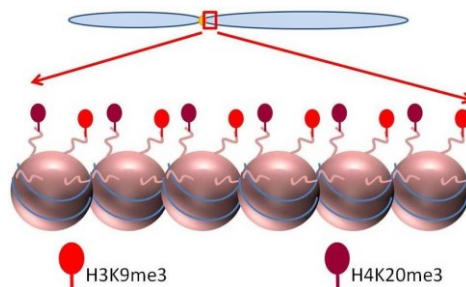


**Figura 5: Epigenética de genes activos y reprimidos en células embrionarias y diferenciadas. A)** Gen inactivo en células embrionarias debido a la bivalencia de su promotor (H3K27me3/H3K4me3) y de su enhancer (H3K4me1/H3K27me3). **B)** El gen en A) puede pasar a ser trivalente (H3K27me3/H3K4me3/H3K9me3) reforzando el silenciamiento en células embrionarias. **C)** El gen en A) puede pasar a ser inactivo en células diferenciadas perdiendo H3K4me3 y probablemente manteniendo el estado bivalente del enhancer. **D)** El gene en A) también puede ser silenciado en células diferenciadas con la pérdida de H3K4me3 y la ganancia de H3K9me3 y probablemente manteniendo también el estado de bivalencia de su enhancer. **E)** El gen en A) puede pasar a ser activo en células diferenciadas mediante la pérdida de H3K27me3 sobre su promotor y su enhancer, la ganancia de H3K9ac sobre su promotor y H3K27ac sobre su enhancer, y la ganancia de H3K36me3 sobre el marco abierto de lectura.

○ *Heterocromatina Constitutiva*

La heterocromatina constitutiva, generalmente presente en centrómeros está conformada mayoritariamente por secuencias satélite, que puede ser fácilmente distinguida tiñendo con DAPI [170], [171]. Los centrómeros constituyen un punto de constricción citológicamente visible cuya función es ser el punto de unión entre dos

cromátidas hermanas y asegurar una correcta distribución de las mismas durante la mitosis ya que éstos son el punto a través del cual los cromosomas se unen al anillo mitótico [172]. En células embrionarias de ratón, esta heterocromatina se caracteriza casi por una total represión aunque existe una mínima transcripción desde estas secuencias satélite [171]. Además, se caracteriza por la presencia de H3K9me3 y de H3K27me1 [170]. En caso de ausencia de H3K9me3, puede existir un reclutamiento de H3K27me3 [170]. En células humanas diferenciadas, se observa que H3K9me3 está asociada a regiones pericentroméricas [173]. Por otro lado, existe otra modificación presente en la heterocromatina pericentromérica como es la H4K20me3 que se localiza exclusivamente en los centrómeros [173], [174] y cuya presencia está ligada a la presencia de H3K9me3 [174].



**Figura 6: Representación de la heterocromatina constitutiva presente en las regiones pericentroméricas.** H3K9me3 and H4K20me3 son las modificaciones de histonas que mayoritariamente participan favorecen en la compactación de la cromatina en estas regiones.

Por otro lado, los telómeros de los cromosomas en eucariotas también son considerados como regiones de cromatina constitutiva. Éstos, en general, consisten en repeticiones en tándem que a través de una ribonucleoproteína, utilizando su propio molde, son alargados a lo largo de sucesivas divisiones celulares evitando así la pérdida de información genética [175]. Los telómeros también presentan enriquecimientos de H3K9me2 y H3K9me3 lo cual contribuye a la compactación de la región evitando así un alargamiento aberrante de los mismos en células embrionarias de ratón [176].

#### 1.4.4. Modificaciones de histonas más relevantes en esta tesis

##### 1.4.4.1. Modificaciones asociadas a cromatina activa

###### ○ Trimetilación de la lisina 4 de la histona 3 (H3K4me3)

Según las evidencias presentadas anteriormente, se deduce que la H3K4me3 es fundamentalmente una modificación de histona asociada a la transcripción activa de un

gen. Desde su descubrimiento en levadura se ligó a genes activamente transcritos [177]. De hecho, un mayor enriquecimiento de la misma en los promotores correlaciona con una mayor actividad transcripcional del gen [136], [137]. Es una modificación que se localiza en el promotor fundamentalmente alrededor del sitio de inicio de la transcripción siempre cerca del extremo 5' del gen [137], [178]. Además, es una modificación casi exclusiva de promotores aunque ciertamente en algún estudio se ha podido detectar también en enhancers [153]. Esta modificación está catalizada por una de las seis enzimas homólogas a SET1 en levadura que existen en mamíferos (SETD1A, SETD1B, MLL1, MLL2, MLL3 y MLL4) [179]. En levadura, SET1 se asocia con la subunidad de iniciación de la polimerasa II y la marca es depositada cotranscripcionalmente [179].

○ *Acetilación de la lisina 9 de la histona 3 (H3K9ac)*

En general, la acetilación de las histonas está ligada a activación transcripcional. En *Drosophila*, los genes más expresados presentan un mayor enriquecimiento de H3ac y H4ac [180]. Además, la acetilaciones globales de H3 y H4 correlacionan con la presencia de H3K4me3 en los mismos genes [180]. En células humanas, la presencia de H3K9ac favorece el reclutamiento del factor de transcripción de TFIID que es esencial para el inicio de la transcripción [181]. La presencia de esta marca sobre el promotor de los genes correlaciona con la iniciación de la transcripción [136] así como con la actividad transcripcional de los mismos [151]. Esta modificación también se detecta sobre enhancers activos [135], [151]. Por otra parte, esta modificación se concentra en torno al sitio de inicio de la transcripción dando lugar a un pico a cada lado del mismo [151]. Esta modificación es depositada por una familia de enzimas denominada histona acetil transferasas (HATs del inglés Histone Acetyl Transferases). Existen dos clases de HATs, clase A y clase B cuyos miembros añaden grupos acetilo a un amplio espectro de residuos [133]. Esta modificación está íntimamente ligada a la presencia de H3K4me3 sobre el promotor de los genes ya que proteínas que reconocen H3K4me3 se encuentran en los complejos HATs y cuya delección conlleva la pérdida de H3K9ac [179].

○ *Acetilación de la histona 4: H4K5ac, H4K8ac, H4K12ac, H4K16ac*

Al igual que la acetilación global de la histona 3, la acetilación global de la histona 4 se encuentra relacionada con la actividad transcripcional. En *Drosophila*, genes enriquecidos en H4ac presentan una mayor actividad transcripcional [180]. En efecto, la presencia de las acetilaciones de la histona 4 como son H4K5ac, H4K8ac, H4K12ac y H4K16ac sobre el promotor de los genes correlaciona con la actividad transcripcional [140], [182]. Por otro lado, H4K16ac también está asociada a enhancers activos junto con H3K4me1 y H3K27ac en células madre embrionarias de ratón [182]. De la misma forma, H4K16ac está ligada a

enhancers activos en células diferenciadas humanas [135]. H4ac es depositada mediante HATs, específicamente los HATs de clase B que catalizan la reacción sobre los residuos K5 y K12 [133].

En resumen, H3K4me3, H3K9ac y H4ac son modificaciones de histonas relacionadas con activación transcripcional.

#### 1.4.4.2. Modificaciones asociadas a cromatina inactiva

##### ○ Trimetilación de la lisina 9 de la histona 3 (H3K9me3)

Esta modificación está siempre ligada a represión. De hecho, es la marca por excelencia de la heterocromatina constitutiva [170] donde la transcripción es casi inexistente [171]. Es una modificación de histona que contribuye durante la diferenciación celular al silenciamiento de grupos de genes a través de la generación de dominios de dicha modificación [163]. Asimismo contribuye a reforzar el silenciamiento génico por parte de H3K27me3 en promotores bivalentes en células embrionarias [161]. Igualmente, promotores bivalentes de células embrionarias que controlan reguladores del desarrollo se pueden resolver hacia el doble marcaje H3K9me3/H3K27me3 en células diferenciadas [163]. También está implicada en el silenciamiento de genes sometidos a impronta génica como *IGFR2* [183], de genes silenciados en cáncer [184] y de genes específicos de línea germinal en células embrionarias [185].

Existen tres enzimas del grupo de las lisina metil transferasas (KMTs del inglés Lysine Methyl Transferases) que pueden catalizar el depósito de H3K9me3 las cuales son Suv39h1/2 (KMT1A/B) y SETDB1 (KMT1E) [179]. El depósito de H3K9me3 que media la formación de heterocromatina pericentromérica es realizado mediante Suv39h1/2 [170] mientras que el depósito sobre los promotores de genes parece estar mediado por SETDB1 [161], [183], [185]. Sin embargo, Suv39h1/2 también puede contribuir al silenciamiento génico como en la represión de genes en fase S tales como la ciclina D1 [186], [187]. De la misma forma, SETDB1 también puede contribuir a la formación de la heterocromatina pericentromérica en neuronas murinas [188]. El hecho de que exista un solapamiento en sus funciones se puede deber a que una parte de la población de estas enzimas forman un complejo con otras metil transferasas como son G9a y GLP que catalizan la reacción de monometilación (H3K9me1) y de dimetilación (H3K9me2) [189].

○ *Trimetilación de la lisina 27 de la histona 3 (H3K27me3)*

Esta modificación de histona también está ligada a represión génica. De hecho, su presencia en los promotores y a lo largo del cuerpo del gen correlaciona con el silenciamiento génico en células diferenciadas y pluripotentes [137], [173], [190]. Los promotores bivalentes, en los que H3K27me3 se encuentra en células embrionarias se caracterizan por una baja expresión [134], [159], [160]. Excepcionalmente, también puede formar parte de la heterocromatina constitutiva en ausencia de H3K9me3 en las regiones pericentroméricas [170] y se localiza también en las regiones subtelo méricas [173]. En células embrionarias humanas y murinas, esta marca co-localiza mayoritariamente con el complejo PRC2 (del inglés Polycomb Repressive Complex 2) en los promotores de genes que son importantes en las distintas etapas del desarrollo tales como neurogénesis, organogénesis, hematopoiesis o morfogénesis [191], [192]. Este complejo es el encargado del depósito de esta marca en dichos promotores [193], [194]. Además, la depleción de uno de los componentes de este complejo, la subunidad Eed, conlleva una pérdida global de H3K27me3/H3K27me2 en el genoma de ratón [195]. De hecho, células embrionarias mutantes para Eed muestran un incremento de la expresión de genes que estaban ocupados por PRC2 y también por H3K27me3 [191]. Este complejo está formado por un núcleo de 4 proteínas: Ezh1/2, Suz12, Eed y RbAp46/48. La presencia de estas proteínas son fundamentales para el desarrollo embrionario temprano, ya que ratones deficientes para Eed, Suz12 y Ezh2 presentan letalidad embrionaria, seguramente debido al tipo de genes que regulan [196]. De hecho, los genes regulados por este complejo, tienden a estar reprimidos en células embrionarias y tienden a expresarse en células diferenciadas [191], [192]. La distribución del complejo sobre aquellos genes que regula es bastante particular ya que se observan dos patrones, un primer patrón donde el complejo se encuentra posicionado en torno al sitio de inicio de la transcripción [197] y un segundo patrón otro patrón donde el complejo puede llegar a ocupar dominios de hasta 35 kb que engloban al gen diana [192] como ocurre con los genes *HOX* donde PRC2 se distribuye a lo largo de 45 kb [197]. De esta forma se explica la típica distribución que adquiere H3K27me3 en algunos casos donde se expande a lo largo de dominios que pueden alcanzar dichas distancias tanto en células embrionarias como diferenciadas [160], [163], [164].

### **1.5. Regulación epigenética de elementos móviles en mamíferos**

○ *Retrovirus endógenos*

En general, los retrovirus endógenos se encuentran asociados con modificaciones de histonas represivas. De hecho, se encuentran decorados por H3K9me3 en el genoma humano [141], [142]. Igualmente, en células embrionarias de ratón, estos retroelementos

estás enriquecidos en H3K9me3 [198]. En células embrionarias de ratón, el depósito de esta modificación parece estar mediado por SETDB1 ya que la ausencia de esta enzima lleva asociada la sobreexpresión de los retrovirus endógenos de clase I y clase II debido a la pérdida de dicha marca sobre los mismos [185], [199]. Por otro lado, Suv39h1/2 también reprimen a los retrovirus endógenos mediante H3K9me3 [117]. Además parece ser que la represión mediante H3K9me3 es el mecanismo encargado de mantener inactivos estos retrotransposones durante el desarrollo embrionario temprano ya que células embrionarias de ratón deficientes para las enzimas DNMT3A/B (del inglés *de novo* DNA methyltransferase 3 A/B) y DNMT1 (del inglés *de novo* DNA methyltransferase 1) no reactivan la expresión de estos retroelementos [185], [199]. Por otro lado, células embrionarias de ratón deficientes para la enzima DNMT1 reactivan la expresión de un retrovirus de clase II (IAP) a medida que se diferencian [200]. Por tanto, la regulación de los retrovirus endógenos en el estado de pluripotencia parece recaer en gran medida sobre H3K9me3. El mecanismo molecular que reprime los retrovirus endógenos a través del depósito de H3K9me3 se desencadena con el reconocimiento de la secuencia denominada PBS (del inglés Primer Binding Site) del retrovirus endógeno, la cual es importante para la replicación, por parte de un complejo represor formado por miembros de la familia KRAB-ZFPs (del inglés Krüppel-associated box zinc-finger proteins) y KAP1 (del inglés KRAB-associated protein 1) [201], [202]. Posteriormente, la interacción de KAP1 con la histona metil transferasa SETDB1 [203] finaliza con la represión del retrovirus endógeno mediante el depósito de H3K9me3 [199], [204], [205].

Además, los retrovirus endógenos también parece estar regulados por H3K27me3, ya que células embrionarias de ratón deficientes en PRC1 (del inglés Polycomb Repressive Complex 1) y PRC2 reactivan transcripcionalmente retrovirus endógenos como el MLV (del inglés Murine Leukemia Virus) lo cual va acompañado del aumento del nº de copias de este retroelemento [206]. Por otro lado, el cultivo de células embrionarias de ratón en medio 2i+Vit C provoca la desmetilación global del genoma que va acompañada de una reactivación transcripcional de los retrovirus endógenos. Estos retroelementos son silenciados mediante el depósito de H3K27me3 [198].

Por último, retrovirus endógenos de clase II como IAP y ETn aparecen enriquecidos en H4K20me3, la cual es una modificación asociada principalmente a heterocromatina pericentromérica donde predominan secuencias satélites mayores ricos en A/T [141], [174]. Además los elementos IAP también muestran un enriquecimiento de H3K64me3 sobre su promotor en células embrionarias de ratón. Esta modificación se encuentra asociada con heterocromatina pericentromérica donde colocaliza con H3K9me3. Además,

a lo largo del genoma, los dominios de ocupación de esta marca se solapan perfectamente con los dominios de H3K9me3 [207].

En principio, la presencia de modificaciones de histonas asociadas a cromatina activa sobre los elementos móviles no debería ser muy frecuente debido a la amenaza que representan para el genoma. Sin embargo, existen datos sobre la relación de retrovirus endógenos con modificaciones asociadas a actividad transcripcional. De hecho, la eliminación genética de una lisina desmetilasa (KDM1A) resultó en el incremento de la expresión de retrovirus endógenos de clase III (MERVLs) acompañada de un enriquecimiento de H3K4me3 y H3K27ac en la vecindad del promotor de este elemento [208]. Por otro lado, la desmetilación casi total del genoma de ratón en células embrionarias mediante su cultivo en medio 2i+Vit C va acompañada de una reactivación transcripcional de los retrovirus endógenos ganando H3K4me3 sobre su promotor [198]. Además, la depleción del represor KAP1 promueve la pérdida de H3K9me3 y la ganancia de H3K4me3 y H3K27ac sobre los retrovirus endógenos que funcionan como enhancer activando la expresión de genes cercanos [209]. Por tanto, en determinadas circunstancias los retrovirus endógenos están regulados por modificaciones asociadas con actividad transcripcional.

○ *LINE-1*

Los elementos LINE-1 a lo largo del genoma humano están asociados con H3K9me3 [142]. Asimismo, en células embrionarias humanas, se detectan enriquecimientos de H3K9me3 sobre elementos LINE-1 pertenecientes a subfamilias evolutivamente más antiguas [210]. Además, en células embrionarias de ratón, estos elementos están enriquecidos en H3K9me3 de un modo dependiente de su longitud siendo los elementos de copia completa los más enriquecidos [117], [198]. El depósito de esta marca sobre estos elementos en células embrionarias de ratón está mediado por las enzimas Suv39h1/2 ya que la ausencia de éstas correlaciona con la pérdida de H3K9me3 sobre la secuencia de dichos elementos y con su reactivación transcripcional [117], [211]. Sin embargo, células diferenciadas deficientes para Suv39h1/2 no reactivan transcripcionalmente los elementos LINE-1 con lo que este mecanismo de silenciamiento mediado por Suv39h1/2 podría ser específico del desarrollo embrionario temprano [117]. Al contrario de lo que ocurre con los retrovirus endógenos, se desconocen cuáles son los intermediarios que permiten el reclutamiento de estas enzimas para la represión de estos retroelementos a través de H3K9me3.

Por otra parte, los elementos LINE-1 se encuentran en dominios del genoma ricos en H3K27me3 tanto en humano como en ratón [154], [164]. El silenciamiento de los



elementos LINE-1 en células embrionarias de ratón, cuya expresión es inducida en unas condiciones de desmetilación general cuando son cultivadas en medio 2i+VitC, es dependiente del depósito de H3K27me3 [198]. Además, H3K64me3, una modificación asociada con heterocromatina pericentromérica, también aparece enriquecida en el promotor de LINE-1 en células embrionarias de ratón [207].

Al igual que los retrovirus endógenos, también existen datos que relacionan los elementos LINE-1 con la presencia de modificaciones de histonas asociadas a actividad transcripcional en determinadas circunstancias. De hecho, existen datos sobre la presencia de H3K4me3 justo aguas arriba del promotor de inserciones de copia completa de LINE-1 en células de cáncer de mama cuyos transcritos son detectados por secuenciación masiva de ARN [71]. Sorprendentemente, estos promotores también presentan justo aguas arriba además una modificación de histona propia de enhancers como H3K27ac [71]. Del mismo modo que los retrovirus endógenos, los elementos LINE-1 también son reactivados transcripcionalmente cuando células embrionarias de ratón son cultivadas en medio 2i+Vit. C lo cual provoca una desmetilación global del genoma. Esta reactivación va asociada al enriquecimiento de H3K4me3 sobre el promotor de los elementos LINE-1 [198].

Todos estos datos sobre regulación epigenética de los elementos LINE-1 provienen de estudios que analizan las copias ya presentes en el genoma humano o murino. Sin embargo, se conoce muy poco acerca de la regulación epigenética de las inserciones de LINE-1 que ocurren *de novo* en el genoma ya que tan sólo existe un estudio que aborde esta cuestión. De hecho, un estudio muy relevante en el campo de la biología de los retroelementos del Dr. José Luís García Pérez, demostró que la inserciones artificiales *de novo* de LINE-1 se silencian epigenéticamente mediante desacetilación de histonas después de la integración en un genoma pluripotente. Estas inserciones eran reactivadas cuando las células eran tratadas con TSA (Tricostatina A) el cual es un pan inhibidor de las histonas desacetilasas [60]. Esta reactivación era medida como porcentaje de células que expresan el gen *GFP*, que fue el reportero usado en este ensayo de retrotransposición [12], [115]. Además, esta reactivación estaba acompañada de un aumento de H4ac y una disminución de H3K9me2 sobre la secuencia de una de las inserciones analizadas. Además también se demostró que este es un mecanismo específico de células pluripotentes. Estos datos sugerían la implicación de modificaciones de histonas como mecanismo epigenético de control de las nuevas inserciones de LINE-1 en un genoma pluripotente.



# OBJETIVOS



## **2. Objetivos de la tesis:**

El objetivo global de la tesis es el estudio de nuevos mecanismos implicados en la regulación del retroelemento humano LINE-1. Para ello se proponen los siguientes dos objetivos:

### **2.1. Estudio del papel del microprocesador en la regulación del retroelemento LINE-1.**

La proteína DGCR8, que forma parte del complejo microprocesador que interviene en la ruta de biogénesis de microRNAs, se une físicamente al ARNm del retroelemento LINE-1 [124]. Por ello, es posible que el microprocesador pueda tener un impacto sobre el ciclo de vida del retroelemento LINE-1. Por tanto, se propone:

1. Determinar el papel regulador del microprocesador (Drosha/DGCR8) sobre la expresión y la movilidad del retroelemento LINE-1.

### **2.2. Estudio de la regulación epigenética de las inserciones *de novo* de LINE-1 en un genoma pluripotente y el impacto epigenético de las mismas en el genoma adyacente.**

Todos los estudios que proporcionan datos sobre regulación epigenética del retroelemento LINE-1 analizan inserciones ya existentes en el genoma [117], [142], [154], [164], [198], [207]. Por otro lado, tan sólo el estudio del Dr. José Luís García Pérez abordó la cuestión de la regulación de las inserciones *de novo* de LINE-1. De hecho, demostró que el silenciamiento epigenético de las nuevas inserciones de LINE-1 está mediado por desacetilación de histonas en células pluripotentes después de su integración [60]. Por tanto, existen pocos datos sobre la regulación epigenética de las inserciones *de novo* de LINE-1. Además es importante conocer esta regulación en un contexto de pluripotencia que mimetice el desarrollo embrionario temprano momento durante el cual se producen la mayoría de las nuevas inserciones *de novo* de LINE-1 [77], [78]. Asimismo, se desconoce si las modificaciones de histonas asociadas con los elementos LINE-1 ya presentes en el genoma tales como H3K9me3 [117], [142], H3K27me3 [138], [198] o H3K4me3 [71], [198] están implicadas en la regulación de una nueva inserción de LINE-1. Igualmente, también se desconoce el si una nueva inserción de LINE-1 puede tener un impacto epigenético sobre el genoma adyacente provocando cambios en el enriquecimiento de determinadas modificaciones de histonas. Por tanto, se propone:

1. Establecer líneas clonales con una inserción *de novo* de LINE-1 en células pluripotentes.
2. Analizar el estado de la cromatina en las inserciones *de novo* de LINE-1.
3. Analizar el posible impacto epigenético de las inserciones *de novo* de LINE-1 sobre el genoma adyacente.
4. Estudiar el estado de la cromatina en inserciones endógenas *de novo* de LINE-1 en un genoma pluripotente.

# MATERIALES Y MÉTODOS





### **3. Materiales y métodos**

#### **3.1. Líneas celulares**

- PA-1: línea de teratocarcinoma humano de ovario aislada del líquido ascítico de una paciente con un teratoma de ovario maligno recurrente [212]. Presenta características de células de carcinoma embrionario con cierta capacidad pluripotente bajo ciertas condiciones de cultivo celular [212]. Presenta un cariotipo diploide muy estable a lo largo de 400-500 pases (aproximadamente entre 400-500 pases) [213]. Ceditas por el laboratorio del Dr. John Moran (University of Michigan, EEUU).
- H9: línea de células madre embrionarias humanas derivada de la masa celular interna de un blastocisto humano femenino [214]. Obtenidas del Banco Andaluz de células madre.
- HeLa: Células de carcinoma de cuello de útero. Es la primera línea celular que se consiguió aislar y mantener en cultivo en 1951. Ceditas por el laboratorio del Dr. John Moran (University of Michigan, EEUU).
- Células embrionarias de ratón V6.5 fueron compradas a Thermo Scientific (MES1402) y sus derivadas mutantes (DGCR8<sup>-/-</sup>) fueron compradas a Novus Biologicals (NBA1-19349).

#### **Mantenimiento de líneas celulares**

Las células HeLa fueron crecidas en DMEM (Dulbecco's Modified Eagle Medium) (Gibco) suplementado con 10% de suero bovino fetal (Gibco) y penicilina-estreptomomicina 1x (Gibco). Las células fueron subcultivadas utilizando tripsina (Gibco) al 0.25% cuando se encontraban en torno al 90% de confluencia.

Las células embrionarias de ratón fueron crecidas en DMEM-high glucose (Gibco) suplementado con 15% de suero bovino fetal (Gibco). Estas células eran cultivadas en placas o frascos de cultivo tratadas con gelatina al 0,1% durante 10 min a 37°C (Sigma-Aldrich). Las células eran subcultivadas utilizando tripsina (Gibco) 0,05%.

Las células PA-1 fueron crecidas en medio MEM (Minimum Essential Media, Gibco) suplementado con 10% de suero bovino fetal descomplementado (Gibco) (56°C 30min), penicilina-estreptomomicina 1x (Gibco) y aminoácidos no esenciales 1x (Gibco). Las células fueron subcultivadas cuando alcanzaban una confluencia del 95% usando tripsina (Gibco) al 0,25%.

Las células PA-1 diferenciadas fueron crecidas en medio MEM suplementado con 10% suero de reemplazo (Gibco), ácido retinoico 1µM (Sigma-Aldrich), aminoácidos no

esenciales 1x (Gibco) y penicilina-estreptomina 1x (Gibco). Las células fueron subcultivadas cuando alcanzaban una confluencia del 95% usando tripsina al 0,25% (Gibco). En este caso era necesario el tratamiento del frasco de cultivo con matrigel (Corning) durante 1 h a temperatura ambiente.

Las H9 fueron crecidas en medio condicionado (suministrado por el Biobanco del Sistema Sanitario Público Andaluz) suplementado con FGF (8ng/ml) sobre matrigel (Corning). Las células fueron subcultivadas cuando alcanzaban una confluencia del 95% despegando las mismas con ayuda de una rasqueta.

Las células H9 fueron adaptadas a un medio denominado E8 que permite una rápida expansión de las mismas [215] para realizar los experimentos de ChIP-qPCR. Para obtener este medio, en primer lugar se parte de medio E4 cuyos componentes son los siguientes: MEDIO DMEM/F12 (Life Technologies), 0.64g L-ascorbic Acid 2-Phosphate Mg<sup>2+</sup> salt (Sigma-Aldrich), 17,43g Sodium Bicarbonate (Sigma-Aldrich), 200µl Sodium Selenite (0,7mg/ml) (Sigma-Aldrich), 10ml NaCl (5M) (Sigma-Aldrich). Posteriormente, se completa con H<sub>2</sub>O Miliq hasta 1 litro. Se hacen alícuotas de 50 ml y se congelan a -80°C. A partir de una alícuota descongelada se prepara el medio E8 cuyos componentes son los siguientes: 50 ml E4; 500µl FGF2 (100µg/ml) (Peprotech); 5µl TGFβ (180µg/ml) (Peprotech); 500µl Holo-transferrin (10,6mg/ml) (Sigma-Aldrich), 1 ml Insulin (Sigma-Aldrich). Dentro de cabina de flujo laminar, se mezclan todos los componentes en una unidad de filtración de 500ml con 450ml de agua Miliq estéril y posteriormente se filtran.

Las células H9 crecidas en medio E8 se subcultivaban cuando alcanzaban una confluencia del 95% mediante el uso de DPBS 1X (Gibco) suplementado con EDTA 0,5mM (Invitrogen) incubando 5 min a 37°C. Al pasar las células, el medio E8 era suplementado con iROCK 10 µM (Sigma-Aldrich) hasta el siguiente cambio de medio que se producía al día siguiente.

Todas las líneas celulares fueron negativas para el test de micoplasma que se realizó de forma rutinaria.

### *Congelación y descongelación de líneas celulares*

Las líneas celulares tuvieron que ser congeladas en multitud de ocasiones. Para congelar las líneas celulares se procedió como sigue:

- Para células PA-1, se preparó una solución de congelación que contiene suero bovino fetal descomplementado y DMSO (Sigma-Aldrich) al 10%. Seguidamente, las células se despegan del frasco de cultivo utilizando tripsina al 0.25%. Posteriormente las células

se centrifugan a 1200 rpm durante 5 min. Después, el medio se aspira completamente y el pellet es resuspendido en solución de congelación. Posteriormente, las células se trasvasan a un criotubo (Thermo Scientific) que es rápidamente congelado a  $-80^{\circ}\text{C}$ . Este proceso de congelación es exactamente igual para las células embrionarias de ratón así como para las células HeLa, preparando la solución de congelación con suero bovino fetal y DMSO 10%.

- Para las células H9 crecidas en medio condicionado en un frasco de  $25\text{cm}^2$  se procede como sigue. En primer lugar, se despegan las células con ayuda de una rasqueta y éstas se centrifugan a 900 rpm 5min. Seguidamente, se aspira el medio y el pellet celular se resuspende en  $500\mu\text{l}$  de solución 1 que contiene 50% de suero bovino fetal y 50% de suero de reemplazo. Justo después, se añaden  $500\mu\text{l}$  de solución 2 que contiene 20% de DMSO y 80% de la solución 1). Después, las células resuspendidas son trasvasadas a un criotubo que es rápidamente congelado a  $-80^{\circ}\text{C}$ .
- Para células H9 crecidas en medio E8 crecidas en un frasco de  $25\text{cm}^2$ , se procede como sigue. En primer lugar, las células son despegadas usando DPBS 1X suplementado con EDTA 0,5mM incubando 5-10 min a  $37^{\circ}\text{C}$ . Luego se centrifugan a 900 rpm durante 5 min y posteriormente se aspira el sobrenadante. Finalmente el pellet celular se resuspende en solución de congelación formada por 10% DMSO en medio E8. La resuspensión celular se trasvasa a un criotubo que es rápidamente congelado a  $-80^{\circ}\text{C}$ .

Para todas las líneas celulares, los criotubos eran congelados utilizando un contenedor con isopropanol (Nalgene) que enlentece el proceso de congelación favoreciendo una mayor viabilidad de las células.

Las células tuvieron que ser descongeladas en multitud de ocasiones. La descongelación tuvo lugar de la siguiente forma:

- El proceso de descongelación es similar en todas las líneas. Descongelación rápida del criovial en un baño a  $37^{\circ}\text{C}$  y posteriormente se trasvasa el contenido del criovial a un tubo de 15ml con 10 ml del medio de cultivo correspondiente. Seguidamente, el tubo se centrifuga (1200 rpm para PA-1, HeLa y células embrionarias de ratón y 900 rpm para H9). Luego, se aspira el sobrenadante y el pellet celular se resuspende en el medio completo correspondiente. Finalmente, las células son trasvasadas al frasco de cultivo correspondiente. En el caso de las PA-1 y HeLa se utiliza un T75 pero para las células H9 se usa un T25 que previamente ha sido tratado con matrigel (1ml 1h temperatura ambiente). Para las células embrionarias de ratón se usan frascos T25 tratados con gelatina al 0,1% durante 10 min a  $37^{\circ}\text{C}$ .

### 3.2. Plásmidos

- pRAM-LRE3: plásmido derivado de pBSKS donde se ha clonado una copia activa de LINE-1 (LRE-3, [11]) seguido del casete de rescate [92] que está flanqueado por dos secuencias FRT (del inglés Flipase Recognition Target). El casete está formado por un gen de resistencia a neomicina clonado en antisentido e interrumpido por el intrón IVS-2 del la  $\gamma$ -globina [12] y el origen de replicación bacteriano ColE1. Después de la segunda FRT, en el extremo 3' hay una señal de poliadenilación SV40 (del inglés Simian Virus 40).
- pCEP4-EGFP: plásmido pCEP4 donde se ha clonado el gen de la proteína verde fluorescente. Expresión dirigida por el promotor CMV (promotor de origen vírico, del citomegalovirus) y aguas abajo de sitio de clonaje hay un señal de poliadenilación SV40 [59].
- CMV-FLP: Plásmido donde se ha clonado el gen de recombinasa “flipasa” cuya expresión está dirigida por el promotor CMV del virus citomegalovirus. Plásmido de expresión también conocido como pOG-Flpe6 descrito en Buchholz et al; en 1998 [216].
- Pk- $\beta$ -arr ( $\beta$ -arr): Contiene el marco abierto de lectura de la  $\beta$ -arrestina humana clonado en el vector pK [217].
- pk- $\alpha$ -Apobec3A (A3A): Contiene el marco abierto de lectura de la APOBEC3A humana clonado en el vector pK [217].
- DN-Drosha: contiene una forma dominante negativa de la proteína DROSHA ( $\Delta C114$ ) [218].
- JJ101/L1.3: contiene una copia retrocompetente de LINE-1 [219] fusionada al casete mBlastI como indicador de la retrotransposición [96] clonado todo ello en un pCEP4 [13].
- JJ101/L1.3D702A: variante del JJ101/L1.3 que contiene una mutación no sinónima (Asp702Ala) en el dominio RT de ORF2 [58].
- TAM102/L1.3: variante del JJ101/L1.3 que carece de 5'UTR [96].
- TAM102/L1.3D702A: variante del TAM102/L1.3 que contiene una mutación no sinónima (Asp702Ala) en el dominio RT de ORF2 [58].
- JJ101.1-300/L1.3: variante del TAM102/L1.3 que sólo posee la porción del nucleótido 1 al 300 de la región 5' UTR clonada en el sitio de restricción NotI.
- JJ101.200-500/L1.3: variante del TAM102/L1.3 que sólo posee la porción del nucleótido 200 al 500 de la región 5' UTR clonada en el sitio de restricción NotI.
- JJ101.400-700/L1.3: variante del TAM102/L1.3 que sólo posee la porción del nucleótido 400 al 700 de la región 5' UTR clonada en el sitio de restricción NotI.

- JJ101.600-900/L1.3: variante del TAM102/L1.3 que sólo posee la porción del nucleótido 600 al 900 de la región 5' UTR clonada en el sitio de restricción NotI.

### Preparaciones plásmidicas para transfección celular.

Los plásmidos fueron transformados en bacterias E.coli DH5 $\alpha$ , aproximadamente 100ng de una preparación plasmídica previa se mezclaron con 18-20 $\mu$ l de bacterias E.coli DH5 $\alpha$  ultracompetentes previamente descongeladas en hielo. La mezcla fue incubada a 4 $^{\circ}$ C durante 20-30 minutos. Posteriormente se dio un choque térmico a 42 $^{\circ}$ C durante 45 segundos. Después, las bacterias se incubaron en hielo nuevamente durante 2 minutos. Luego se añadieron 800 $\mu$ l de medio de cultivo LB (Luria Broth, Sigma-Aldrich) precalentado a 37 $^{\circ}$ C, y justo después, se sembraron aproximadamente 100 $\mu$ l en una placa de LB-Agar (Luria-Broth con Agar, Sigma-Aldrich) con ampicilina 100 $\mu$ g/ml. Las placas fueron incubadas a 37 $^{\circ}$ C durante la noche.

Al día siguiente, se picó una de la colonias resistentes y se cultivó en 50ml de medio de LB líquido con ampicilina 100  $\mu$ g/ml que se incubó en agitación a 37 $^{\circ}$ C durante la noche. Al día siguiente, se procedió con la preparación de ADN plasmídico mediante el kit Quiagen Plasmid Midi kit (Quiagen), el cual se usa para obtener altas concentraciones de ADN plasmídico. Por tanto, se procedió según las instrucciones del fabricante y finalmente siempre se obtuvieron concentraciones que oscilaron entre 300ng/ $\mu$ l y 800ng/ $\mu$ l. Antes de realizar la transfección, se comprobó el superenrollamiento del plásmido en un gel de agarosa al 1%. Solo aquellas preparaciones con al menos el 90% del plásmido superenrollado fueron utilizadas para transfección celular.

Se procedió de la misma forma para el resto de plásmidos usados en el estudio. Todos los plásmidos presentan un gen de resistencia a ampicilina como marcador de selección excepto DN-Drosha cuyo marcador es un gen de resistencia a kanamicina.

### **3.3. Células E.coli X-Gold ultracompetentes para el rescate de la inserción**

Las bacterias ultracompetentes, para la transformación de las ligaciones de plásmidos de ADN genómico producidos durante el proceso de rescate de la inserción de LINE-1, fueron preparadas de la siguiente forma:

- Día 1: Crecer una colonia de E.coli X-Gold (Stratagene) en 4 ml de medio de cultivo LB líquido.
- Día 2: Inocular 250ml de medio SOB (Super Optimal Broth) en un matraz erlenmeyer con 4ml de cultivo bacteriano de la noche anterior. El medio SOB contiene 20 g

de Triptona, 5g de extracto de levadura, 0,5g de NaCl, 10 g de glicina en 1L de agua destilada. Luego se autoclava y después es suplementado con 10 ml de MgCl<sub>2</sub> (1M) (previamente autoclavado) y 625 µl de KCl (4M). Dejar crecer a 18°C durante 48 horas aproximadamente a unas 180-200rpm.

- Día 4: Preparar 100ml de TB (Terrific Broth): 0,30g PIPES, 0,22g CaCl<sub>2</sub>\*2H<sub>2</sub>O, 1,86g KCl, 1,09g MnCl<sub>2</sub>H<sub>2</sub>O. Seguidamente filtrar para esterilizar y guardar a 4°C. Comprobar la densidad óptica del cultivo celular por espectrofotometría y cuando haya alcanzado una densidad óptica igual a 0,94, el matraz se coloca en hielo durante 10 min. Dividir el contenido del matraz en 5 tubos de 50 ml para centrifugar a 2500g durante 10 min a 4°C. Posteriormente, desechar el sobrenadante y resuspender en 16ml de TB por tubo (mantenido en hielo). Seguidamente, dejar en hielo durante 10 min. Después, centrifugar a 2500g a 4°C y desechar sobrenadante. Resuspender en 2ml de TB frío por tubo, seguidamente juntar todo en un tubo (10ml) y añadir 10 ml de TB con DMSO al 14%. Esperar 10 min y después hacer alícuotas de 500µl en tubos de 1,7 ml y congelar rápidamente en hielo seco. Importante, todo el material que entre en contacto con las células debe estar previamente enfriado a -80°C.

Se utilizó esta cepa ya que es capaz de replicar plásmidos de gran tamaño (>10kb) y además es capaz de replicar el ADN metilado humano.

### **3.4. Antibióticos**

Para los distintos experimentos se utilizaron una serie de antibióticos cuya forma de preparación se detalla abajo:

- Neomicina: se prepara una solución madre de 50mg/ml en agua libre de nucleasas y se mantiene a 4°C evitando su exposición a luz (Gibco).
- Kanamicina: se prepara en agua Miliq a 25mg/ml, se conserva a -20°C y se añade a LB líquido o LB agar a una dilución 1:1000 obteniéndose una concentración final de 25 µg/ml (Amresco).
- Ampicilina: se prepara en agua Miliq a 100mg/ml, se conserva a -20°C y se añade a LB líquido o LB agar a una dilución 1:1000 obteniéndose una concentración final de 100 µg/ml (Sigma-Aldrich).
- Blastidina: se prepara en agua Miliq a 10mg/ml, en condiciones de esterilidad y se guarda a -20°C (Invitrogen).

### **3.5. Protocolos agrupados según experimentos**

#### **3.5.1. Western blot**

Las células fueron lisadas en hielo durante 10 min utilizando el siguiente tampón de lisis: 2ml de Ripa Buffer (Sigma-Aldrich) suplementado con 20µl inhibidores de fosfatasa 1 y 2 (Sigma-Aldrich), 100µl inhibidores de proteasas 1x (Roche), 20µl PMSF y 5µl mercaptoetanol (Sigma-Aldrich). Seguidamente, las muestras eran centrifugadas a velocidad máxima durante 10 min a 4°C siendo los sobrenadantes trasvasados a tubos nuevos. Seguidamente, las proteínas fueron cuantificadas mediante absorbancia utilizando el Micro BCA Protein Assay Kit (Thermo Scientific) según las indicaciones del productor.

Posteriormente, 50µg de proteína fueron cargados por cada condición en un gel de poliacrilamida al 10%. Las proteínas fueron separadas durante 1h 30 min a 100v. Seguidamente, las proteínas fueron transferidas a una membrana de nitrocelulosa mediante transferencia húmeda durante 1h a 250mA. Por último, la membrana fue bloqueada con una solución de TBS-tween 0,1% + leche no grasa 5% en agitación. Después de 1 h de bloqueo, la membrana fue incubada durante la noche a 4°C en agitación suave con el anticuerpo primario para ORF1p (donado por S. Martin de la Universidad de Colorado) a una dilución 1:2000. Al día siguiente, después de 3 lavados de 5 min, entonces, la membrana fue incubada con el anticuerpo secundario anti-mouse 1:10000 (Cell Signaling, 7076). Después de 1 h de incubación a temperatura ambiente en agitación suave, la membrana se lavó 3 veces durante 5 min con TBS-tween 0,1%. Posteriormente, la membrana fue revelada usando el kit de quimioluminiscencia de Biorad según las indicaciones del fabricante. Por último, se repitió el proceso incubando ahora con el primario anti-actina producido en ratón 1:5000 (Sigma A1978) a temperatura ambiente durante 45min. Después de lavar, incubación con el anticuerpo secundario anti-mouse 1:10000 (Cell Signaling, 7076) durante 45 min, lavado y revelado como se describe arriba.

#### **3.5.2. Ensayo de retrotransposición en células HeLa**

Las células HeLa se crecieron en un frasco de cultivo de 75cm<sup>2</sup> a una confluencia entre el 80% y el 95%. Se despegaron usando tripsina al 0,25%. Se sembraron 2\*10<sup>4</sup> células/pocillo en placas de 6 pocillos. Las células se transfectaron a las 14-16 horas después de la siembra. La transfección se llevó a cabo utilizando una mezcla compuesta de 97 µl de Optimem (Gibco), 3 µl Eugene 6 (Promega) y 1 µg total de plásmido (0,5µg DN-Drosha o β-arr + 0,5 µg vector de retrotransposición) por pocillo. Para medir la eficiencia de transfección se realizó una tri-transfección (0,25µg DN-Drosha o β-arr + 0,25 µg vector de retrotransposición+ 0,5 µg de pCEP4-EGFP). En primer lugar, se mezcló el Optimem

junto el Fugene 6 y la mezcla fue incubada 5 minutos a temperatura ambiente. Posteriormente, se añadió 1µg de plásmido y la mezcla resultante se incubó durante 20 min a temperatura ambiente. Después, se repartieron los 100µl cuidadosamente sobre cada pocillo. El medio fue reemplazado por medio completo a las 24h después de la transfección. A las 48 horas post-transfección se midió a eficiencia de transfección mediante citometría de flujo. A las 120 horas post-transfección, se comenzó la selección con antibiótico. Para ello, el medio de cultivo fue reemplazado por medio completo suplementado con blasticidina a una concentración de 10µg/ml. A partir de aquí, el medio era reemplazado por medio fresco cada 5 días y la selección se mantuvo durante 12-14 días. Al finalizar el tiempo de selección, las células son fijadas usando la siguiente solución de fijación (54 ml de formaldehído 37% (Sigma-Aldrich), 4 ml de glutaraldehído 50% (Sigma-Aldrich), 100 ml de PBS 10X (Gibco), 842 ml de agua destilada) durante 30 min a temperatura ambiente. Posteriormente, son teñidas con cristal violeta (0,1% peso/volumen en agua durante) 10 min a temperatura ambiente. Finalmente se retiran los restos de cristal violeta lavando con agua. Se deja secar y las colonias se cuentan manualmente. El nº de colonias normalizado respecto a la eficiencia de transfección equivale a la tasa de retrotransposición en cada caso.

### 3.5.3. Generación de líneas clonales resistentes en PA-1

#### – Ensayo de retrotransposición

Las células PA-1 se crecieron en un frasco de cultivo de 75 cm<sup>2</sup> a una confluencia entre el 80% y el 95%. Se despegaron usando tripsina al 0.25%. Se sembraron 8\*10<sup>4</sup> células/placa en 6 placas de 100 mm. Además, se añadió una placa extra para llevar a cabo el análisis de eficiencia de transfección mediante una co-transfección con pCEP4-EGFP. Las células se transfectaron a las 14-16 horas después de la siembra. La transfección se llevó a cabo utilizando una mezcla compuesta de 388 µl de Optimem, 12 µl Fugene 6 y 4 µg de plásmido pRAM-LRE3 por placa. Para la co-transfección, 388 µl de Optimem, 12 µl Fugene, 2µg de plásmido pRAM-LRE3 y 2µg de plásmido pCEP4-EGFP. A partir de aquí, se procedió como en el apartado 3.5.2 repartiendo 400 µl en cada placa. El medio fue reemplazado por medio completo a las 8 horas después de la transfección. A las 48 horas post-transfección se midió a eficiencia de transfección mediante citometría de flujo. A las 72 horas post-transfección, se comenzó la selección con antibiótico. Para ello, el medio de cultivo fue reemplazado por medio completo con neomicina a una concentración de 200µg/ml. A partir de aquí, el medio era reemplazado por medio fresco con neomicina 200µg/ml cada dos días durante 14 días.



– *Aislamiento de colonias resistentes a neomicina*

En primer lugar, el medio de cultivo fue aspirado y los restos de medio fueron eliminados lavando una vez con 8 ml de DPBS 1X. Las colonias resistentes a neomicina fueron levantadas de las placas de 100mm utilizando pequeños trozos de papel de filtro autoclavados que previamente habían sido sumergidos en tripsina 0.25%. De esta forma, usando unas pinzas de metal previamente esterilizadas, se colocó un pequeño trozo de papel de filtro sobre cada una de las colonias dejando actuar a la tripsina durante unos 5 minutos. Después de los 5 min, cada trozo de papel fue recogido con las pinzas y llevado a un pocillo de una placa de 24 pocillos que contenía 500µl de medio de cultivo completo.

– *Expansión de las líneas clonales resistentes a neomicina*

A medida que las colonias fueron creciendo, las células fueron pasadas a frascos de cultivo de superficie creciente hasta alcanzar un suficiente número de células para la extracción de ADN genómico y para establecer un stock de la línea clonal. Las líneas se expandieron con medio completo suplementado neomicina a una concentración 200µg/ml. Proceder al apartado 3.5.6 para detalles sobre la recuperación de la inserción.

**3.5.4. Generación de líneas sensibles en células PA-1**

– *Ensayo de retrotransposición*

Las células se crecieron en un frasco de cultivo de 75cm<sup>2</sup> a una confluencia entre el 80% y el 95%. Se despegaron usando tripsina al 0.25%. Se sembraron  $2 \times 10^5$  células en un pocillo de una placa de 6 pocillos. Por otro lado, se añadió un pocillo extra para llevar a cabo el análisis de eficiencia de transfección mediante una co-transfección con pCEP4-EGFP. Las células se transfectaron a las 14-16 horas después de la siembra. La transfección se llevó a cabo utilizando una mezcla compuesta de 97 µl de Optimem, 3 µl Fugene 6 y 1 µg de plásmido pRAM-LRE3 por pocillo y para la co-transfección 100 µl de Optimem, 3 µl Fugene 6, 500ng de plásmido pRAM-LRE3 y 500ng de plásmido pCEP4-EGFP por pocillo. A partir de aquí se procedió como en el apartado 3.5.2. El medio fue reemplazado por medio completo a las 8 horas después de la transfección. A las 48 horas, se midió la eficiencia de transfección mediante citometría de flujo.

– *Dilución clonal*

120 horas después de la transfección, las células PA-1 fueron levantadas usando tripsina al 0.25% incubando durante 10 min a 37°C. Posteriormente, las células fueron contadas utilizando una cámara de Neubauer y se sembraron en placas de 96 pocillos. Para

asegurar la presencia de una sola célula en un pocillo y por tanto que se generen líneas clonales puras, se sembraron 48 células por placa, es decir, 1 célula cada dos pocillos. Para ello, se preparó una mezcla que contenía 48 células más 9600 $\mu$ l (100 $\mu$ l/pocillo) de medio completo por placa de 96 pocillos. En este caso, se sembraron 40 placas de 96 pocillos con 100  $\mu$ l de medio completo cada uno. Al día siguiente se completó cada pocillo con 100 $\mu$ l de medio completo.

– *Generación de placas espejo de las líneas clonales*

Dos semanas después de la dilución clonal, se comenzó la búsqueda de nuevos eventos de retrotransposición. Para ello, se procedió de la siguiente forma, las colonias resultantes fueron levantadas usando 50 $\mu$ l de tripsina al 0.25% por pocillo, posteriormente la reacción fue bloqueada usando 50  $\mu$ l de medio completo resultando un volumen final de 100 $\mu$ l. El contenido celular del pocillo fue dividido de la siguiente forma: 70  $\mu$ l fueron traspasados a un pocillo de una placa de 48 pocillos mientras que los 30 restantes fueron llevados a un pocillo de una placa de 96 pocillos. Posteriormente, la placa de 48 pocillos fue destinada a la congelación cuando las células alcanzaron el 95% de confluencia. Para ello, se retiró el medio de cultivo y se añadió 200 $\mu$ l de solución de congelación (Suero bovino feta descomplementado + DMSO 10%). Finalmente, la placa fue sellada con parafilm y rápidamente congelada a -80°C. La placa de 96 pocillos fue destinada a la extracción de ADN genómico.

– *Extracción de ADN genómico desde la placa de 96 pocillos.*

El ADN genómico fue extraído directamente sobre la placa de 96 pocillos añadiendo 50 $\mu$ l de Quick Extract DNA Solution (Epicentre). En primer lugar, la placa se incubó a 65°C durante 6 min y posteriormente a 98°C durante 2 min. Por último, la placa fue sellada con parafilm y congelada a -20°C.

– *Identificación de eventos nuevos de retrotransposición mediante PCR de NEO*

La PCR de *NEO* fue llevada a cabo usando KAPPA 2G FAST ReadyMix PCR Kit (Kappa Biosystems) según el protocolo: 12,5 $\mu$ l de mezcla, 0,5 $\mu$ l del oligonucleótido Fw; 0,5 $\mu$ l del oligonucleótido Rev; 2 $\mu$ l de la extracción de ADN genómico y se completa con agua libre de nucleasas hasta un volumen final de 25 $\mu$ l (oligonucleótidos disponibles en Anexo I). Para el control negativo se usaron 20ng de plásmido pRAM-LRE3 y como control positivo se cargaron 150 ng del ADN genómico de una línea clonal con un evento de retrotransposición. El programa usado en el termociclador fue el siguiente:

Temperatura	Tiempo
<b>Paso 1 (Desnaturalización)</b>	
95°C	3min
<b>Paso 2</b>	Repetir 40 ciclos
95°C	15 seg
59°C	15seg
72°C	15seg
<b>Paso 3 (Extensión Final)</b>	
72°C	1min

**Tabla 3:** Programa de temperaturas para PCR de NEO

Por último, los productos de PCR fueron resueltos en un gel de agarosa al 1%.

– *Descongelación y expansión de líneas clonales positivas para la PCR de NEO.*

La placa de 48 pocillos con las líneas positivas fue colocada a 37°C en un baño durante unos minutos. Posteriormente, se rellenaron los pocillos positivos completamente con medio completo suplementado con un 10% extra de suero bovino fetal. Al día siguiente, el medio fue reemplazado por medio completo nuevo suplementándolo con un 10% extra de suero bovino fetal. A partir de aquí, las células fueron pasadas a frascos de superficie creciente hasta tener un número de células para generar un stock de la línea así como para extraer ADN genómico para la recuperación de la inserción (proceder según 3.5.6).

**3.5.5. Generación de líneas clonales con inserciones activas en células H9**

– *Ensayo de Retrotransposición*

En este caso, las células fueron nucleofectadas según el protocolo Amaxa Human Stem Cell Nucleofector Kit2 LONZA:

Se parte de células H9 en un frasco de cultivo de 25cm<sup>2</sup> creciendo sobre matrigel con medio condicionado. 1 hora antes de la transfección, se reemplazó el medio por medio condicionado suplementado con FGF (8ng/ml) e iROCK (10µM). Además se añadió 1-1,5ml de matrigel por cada placa de 100mm destinada a la siembra de las células después de la nucleofección. Las placas se incubaron a temperatura ambiente durante 1 hora antes de su uso. En este caso se sembraron 6 placas.

Pasada esta hora, las células fueron despegadas utilizando 1ml de Tryple Select (Gibco). Éste se dejó actuar 2-4 minutos a 37°C y posteriormente se despegaron las células con ayuda de una rasqueta de forma suave. El contenido de un frasco de 25cm<sup>2</sup>

(aproximadamente al 95% de confluencia) se dividió en dos para hacer dos transfecciones independientes en dos tubos de 15ml. Posteriormente, estos tubos de 15ml fueron completados con DPBS 1X y las células se centrifugaron a 900rpm durante 5min. Después de la centrifugación, el sobrenadante fue aspirado completamente. Posteriormente, los pellets fueron mezclados con 4µg de plásmido pRAM-LRE3 y 100 µl de la solución de nucleofección (LONZA) que debe estar a temperatura ambiente antes de su uso. De aquí, las mezclas fueron trasvasadas a las cubetas propias del kit compatibles con el nucleofector donde se nucleofectaban las células. Rápidamente, se añadió a la cubeta 500µl de medio RPMI (Gibco) previamente calentado a 37°C. Se recogió la nucleofección en este volumen de 500 µl de RPMI en un tubo limpio y se colocó a 37°C durante 15 min. Mientras tanto, se retiró el matrigel de las placas de 100mm y se añadió medio condicionado suplementado con FGF (8ng/ml) e iROCK (10µM). Pasados los 15 min, las células fueron repartidas sobre la placa de 100mm cuidadosamente. Después, de unas 8-12 horas se reemplazó el medio por medio fresco para retirar los restos de solución de nucleofección. Uno de los pellets fue nucleofectado con la mezcla pRAM-LRE3+pCEP4-EGFP (2µg+2µg) para comprobar la eficiencia de transfección mediante citometría de flujo a las 48h.

– *Selección con neomicina*

Después de 72 horas, se comenzó con la selección de neomicina. En este caso, se comenzó con 75µg/ml durante la primera semana de selección mientras que la segunda semana se subió la concentración a 150µg/ml. Para la selección se reemplazó el medio cada dos días utilizando medio condicionado suplementado con FGF (8ng/ml) y neomicina.

– *Aislamiento de colonias resistentes a neomicina*

Una vez la selección finalizó, se comenzó con el levantamiento de colonias resistentes a neomicina para establecer a partir de ellas las líneas clonales. En primer lugar, se retiró el medio de cultivo y se lavó la placa con DPBS 1X. Posteriormente, se añadió medio de cultivo de nuevo, y en la lupa, con ayuda de un minipipeta Pasteur de plástico de punta fina se despegó la colonia y se aspiró, pasando la misma a un pocillo de una placa de 24 pocillos que contenía 500 µl de medio de cultivo condicionado suplementado con iROCK (10µM) y FGF (8ng/ml). Los pocillos fueron previamente tratados con 200 µl de matrigel durante 1h a temperatura ambiente. Para evitar contaminaciones cruzadas, después del levantamiento de cada colonia, la placa fue lavada con DPBS 1X para eliminar cualquier resto celular que hubiera en suspensión. Se repitió el proceso para cada colonia.

– *Expansión de las líneas clonales*

Las colonias fueron pasadas a frascos de cultivo de superficie creciente, hasta alcanzar un número suficiente de células para establecer la línea clonal y extraer ADN genómico para la recuperación de la inserción (seguir pasos descritos en 3.5.6).

3.5.6. Recuperación de la inserción

– *Extracción de ADN genómico*

La recuperación de la inserción comienza con la extracción de ADN genómico. Se utilizaron aproximadamente  $4 \cdot 10^6$ - $5 \cdot 10^6$  millones de células. Las células fueron despegadas en cada caso según los requerimientos de cada línea celular. Finalmente, una vez obtenido el pellet celular, se procedió como sigue:

- Extracción mediante Midi Kit gDNA de Quiagen. Para las líneas resistentes de PA-1. El ADN genómico fue extraído siguiendo en todo momento las indicaciones del fabricante.
- Extracción mediante fenol-cloroformo. Para las líneas clonales sensibles de PA1 y para las líneas resistentes de H9. Este protocolo consta de las siguientes partes:
  - Lisis celular durante 3 horas a 56°C mediante el uso de un tampón de extracción que contiene los siguiente elementos: Tris pH=8,2 10mM, EDTA 10mM, NaCl 200mM, SDS 0,5%, Proteinasa K 200µg/ml (Roche). Se añadieron entre 500-800µl de tampón, en función del tamaño del pellet.
  - Posteriormente, se añade el mismo volumen de Fenol:Cloroformo:Isoamílico (25:24:1) (Life Technologies), se agita vigorosamente unos segundos a mano y posteriormente se centrifuga a máx. velocidad durante 1 min. Se extrae la fase acuosa y se transfiere a u nuevo tubo de 1,7ml. Ahora, se precipita el ADN utilizando el doble de volumen de etanol puro (100%) y se incuba a -20°C hasta el día siguiente.
  - El siguiente día, se centrifugan los tubos a máx. velocidad, se descarta el sobrenadante, y se lava con etanol al 75% y se centrifuga a máx. velocidad. Se descarta el sobrenadante, y se dejan secar los pellets a temperatura ambiente. Cuando los restos de etanol hayan desaparecido por completo, se añaden entre 200-400µl de TE (Ambion) y se coloca a 56°C a 1.400rpm para solubilizar el ADN. Después de una hora, se retiran los tubos del termobloque y se mide la concentración resultante de la extracción utilizando el nanodrop.

– *Digestión ADN genómico*

8 µg de ADN genómico de cada línea fue digerido en paralelo con las enzimas de restricción SSPI (NEB) y HindIII (NEB) según la siguiente reacción:

8µg de ADN genómico, 10µl de tampón de digestión, 1µl de BSA, 5µl de enzima y agua libre de nucleasas hasta llegar a un volumen final de 100µl. La digestión se deja actuar durante la noche a 37°C.

– *Ligación*

Al día siguiente, se añadió 1µl de SSPI o HindIII a la reacción y se dejó actuar a temperatura ambiente durante 2 horas. Después, las enzimas de restricción fueron inactivadas colocando la reacción a 65 grados durante 25 min. Posteriormente, se montó la reacción de ligación añadiendo los siguiente componentes: 8µl de Ligasa T4 (NEB), 40 µl de tampón de ligación y se completó hasta 500 µl con 351µl de agua libre de nucleasas. En el caso de las digestiones con HindIII, la reacción de ligación se llevó a cabo a 16 grados mientras que para las digestiones realizadas con SSPI, la reacción de ligación se llevó a cabo a 25 grados. La reacción de ligación se dejó incubar durante la noche.

– *Concentración, transformación y recuperación*

Al día siguiente, se añadió 1µl de Ligasa T4 y se dejó a temperatura ambiente durante 4 horas. Después, las reacciones se concentraron usando amicones o concentradores (Millipore). Cada una de las reacciones fue trasvasada a un concentrador y se centrifugó a 7500g durante 5 min. Después, se añadieron 450 µl de agua libre de nucleasas y se volvió a centrifugar a 7500g. Finalmente, los concentradores se colocaron dados la vuelta en un tubo limpio y mediante un spin corto el contenido del concentrador fue vertido dentro del tubo.

Una tercera parte del concentrado (aprox. 20µl) fue mezclado con 500µl de bacterias E.Coli X-Gold (producidas en 3.3) ultracompetentes en un tubo de 15ml. Esta mezcla se incubó durante 30 min en hielo. Posteriormente, los tubos se introdujeron en el baño a 42 grados durante 38 segundos. Seguidamente, los tubos se colocaron de nuevo en hielo durante 2min. Finalmente, 1ml de LB líquido precalentado a 37°C se añadió a cada uno de los tubos. Los tubos permanecieron en agitación durante la noche a 550rpm a temperatura ambiente para dejar recuperar a las bacterias.

– *Siembra*

Los tubos con la transformación bacteriana fueron centrifugados a 550g durante 8 min. Posteriormente, se retiró únicamente 1ml de LB y se dejaron 500 ml aprox. en el tubo. Finalmente, el pellet bacteriano fue resuspendido en los 500µl restantes que fueron sembrados en placas de LB-Agar con kanamicina (25µg/ml) de 140mm precalentadas a 37 grados. Las placas fueron incubadas a 37 grados durante 18-24 horas.

– *Picar colonias*

Al día siguiente, las colonias resistentes se crecieron en 2ml de LB líquido con kanamicina (25µg/ml) durante la noche a 200 rpm y 37 grados.

– *Extracción de plásmido*

Al día siguiente, se realizó la extracción de plásmido utilizando el kit Wizard® Plus SV Minipreps DNA Purification System (Promega) según las directrices del fabricante. Todas las preparaciones plasmídicas fueron digeridas con ECORI (NEB) (según la reacción: 0,3 µl enzima, 1µl de tampón de digestión y 8,7 µl de la preparación plasmídica) durante 1hora a 37 grados. Después, los productos de la digestión se resolvieron en un gel de agarosa al 1%.

– *Secuenciación y alineamiento*

Los plásmidos resultantes fueron secuenciados mediante secuenciación Sanger usando dos oligonucleótidos: NEO210AS y RECO3. Con el primero se secuencia hacia el extremo 5' de la inserción ya que anilla sobre el gen *NEO* en antisentido mientras que con el segundo se secuencia hacia el extremo 3' de la misma ya que anilla sobre el origen de replicación bacteriano (ColE1) en sentido. En primer lugar, la secuencia resultante de la secuenciación con NEO210AS se alinea con la secuencia del plásmido pRAM-LRE3 para determinar el punto de truncación de la inserción de LINE-1. En caso de no encontrarse el punto de truncación, se sigue secuenciando hacia el extremo 5' de la inserción con oligonucleótidos solapantes sobre la secuencia de LINE-1 (ver Anexo I) hasta encontrar el punto de truncación y la región genómica situada en 5'. En segundo lugar, la secuencia resultante de la secuenciación con RECO3 se mapea en el genoma de referencia para encontrar la región genómica situada en 3'. En la mayoría de las ocasiones existen problemas técnicos asociados a este proceso. En ocasiones no se puede identificar la región genómica en el extremo 5' porque la inserción ha sido cortada por las enzimas de restricción HINDIII (posición 3662) o SSPI (posición 2335) lo que favorece la unión de otros fragmentos de ADN genómico. En caso de no conocer la región genómica ubicada en 3',

hay que caminar por el plásmido hasta encontrar el extremo 3' diseñando oligonucleótidos solapantes sobre la secuencia del plásmido. Posteriormente la inserción se resuelve por PCR. En ocasiones también ocurre que no se puede identificar la región en el extremo 3' debido a problemas durante la secuenciación ocasionados por la longitud de la cola de Poli A, por lo que hay que identificar el extremo 5' para posteriormente resolver la inserción por PCR. Una vez se han identificado las secuencias genómicas localizadas en el extremo 5' y en el extremo 3', éstas se mapean en el genoma de referencia usando UCSC Genome Browser. En este caso se usó el genoma de referencia de Feb.2009 (GRCH37/hg19).

### 3.5.7. Inmunoprecipitación de cromatina seguida de PCR cuantitativa a tiempo real (ChIP-qPCR).

Para la inmunoprecipitación de cromatina seguida de PCR cuantitativa se procedió como sigue:

#### – *Preparación de cromatina*

Siempre que fue posible, la preparación de cromatina de la línea control y de la línea clonal se hizo en paralelo. Para células PA-1 se utilizó el contenido celular de 5 frascos de cultivo de 175cm<sup>2</sup>, para células PA-1 diferenciadas 10-15 frascos de 175cm<sup>2</sup> y para células H9 6-7 frascos de 75cm<sup>2</sup>. En todos los casos, las células se dejaron crecer hasta alcanzar un 80%-95% de confluencia. Una vez alcanzada dicha confluencia, las células eran despegadas según los requerimientos de cada línea celular. Posteriormente, el contenido de todos los frascos de cada línea era trasvasado a un tubo de 50ml. De aquí se guardaba 1ml de suspensión celular para las PCRs de genotipado. Después, el contenido celular era centrifugado según los requerimientos de cada línea celular. Seguidamente, el pellet de células era resuspendido en 18 ml del medio completo correspondiente. Posteriormente se añadía 2ml de solución de fijación (50mM HEPES pH=8, 1mM EDTA pH=8, 0,5mM EGTA pH=8, 100mM NaCl, 10% Formaldehído (Formaldehído 37%, Sigma-Aldrich)) y se incubaba 10 min a temperatura ambiente en una rueda giratoria. Después, se paraba la reacción de fijación con 2,5 ml de glicina (concentración final 140mM) y de nuevo se incubaba en la rueda giratoria durante 5min. Pasados los 5min, los tubos eran centrifugados a 550g a 4°C durante 5 min. Después, el sobrenadante era desechado y los pellets fueron lavados dos veces con 10 ml de DPBS 1X, previamente enfriado en hielo, centrifugando a 550g a 4°C durante 5min. Después del primer lavado, los pellets resuspendidos en DPBS 1X fueron a trasvasados a tubos de 15 ml.



Después del segundo lavado con DPBS 1X, los sobrenadantes se desecharon y los pellets se resuspendieron en 1 ml de tampón de lisis celular (5mM PIPES pH=8, 85mM KCL y 0,5% NP-40, Inhibidores de Proteasas 1X (Roche). El tampón de lisis celular era suplementado con 5mM de butirato de sodio para inhibir las histonas deacetilasas en las inmunoprecipitaciones para H3K9ac y H4ac. La lisis celular tuvo lugar durante 20 min en hielo. Posteriormente, el contenido era trasvasado a un tubo de 1,7ml para ser centrifugado durante 10 min a 900g a 4°C. Después, el sobrenadante era desechado y el pellet se resuspendió en 1 ml de tampón de lisis nuclear (50mM Tris-Cl pH=8, 10mM EDTA pH=8, 1% SDS, Inhibidores de Proteasas 1x). Del mismo modo, el tampón de lisis celular era suplementado con 5mM de butirato de sodio para inhibir las histonas deacetilasas en las inmunoprecipitaciones para H3K9ac y H4ac. La lisis nuclear tuvo lugar en hielo durante un período mínimo de 5 min.

Por último, se procedió con la sonicación. En primer lugar, se transfirieron los lisados nucleares a tubos de 15 ml (Falcon) compatibles con el sonicador. Seguidamente, dichos lisados se sonicaron en el sonicador Bioruptor Plus de Diagenode usando 15 ciclos de 30 seg on/50 seg off para células PA-1, 2X de 15 ciclos de 30 seg on/50 seg off para células H9 y para células PA-1 diferenciadas. Posteriormente, la eficiencia de sonicación era determinada en un gel de agarosa al 1% cargando 8 µl del sonificado junto con 2µl de sarcosyl 5% y tampón de carga 6x (NEB). La mayoría de las sonicaciones dieron lugar a una población mayoritaria de fragmentos de ADN que se encontraban en torno a 800-1200pb. Los geles de las sonicaciones más representativas de algunos experimentos de ChIP-qPCR se muestran en el Anexo II al final de esta sección. Una vez comprobado el tamaño de los fragmentos de ADN generados, se procedió a eliminar todo el material insoluble generado durante la sonicación mediante dos pasos de centrifugación a velocidad máxima (15.000g 20min a 10°C). Después de la segunda centrifugación, se cuantificaba la cantidad de cromatina obtenida en el nanodrop. En general, las concentraciones de cromatina oscilaron entre 2µg/µl y 3 µg/µl. Después, la cromatina se diluyó 10 veces usando tampón RIPA (10mM Tris-Cl pH=7,5, 1mM EDTA pH=8, 0,5mM EGTA, 1% Tritón X-100, 0,1% Sodio Deoxycolato, 140mM NaCl, Inhibidores de Proteasas 1X). Por último, la cromatina fue congelada en alícuotas de 300µg para las células PA-1, 200µg para células H9 y 150µg para células PA-1 diferenciadas.

#### – Inmunoprecipitación

La segunda parte del protocolo comienza con la preparación de las bolas magnéticas, que idealmente, se deben preparar el día anterior a su uso.

Para ello, se utilizan 50µl de bolitas magnéticas (Life Technologies) por cada una de las inmunoprecipitaciones que se vayan a realizar. Posteriormente, la bolitas magnéticas se lavan dos veces con tampón RIPA estándar previamente enfriado en hielo (10mM Tris-Cl pH=7,5; 1mM EDTA pH=8; 0,5mM EGTA; 1% Tritón X-100; 0,1%SDS; 0,1% Na Deoxycolato; 140mM NaCl, Inhibidores de Proteasas 1X). Para realizar los lavados, el tubo con las bolitas magnéticas se posiciona en el imán (Thermo Scientific) durante 1 min, y una vez que las bolitas han migrado hacia la pared del tubo, entonces se pipetea el contenido líquido del tubo. Después del segundo lavado, las bolitas magnéticas se resuspenden en 1ml de tampón RIPA estándar junto 10 µl de BSA (Bovine Serum Albumine, NEB) y 4 µl de ADN de esperma de salmón (Invitrogen) para bloquear. La reacción de bloqueo se hace en movimiento a 4°C usando una rueda giratoria durante 30 min como mínimo. Pasado dicho tiempo, los tubos se colocan en el imán y se procede con 3 lavados de tampón RIPA estándar. Por último, las bolitas se resuspenden en el mismo volumen original usando tampón RIPA estándar y se guardan a 4°C hasta el día siguiente. Mantener el tampón RIPA en hielo así como enfriar en hielo el imán antes de ser usado.

Para llevar a cabo la inmunoprecipitación propiamente dicha, el primer paso, es la descongelación a 4°C de las alícuotas de cromatina que estaban congeladas a -20°C. Una vez descongeladas, se añadían a cada alícuota el anticuerpo correspondiente en la cantidad indicada: IgG 10µg (Dako Z0259); anti-H3 5µg (Abcam 1791); anti-H3K4me3 2,5µg (Abcam 8580); anti-H3K9me3 5µg (Abcam 8898); anti-H3K27me3 1µl (Active Motif 39535); anti-H4ac 2,5µg (Millipore 06-866); anti-H3K9ac 2,5µg (Abcam 4441). Antes de añadir el anticuerpo, se extrajo 1/100 del control negativo (IgG) para el input. Posteriormente, las inmunoprecipitaciones se incubaban durante la noche (15-17 horas) a 4°C en la rueda giratoria. Al día siguiente, en cada una de las inmunoprecipitaciones se añadían 50µl de bolitas magnéticas (preparadas el día anterior) y se incubaban durante 5 horas a 4°C en la rueda giratoria. Pasado dicho tiempo, se procedía con los lavados de las inmunoprecipitaciones usando tres tipos diferentes de tampón RIPA y TE, los cuales eran enfriados en hielo y mantenidos en hielo durante todo el proceso. Las inmunoprecipitaciones eran lavadas durante 5 min a 4°C en la rueda giratoria y para eliminar los restos líquidos las inmunoprecipitaciones se posicionaban en el imán, previamente enfriado en hielo antes de cada lavado. En primer lugar, se lavaban las inmunoprecipitaciones dos veces (2x500µl-700µL) con tampón RIPA estándar (10mM Tris-Cl pH=7,5; 1mM EDTA pH=8; 0,5mM EGTA; 1% Tritón X-100; 0,1%SDS; 0,1% Na Deoxycolato; 140mM NaCl, inhibidores de proteasas 1x). Después, dos lavados (2x500µl-700µl) con tampón RIPA con alto contenido en sal (10mM Tris-Cl pH=7,5; 1mM EDTA pH=8; 0,5mM EGTA; 1% Tritón X-100; 0,1%SDS; 0,1% Na Deoxycolato; 550mM NaCl; inhibidores de proteasas 1x). Seguidamente, un único lavado (1x 500-700µl) con tampón RIPA con

cloruro de litio (10mM Tris-Cl pH=7,5; 1mM EDTA pH=8; 0,5mM EGTA; 1% Tritón X-100; 0,1%SDS; 0,1% Na Deoxicolato; 250mM LiCl; inhibidores de proteasas 1x). Por último dos lavados con TE pH=8. Después del último lavado, las inmunoprecipitaciones se resuspendieron en 150 µl de tampón de elución (20mM Tris-HCl pH=7,5; 5mM EDTA pH=8; 50mM NaCl; 1%SDS; 50 µg/ml Proteínasa K (Roche); 100 µg/ml de ARNasa A (Sigma). Se revirtió la unión bola/anticuerpo en un termobloque incubando durante la noche a 68°C a 1100rpm. El input, que había sido congelado el día anterior, fue resuspendido en 150µl del tampón de elución y colocado también en termobloque junto con el resto de resuspensiones.

A día siguiente, las inmunoprecipitaciones fueron posicionadas en el imán para extraer la parte líquida de la suspensión y trasvasarla a un tubo limpio de 1,7ml. Seguidamente, las bolas magnéticas se volvieron a incubar con tampón de elución durante 5 min a 68° a 1100rpm (20mM Tris-HCl pH=7,5; 5mM EDTA pH=8; 50mM NaCl; 1%SDS; 50 µg/ml Proteínasa K; 100 µg/ml de ARNasa A). Posteriormente, se colocaron las muestras en el imán y se volvió a pipetear la parte líquida que se mezcla con los anteriores 150µl. En este paso, los tubos con las bolas magnéticas eran desechados. Posteriormente, se añadieron a cada tubo 200µl de tampón de elución sin SDS, Proteínasa K y ARNasa A.

Seguidamente, se procedió con la extracción del ADN inmunoprecipitado. Para ello, se añadió a cada muestra 500µl (1 volumen) de fenol/cloroformo/isoamílico (25:24:1) (Life Technologies). Los tubos eran agitados vigorosamente y se centrifugaron a máxima velocidad durante 1min a temperatura ambiente. La fase superior, aprox. 400µl, fue pipeteada en un tubo de 1,7ml limpio. Se añadieron a la fase superior, 500µl de Cloroformo/Isoamílico 24:1 (Acros), agitando los tubos vigorosamente. Posteriormente, los tubos se centrifugaron a máxima velocidad a temperatura ambiente. De nuevo, la fase superior, aprox. 300 µl, era trasvasada a un tubo limpio de 1,7ml. Ahora, se añadió a ésta, 600µl de etanol 100% (2 volúmenes), 50µl de Acetato Sódico 3M pH=5.2 y 1,5 µl de Glicógeno (20mg/ml) (Affymetrix). Para la precipitación del ADN se incubaron los tubos a -80°C durante 30 min. Después, se centrifugaron a máxima velocidad a 4°C y se desechó el sobrenadante. Seguidamente, se lavaron con etanol 75% y de nuevo se centrifugaron a máxima velocidad a 4°C. Por último, se desechó el sobrenadante y se colocaron los tubos, dados la vuelta, a secar a temperatura ambiente. Pasadas unas 2-3 horas, se resuspendieron los inmunoprecipitados en 150µl de TE (Ambion).

#### – *PCR cuantitativa a tiempo real*

Las reacciones de PCR se montaron en placas de 96 pocillos (Applied Biosystems) de la siguiente forma: 10µl de 2x Go Taq qPCR Master Mix (Promega), 0,4µl del oligonucleótido

Fw, 0,4µl del oligonucleótido Rev y 2µl del inmunoprecipitado completando hasta 20µl con agua libre de nucleasas por reacción. Cada reacción se montó en duplicado. La PCR cuantitativa a tiempo real fue llevada a cabo en la máquina STEPONEPLUS SYSTEM de Applied Biosystems siguiendo el programa por defecto para cuantificación relativa.

Todas las parejas de oligonucleótidos utilizadas en los experimentos de ChIP-qPCR (disponibles en el Anexo I de esta sección) se diseñaron para tener una temperatura de anillamiento de 60°C. Posteriormente, para calcular los enriquecimientos de cada modificación en cada una de las posiciones analizadas, se normalizó respecto H3 según la fórmula  $\Delta Ct: 2^{Ct_{H3} - Ct_{Anticuerpo}}$ . Cuando fue posible se llevó a cabo un contraste estadístico para determinar la significancia estadística de los cambios epigenéticos observados. El test estadístico utilizado fue la T de student asumiendo varianzas iguales entre la muestras pues los experimentos están realizados en las mismas condiciones. Se consideraron significativos aquellos p-valores menores de 0.05 en una distribución de dos colas. El contraste estadístico fue realizado en Excel 2007 o en Graph Pad Prism 6.

La eficiencia de todas las parejas de oligonucleótidos usadas en el estudio fue determinada y se usaron aquellas parejas cuya eficiencia estaba comprendida entre 1,6 y 2. Para determinar la eficiencia de las parejas de oligonucleótidos previamente se hizo una curva estándar para cada una de ellas partiendo de 2µl de Input y haciendo diluciones sucesivas 1:2. Posteriormente, esta se calculó según la fórmula  $10^{1/\text{pendiente de la curva estándar}}$ .

### 3.5.8. Recombinación mediada por flipasa

#### – *Co-transfección pCEP4-EGFP/CMV-FLP*

En primer lugar, se sembró un placa completa de 6 pocillos poniendo  $2 \times 10^5$  células por pocillo. A las 14-16 horas, las células fueron co-transfectadas usando pCEP4-EGFP y el plásmido CMV-FLP. Para la co-transfección se usó la siguiente mezcla: de 97 µl de Optimem, 3 µl Fugene 6 y 1 µg de ADN plasmídico por pocillo. La proporción de plásmido fue la siguiente: 0,25 µg de pCEP4-EGFP y 0,75 µg de CMV-FLP. A partir de aquí, se procedió como 3.5.2. A las 6-8 horas, el medio de la transfección fue reemplazado por medio completo.

#### – *Separación de células transfectadas*

A las 48 horas, las células se despegan poniendo 500µl/pocillo de tripsina al 0.25%. La reacción se deja actuar durante 10min a 37°C. La reacción se bloquea con 500ul de medio completo. Posteriormente, se pasan por un filtro de 0.40µm. Luego, se centrifugan a 1200rpm durante 5 min. Se desecha el sobrenadante y el pellet celular es resuspendido en

tampón de separación (DPBS 1X, Suero Bovino Fetal Descomplementado 3%-5%, EDTA 5mM). Seguidamente, se procede con la separación de células por citometría de flujo. Las células GFP+ recogidas se siembran directamente en una superficie de cultivo dependiente del nº de células obtenido. Al día siguiente el medio se reemplaza por medio completo suplementado con un 10% extra de suero bovino fetal.

– *Dilución clonal*

Proceder como en 3.5.4. En este caso se sembraron 5 placas de 96.

– *Expansión de clones*

Dos semanas después, 20 subclones eran seleccionados. Éstos fueron expandidos hasta tener dos pocillos de una placa de 6 pocillos, destinando uno de ellos para extracción de genómico y otro para congelación de un vial.

– *Extracción de ADN genómico*

Extracción mediante fenol:cloroformo, según apartado 3.5.6 de esta sección.

– *PCR de NEO*

Según apartado 3.5.4 de esta sección, con la salvedad de que se usaron 150ng de ADN genómico por reacción.

– *Ensayo funcional de resistencia a Neomicina.*

La línea clonal con inserción activa junto con la línea subclonal FLP+, fueron cultivadas con medio completo suplementado con neomicina 200µg/ml durante 14 días.

– *PCRs alternativas para demostrar la escisión del casete RECO*

En función del tamaño del amplicón se usó bien Expand Long Template PCR system (Roche) para amplicones grandes (aprox. 3000 pares de bases) y para amplicones más pequeños (≤1000bp) KAPPA 2G FAST ReadyMix PCR Kit (Kappa Biosystems) (ver tabla para más detalles). Los oligonucleótidos están disponibles en el anexo 1 de esta sección.

Expand Long Template	Cantidad	Kappa 2G FAST Ready Mix	Cantidad
Buffer 1	5µl	KAPPA 2G FAST ReadyMix	12,5µl
Oligonucleótido Fw 20uM	1µl	Oligonucleótido Fw 10µM	0,5µl
Oligonucleótido Rev 20uM	1µl	Oligonucleótido Rev 20µM	0,5µl
dNTPs 10 Mm (cada uno)	4µl	ADN genómico	150ng
Expand Long Template	0,7µl	Agua libre de nucleasas	Hasta 25µl
ADN genómico	300ng		
Agua libre de nucleasas	Hasta 50µl		

**Tabla 4:** Preparación de reacciones de PCR para las polimerasas Expand Long Template) y Kappa 2G Fast

Las condiciones de temperatura y tiempo para estas PCRs son las siguientes:

Expand Long Template PCR		KAPPA 2G FAST ReadyMix PCR	
Temperatura	Tiempo	Temperatura	Tiempo
<b>Paso 1 (Desnaturalización)</b>		<b>Paso 1 (Desnaturalización)</b>	
96°C	1min	95°C	3min
<b>Paso 2</b>		<b>Paso 2</b>	
94°C	40 ciclos	95°C	40 ciclos
	15seg	95°C	15seg
Tª hibridación <sup>1</sup>	30seg	Tªhibridación <sup>1</sup>	15seg
68°C	4min	72°C	15seg <sup>2</sup>
<b>Paso 3 (Extensión)</b>		<b>Paso 3 (extensión)</b>	
68°C	10min	72°C	1min

**Tabla 5:** Programas de temperaturas para cada tipo de polimerasa.<sup>1</sup>En función de la pareja de oligonucleótidos utilizada.<sup>2</sup>Desde 1 seg hasta 15 seg en función del tamaño del amplicón (1seg por cada 100 pares de bases).

### 3.5.9. Cuantificación por PCR cuantitativa a tiempo real de las inserciones endógenas

Se utilizó el método de la curva relativa estándar.

- Extracción de ADN genómico de H9-pRAM4/H9-pRAM4 subclon 13/H9-pRAM4 subclon 9 (Extracción fenol:cloroformo, según apartado 3.5.6 de esta sección).
- Los ADN genómicos fueron diluidos a una concentración de 25ng/µl.
- Preparación de la curva estándar. Se hizo una curva estándar para cada pareja de oligonucleótidos siguiendo una dilución 1:5 utilizando como primer punto de la curva 2µl de ADN genómico (25µg/µl) lo que equivale a 50ng de ADN genómico.
- Las reacciones para la curva estándar fueron preparadas de la siguiente forma: 10µl de 2x Go Taq qPCR Master Mix (Promega), 0,4µl del oligonucleótido FW, 0,4µl del oligonucleótido RV y 2µl de cada una de las diluciones de ADN genómico en un volumen final de 20µl por reacción (completar con agua libre de nucleasas).

- Las reacciones para las muestras problema fueron preparadas de la siguiente forma: 10µl de 2x Go Taq qPCR Master Mix, 0,4µl del oligonucleótido FW, 0,4µl del oligonucleótido RV y 1µl de ADN genómico (25µg/µl) en un volumen final de 20µl por reacción (completar con agua libre de nucleasas). Se hicieron tres réplicas por cada reacción. Los oligonucleótidos usados están disponibles en Anexo I de esta sección.
- Las reacciones fueron llevadas a cabo en la máquina STEPONEPLUS SYSTEM de Applied Biosystems utilizando el programa de temperaturas por defecto para cuantificación relativa con curva estándar. Las temperaturas de hibridación fueron las siguientes: 63°C para 1-927; 55°C para X-810; 60°C para 4-144 y 64°C para subunidad 25 de la ribonucleasa P.
- El nº de copias relativo obtenido para cada una de las inserciones según la curva estándar de cada pareja de oligonucleótidos se normalizó respecto al nº de copias relativo del gen referencia (Subunidad 25 de la ribonucleasa P) que toma el valor 1.

#### 3.5.10. Generación de subclones en H9-pRAM4

- Dilución clonal. Descrito en apartado 3.5.4., aunque se realizó con alguna salvedad. Tratamiento de las placas de 96 pocillos con matrigel durante 1 hora a temperatura ambiente. Cada pocillo fue tratado con 50µl de matrigel. En este caso se sembraron 5 placas de 96 pocillos. Para sembrar las células, el medio E8 fue suplementado con iROCK (10µM).
- Generación de placas espejo. Descrito en 3.5.4 pero con alguna salvedad. Las placas de 24 y 96 pocillos fueron tratadas con matrigel 200 µl y 50µl respectivamente durante 1h a temperatura ambiente. Extracción de ADN genómico sobre las placas de 96 pocillos según el apartado 3.5.4. En este caso, las placas de 24 pocillos no se congelaron sino que se mantuvieron en cultivo hasta la identificación de los subclones positivos por PCR.
- Identificación por PCR de genotipado de los subclones positivos usando KAPPA 2G FAST ReadyMix PCR Kit usando las indicaciones sobre la preparación de reacciones en apartado 3.5.4. Los productos de PCR fueron resueltos en un gel de agarosa al 1%. Los oligonucleótidos de genotipado están disponibles en el Anexo I de esta sección. Las temperaturas de hibridación fueron 55°C para la inserción X-810 y 60°C para la inserción 4-144.
- Expansión de los subclones positivos para generar una línea clonal.
- Cuantificación de las inserciones por PCR cuantitativa a tiempo real en los subclones de H9-pRAM4.9/H9-pRAM4.13 según apartado 3.5.9.
-

### 3.5.11. PCR cuantitativa a tiempo real del gen NEO

- Extracción de ARN. Para la extracción de ARN, las células PA-1 fueron despegadas, contadas y se sembraron  $3 \times 10^5$  células/pocillo de una placa de 6 pocillos siguiendo las recomendaciones para las células PA-1. Cuando las células, estaban a un 90-95% de confluencia, el medio de cultivo fue retirado y se añadió 1 ml de Trizol (Life Technologies) y se prosiguió según recomienda el fabricante.
- Tratamiento del ARN con DNAasa I (Invitrogen). En primer lugar las muestras eran tratadas durante 15 min a temperatura ambiente con DNAasa I según la reacción: 2µg de RNA, 4µl Buffer DNAasal 10X, 4µl enzima y completar con agua libre de nucleasas hasta 40µl. Seguidamente, la reacción era bloqueada a 65°C 10min añadiendo 4µl de EDTA 25mM.
- Reacción de reverso transcripción con un oligonucleótido específico para el gen *NEO* y el gen *GAPDH* usando la reverso transcriptasa del virus de la leucemia murina de Moloney (M-MLV RT, del inglés Moloney Murine Leukemia Virus Reverso Transcriptase) (Promega).

Las reacciones de reverso transcripción se prepararon como sigue:

5µl Buffer 5X MMLV; 10µl ARN tratado con DNAasa I; 1µl oligonucleótido de reverso transcripción (20µM); 1,25µl dNTPs (10mM) (Invitrogen); 0,62µl RNAsin (inhibidor de ARNasas, Promega); 1 µl reverso transcriptasa M-MLV y completar con agua hasta 25µl. Se preparó en cada caso la reacción correspondiente sin reverso transcriptasa como control negativo. Posteriormente, las reacciones se llevaron a cabo a 42°C durante 1h.

- PCR cuantitativa a tiempo real del cDNA.

Las reacciones de PCR cuantitativa se prepararon como sigue: 10µl de Go Taq 2x qPCR Master Mix (Promega), 0,4µl del oligonucleótido FW, 0,4µl del oligonucleótido RV y 2µl de cDNA (dilución 1:5) y completar hasta de 20µl por reacción con agua libre de nucleasas. Los oligonucleótidos están disponibles en el Anexo I de esta sección.

- Las reacciones de PCR cuantitativa a tiempo real se llevaron a cabo en la máquina STEPONEPLUS SYSTEM de Applied Biosystems utilizando el programa por defecto para cuantificación relativa. En este caso la temperatura de hibridación utilizada fue 60°C.
- Análisis de resultados. La expresión del gen NEO en cada línea clonal fue determinada mediante la fórmula  $\Delta Ct: 2^{Ct_{GAPDH} - Ct_{NEO}}$ . La expresión relativa del gen NEO en cada línea clonal fue normalizada respecto a la expresión relativa del gen NEO en la línea PA-1-pRAM4 tomando ésta valor 1.



### 3.5.12. Cuantificación del nivel de expresión de genes receptores de inserciones de LINE-1

El protocolo descrito también fue utilizado para la cuantificación del ARNm de LINE-1 mostrado en figura 19.

- Extracción de ARN. Para la extracción de ARN, las células PA-1 y H9 fueron despegadas, contadas y sembradas a una densidad de  $3 \times 10^5$  células/pocillo en una placa de 6 pocillos siguiendo las recomendaciones para cada tipo celular. Cuando las células, estaban a un 90-95% de confluencia, el medio de cultivo fue retirado y se añadió 1 ml de Trizol (Life Technologies) y se prosiguió según recomienda el fabricante.
- Tratamiento del ARN con DNAasa I (Invitrogen). En primer lugar las muestras eran tratadas durante 15 min a temperatura ambiente con DNAasa I según la reacción:  $2 \mu\text{l}$  de RNA,  $2 \mu\text{l}$  Buffer DNAasal 10X,  $2 \mu\text{l}$  enzima y completar con agua libre de nucleasas hasta  $20 \mu\text{l}$ . Seguidamente, la reacción era bloqueada a  $65^\circ\text{C}$  10min añadiendo  $2 \mu\text{l}$  de EDTA 25mM.
- Reverso transcripción de cDNA usando High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Las reacciones de reverso transcripción se prepararon de la siguiente forma:  
 $2 \mu\text{l}$  RT Buffer 10X;  $0,8 \mu\text{l}$  dNTPs 20mM; Random Primers  $2 \mu\text{l}$ ;  $0,5 \mu\text{l}$  RNAsin (inhibidor de ARNasas);  $1 \mu\text{l}$  reverso transcriptasa;  $10 \mu\text{l}$  ARN procedente del paso anterior y completar con agua hasta  $20 \mu\text{l}$ . Se preparó la reacción correspondiente sin reverso transcriptasa como control negativo. El programa de temperaturas es el siguiente:  $25^\circ\text{C}$  10 min,  $37^\circ\text{C}$  2h,  $85^\circ\text{C}$  5min.
- PCR cuantitativa a tiempo real del cDNA  
Las reacciones de PCR cuantitativa se prepararon como sigue:  $10 \mu\text{l}$  de 2x Go Taq qPCR Master Mix (Promega),  $0,4 \mu\text{l}$  del oligonucleótido Fw,  $0,4 \mu\text{l}$  del oligonucleótido Rev y  $1 \mu\text{l}$  de cDNA (dilución 1:3) completando hasta  $20 \mu\text{l}$  con agua libre de nucleasas por cada reacción.
- Las reacciones de PCR se llevaron a cabo en la máquina STEPONEPLUS SYSTEM de Applied Biosystems siguiendo el programa por defecto para cuantificación relativa. Los oligonucleótidos fueron diseñados para tener una temperatura de hibridación de  $60^\circ\text{C}$ .
- Análisis de los resultados. La expresión del gen de interés en cada línea clonal fue determinada mediante la fórmula  $\Delta\text{Ct}: 2^{\text{Ct}_{\text{GAPDH}} - \text{Ct}_{\text{Gen Interés}}}$ .

### 3.5.13. Ensayo de resistencia a neomicina en células PA-1 y células H9.

#### – Ensayos de resistencia a neomicina en PA-1

Las células fueron despegadas, contadas y sembradas a una densidad de  $4 \times 10^4$  células por pocillo en una placa de 6 pocillos. Al día siguiente comenzó la selección usando medio completo suplementado con  $200 \mu\text{g/ml}$  de Neomicina. El medio fue reemplazado cada dos días con medio fresco suplementado con neomicina durante 14 días. Pasados los 14 días, el medio fue retirado y se añadieron 2 ml de solución de fijación (descrita en 3.5.2). Después de 30 min a temperatura ambiente, la solución fijadora fue retirada y se añadieron 2 ml de solución de cristal violeta (descrito en 3.5.2). Después de 10 min, la solución de cristal violeta fue retirada y los pocillos fueron lavados con agua para eliminar los restos de cristal violeta.

#### – Ensayo de resistencia a neomicina en células H9

Las células fueron despegadas, contadas y sembradas a una densidad de  $8 \times 10^4$  células en una placa de 100 mm tratada previamente con 1ml de matrigel durante 1 hora a temperatura ambiente. Al día siguiente, el medio de cultivo fue reemplazado por medio suplementado con  $100 \mu\text{g/ml}$  de neomicina. El medio era reemplazado cada día con medio fresco suplementado con neomicina. Después de 7 días, el medio es retirado y las células son fijadas con solución de fijación (descrita en 3.5.2) y teñidas con cristal violeta (descrita en 3.5.2).

### 3.5.14. Relación entre los niveles de expresión del gen NEO y los enriquecimientos de H3K9me3 y H3K4me3

Los valores relativos de expresión del gen *NEO* en cada una de las líneas fueron representados junto con los valores relativos de enriquecimiento de ambas modificaciones de histona. Los valores relativos de expresión del gen *NEO* fueron obtenidos en el apartado 3.5.11. de esta sección. El valor relativo de cada modificación de histona fue calculado normalizando el valor medio de enriquecimiento sobre la inserción de LINE-1 respecto al valor medio del control positivo para dicha modificación en las réplicas biológicas correspondientes. Por tanto, los valores relativos de enriquecimiento de H3K9me3 y H3K4me3, fueron calculados según la fórmula:

A)

$$1. \text{Enriquecimiento medio de H3K9me3 o H3K4me3 sobre la inserción} / 2. \text{Enriquecimiento medio del control positivo para H3K9me3 o H3K4me3}$$

La desviación estándar resultante se calculó, según la fórmula:

B)

$$\left( \text{RAIZ} \left( \left( \frac{DS1}{1} \right) + \left( \frac{DS2}{2} \right) \right) \right) * \text{enriquecimiento relativo calculado en A).$$

DS significa desviación estándar. Por otro lado, cabe destacar, que se utilizaron todas las líneas con inserciones artificiales independientemente del número de réplicas biológicas realizadas en los experimentos de CHIP-qPCR.

### 3.5.15. Relación entre el reclutamiento de H3K4me3 y H3K9me3 y la expresión del gen receptor

Los valores relativos de enriquecimiento de H3K9me3 y H3K4me3 fueron obtenidos según el apartado 3.5.14. Posteriormente, éstos se representaron junto a la expresión relativa de los genes receptores de dichas inserciones calculada según el apartado 3.5.12 de esta sección. Finalmente, la expresión de los genes fue normalizada respecto a la expresión del gen receptor de la inserción de LINE-1 de la línea PA-1-pRAM4 que tomó el valor 1.

### 3.5.16. Regresión lineal entre la longitud de inserción y el enriquecimiento de H3K9me3.

Para ello, se utilizaron todas las inserciones artificiales de LINE-1. Por tanto, se utilizaron 10 inserciones artificiales. En primer lugar, se calculó el número de nucleótidos exactos pertenecientes a la propia secuencia de LINE-1 para cada una de las inserciones alineando la secuencia de la inserción con la secuencia del vector pRAM-LRE3. Posteriormente, se representó la longitud de cada inserción frente al enriquecimiento de H3K9me3 en cada una de las inserciones usando el software Graph Pad Prism 6. Se utilizaron los enriquecimientos de H3K9me3 mostrados en la figura 31 calculados según el apartado 3.5.7 de esta sección. Seguidamente, se llevó a cabo una correlación de Pearson usando del software Graph Pad Prism 6. Para determinar la significancia estadística de la correlación se aplicó un intervalo de confianza del 95% para una distribución de dos colas.

### 3.5.17. PCRs de genotipado

Como paso previo a cada inmunoprecipitación, se llevó a cabo una PCR de genotipado para demostrar la presencia de la inserción de LINE-1 correspondiente. En primer lugar, se realizó una extracción de genómico mediante fenol:cloroformo según el apartado 3.5.6 de

esta sección. Posteriormente se llevaron a cabo las PCR de genotipado utilizando KAPPA 2G FAST ReadyMix PCR Kit según tabla 4 y 5 del apartado 3.5.8 de esta sección. Temperaturas de hibridación y tiempo de extensión dependientes de la pareja de oligonucleótidos utilizada.

### 3.5.18. Anexo I. Oligonucleótidos



**Oligonucleótidos PCR NEO**

<b>Nombre Oligonucleótido</b>	<b>Secuencia</b>
Neo Fw	TGGATACTTTCTCGGCAGGAGC
Neo Rev	CATTGAACAAGATGGATTGCACGC

**Oligonucleótidos Rescate o Recovery**

<b>Nombre Oligonucleótido</b>	<b>Secuencia</b>
RECO3	CTAAAGTATATATGAGTAACC
NEO210AS	GACCGCTTCTCTGCTTTACG
JORF2JAS	GGTGCGCTGCACCCACTAATG
JORF2IAS	CACATTTTCTTAATCCAGTC
JORF2HAS	CTAACTGGTGTGAGATGATATC
JORF2GAS	GAATTGATTTTTGTATAAGGTG
JORF2FAS	CATTTTCACGATATTGATTCTTCC
JORF2EAS	GTCCATCAATACCTAATTTATTG
JORF2DAS	GAATTCGGCTGTGAATCCATCTGG
JORF2CAS	TTAATTGTGATGTTAGGGTGTC
ORF2BAS	GATTTGGGGTGGAGAGTTCTG
PRAM4/Chrm2a	GCTGTTGAGTGTATGAGTACAG
PRAM4/Chrm2b	ACTGTGCATACACTGGATG
PRAM8/Chrm1a	AACAAAGAGCCATAAAGTGATC
PRAM8/Chrm1b	CTGTGCTTCTTTGGTGTTTTGC
PRAM8/Chrm1c	GTGTCTACCTACAAATTAACCAT
PRAM8/Chrm1d	GCAGTGACTTAAGCCCAGTC
PRAM8/Chrm1e	CACATTGGTCTGAGTACATTAC
PRAM8/Chrm1f	GTTCTGCTGCTATCTGACTGCTC
PRAM8/Chrm1g	GACGTTTATAGATGCTGCAC
PRAM8/Chrm1h	CCTCCCGGCATTTTATTCTG
PRAM8/Chrm1i	CAGGTTGTGGTTAGGCTTAGTG
PRAM8/Chrm1PCR	CATCTGGGGCTCTCTCTGTTCTC
PRAM11/Chrm2a	CACATTTCTTCTCATGGCAC
PRAM26/Chrm53'	AAGTATCCAGCTTGGGTGAGTC
PRAM27/Chrm14a	CAGCAATCGAAGACGAACATC
PRAM27/Chrm14b	AATGTCTAAGTTCTCTGAAATGC
PRAM27/Chrm14c	CTCCAGTCTCCCATCAAGTAG
PRAM27/Chrm14d	AGGCAGAGCACATCCCACAAG
PRAM27/Chrm14e	AGACAAGAGTTGGGGAAGAG
pPRAM27/ChrmX	AGTGGTACAATCACAGCTCAC
PRAM27/ChrmXb	CTCATGCTGGTTAACTGCGTC
PRAM27/ChrmXc	TTCTCTCTCTCCTTCTCTTC
PRAM27/ChrmXPCR	CAGAATTGAGAGCTGGAAGGAATAC
PR29H/Chrm1	GGAGTAGCCCTGTATGAGACTTG

PRAM29/Chrm1b	CATTAATAAGCCTGTTACATG
PRAM29/Chrm1c	CTCACAGAAGCAGATGTAAGTTTTC
PRAM29/ChrmXb	ATTTCTCTGGTAATCCTTGACCCTC
PRAMN29/ChrmXc	GTTTTGGCGCAACACCATTACAC
PRAM29ChrmXPCR	CATAGAGCCTTGCTGACCACCTC
PRAM31/Chrm3	AACAGAAGGGCTGGCTTTG
PRAM33/Chrm16	GTGTTGGGATTACAGGAGTGAG
PRRAM33/Chrm16b	TTGTGGTTTGGCCTGATTCTC
PRAM33/Chrm16PCR	CATTGCTGGCTCTGTTGCTTTC
PRAM41/Chrm4PCR	GATATTAGCTACATGCTTCCAGCC
PRAM44/Chrm5a	CTTGATAAGGTAGATTGAGAAG
PRAM44/Chrm5b	CATTTGCACAAGAACACTCTC
PRAM44/Chrm12	TTAGCTGAGTTTGGTGGTGC
PRAM44/Chrm12b	CTGGAGCACAGCATTGTCTTAG
PRAM45/Chrm9	TGCTCTGATATGAACTTAGG
PRAM45/Chrm9b	GGTATTAGGAGGTGATTTTGC
PRAM45/Chrm9c	GATCTCCCTAGCCATGCTGAAC
PRAM45/Chrm9d	GCATATCAGAGGTCTTCATG
PRAM45/Chrm9e	CATGGGAGCCTACTTCTTG
PRAM45/Chrm9f	GAGGAAAGTGAGCTATGCCTAC
PRAM45/Chrm4	AGTTGGGCAAGAGACATCTTG
PRAM45/Chrm4PCR	TAGACGGCTCTTTCCTGTACTG
PRAM221/Chrm14	AAGCAACCCTGACAGTCTG
PRAM221/Chrm14b	GGAGGCACACTTTGTCTGAC
PRAM221/Chrm14c	TAGCAGGAGGTGACCTTCCAG
PRAM221/Chrm14d	GTATAGTCCTCCCTCTTCCAAATG
PRAM221/Chrm14e	CATGGCAGACCTGAGAGCGCATAc
PRAM221/Chrm14f	ACCACTCAAAATACTGCATCAAAAC
PRAM221/Chrm14g	CTAGAACCCCATCCCTGCATG
PRAM221/Chrm20	AATCAACCCAATTTTGTGTCCCTG
PRAM221/Chrm20c	GGATGGCTGGACAACCGAAGTGC
PRAM221/Chrm20PCR2	AGCCGTCTGACCAAAAGAGAAGGAG
PRAM272PCR	AACTACAAAGGGGCATATCATAGC
PRAM286/Chrm1	TCTGTGGTTCTTGCTGCCCTTTG
PRAM286/Chrm1b	CTACTGGTGGCAGGCATAGTG
PRAM286/Chrm1c	GCTCCCCATATGTGAATATCAG
PRAM286/Chrm1d	GTTTGCTCACAGAGACAGAAGC
PRAM286/Chrm1e	CTTTGCTGAATATCTGACTTC
PRAM286/Chrm1f	TGAAGTTTGGAGCACCAAGTGGCAG
PRAM286Chrm1g	CTGTTCTCATTATGAGGGATATG
PRAM286Chrm1PCR	AACCCACAACATGAGTAAATTCATG



**Oligonucleótidos Recombinación Flipasa**

<b>Nombre Oligonucleótido</b>	<b>Secuencia</b>
2H sentido	GATATCATCTCACACCAGTTAG
SV40REV FLP2	TAAAGCAATAGCATCACAAATTTAC
PRAM15 PCR INSERT2	CTGGAAGTTCTGTCTCAAGGTCCTG
PRAM272PCR	AACTACAAAGGGGCATATCATAGC

**Oligonucleótidos utilizados en los experimentos de CHIP-qPCR**NEO

<b>Nombre Oligonucleótido</b>	<b>Secuencia</b>
Neo Fw1	AGCTGCGCAAGGAACGCCCGTC
Neo Rev1	GCGCCCGTTCTTTTGTCAAG

RECO

<b>Nombre Oligonucleótido</b>	<b>Secuencia</b>
FLPRECO5	CACCAGCATGGCACATGTATACGAAG
SV40REVFLP	TTCTAGTTGTGGTTTGTCCAAACT

Línea PA-1-pRAM4

<b>Nombre Oligonucleótido</b>	<b>Secuencia</b>
PRAM4-9Fw	GTGTCTAAGTTTCTGCTGCGAC
PRAM4-9Rev	GAGGTGCTGAGGCATGAGGTG
PRAM4-545Fw	GCACATACGGATCATGCTTCATTAC
PRAM4-545Rev	CTGCCTTTAAACTGACTCTACATG
PRAM4-1171Fw	CGGCATAGATTGCATCCTTCTG
PRAM4-1171Rev	GCCATGGCCTTTTGGCTTGGTG
PRAM4-1976Fw	CATATCACAAAGCATTCTTTACCAC
PRAM4-1976Rev	AAGATTGTACCATATTTAAGAGTGC
PRAM4-2366Fw	CGCATGTATCTCTCTAGTTATAGC
PRAM4-2366Rev	TGTCATCATTATCATCAACACTATC
PRAM4-2687Fw	ACGCGGAGAGAAACGTGATTCTC
PRAM4-2687Rev	AGCCTCAGTCTAAAACATTTCCAGC
PRAM4-3359Fw	GCGGATTAATCACAGAACTCTTAC
PRAM4-3359Rev	GAACTGATTTTTATAACTGAAGATG
PRAM4-4193Fw	GAAGCCAACAGCTCTTGTAGAAACTC
PRAM4-4193Rev	CCATGAGTTTGGAGAAAATTCAG
PRAM4+0Fw	CAAGCTCCTATGTGGCATAATATC

PRAM4+0Rev	AAATTGCAAACCTACCTCGATAAG
PRAM4+679Fw	TTCAATCCACGGCAGTTACATAAAC
PRAM4+679Rev	ACTTTTAAACCTCTGTAGCCACAG
PRAM4+1187Fw	GGAGTTGAGGTCAGGAAGTCTGGC
PRAM4+1187Rev	CTGTAACCTCCTGGACTGGAAG
PRAM4+1897Fw	TTAAAGCTACACACACAAAACCTC
PRAM4+1897Rev	G TTCAGTCCATGTCAAATGTGCTG
PRAM4+2620Fw	TTGCTGGTGTATTCTATTGGTTTG
PRAM4+2620Rev	TCAGACTGGGACCAATTTTTGTTAC
PRAM4+3322Fw	CATAAAGGATAGATGTTGATGAGAG
PRAM4+3322Rev	GCCATTCACAGCAGATTCAGC

Línea PA-1-PRAM26

Nombre Oligonucleótido	Secuencia
PRAM26-375Fw	GAAACTCCAATGCAATCTTAAAGAC
PRAM26-375Rev	CTTCTTGGACAGTGGGAGTGGGGAC
PRAM26-1069Fw	TCACCCCTACCACTTACTACCAAAC
PRAM26-1069Rev	GATATTTGGGATCTTCGATAACTTC
PRAM26-4152Fw	ACTTCATCACCCGCTGGAGTTTCAG
PRAM26-4152Rev	CAGAGATACTGAGGGGAAGAACAG
PRAM26-5021Fw	TGTGATTGCTGTCTTGGGTCTATAG
PRAM26-5021Rev	CTCTTGCCAGCAACGTGGAGCTC
PRAM26+706Fw	TCCAGCCCTGTCAATCAGCAGGAC
PRAM26+706Rev	TATAAAACCTCTTAGGCAAAAGTTG
PRAM26+3413Fw	TAGGCCTCATTTATTTCTTTGCTTC
PRAM26+3413Rev	TATTACCTGGATCACTATTAC
PRAM26+4644Fw	AGGCCACAGAAGTAATAAGATTTGC
PRAM26+4644Rev	ACCTCACAAGTTCAATGCACAATG
PRAM26+5399Fw	GAGAGTTTCACTAAGACTGTACCTC
PRAM26+5399Rev	ATTAAGTATGATTTCAAGGCAAGC

Línea PA-1-PRAM8

Nombre Oligonucleótido	Secuencia
PRAM8-28Fw	ACATCTGGGGCTCTCTGTCTC
PRAM8-28Rev	CTCCCCACCCTACCTCCAAAAC
PRAM8-1628Fw	CCACTAGCCATTGGTTGCCTGTC
PRAM8-1628Rev	GCAGCATTCTTCAGGGCAGAACCG
PRAM8-2119Fw	GAAGAAGTCTGAACCACTATTAC
PRAM8-2119Rev	GAAATGCCTCATGGGTAGTTCAC
PRAM8-2999Fw	GTAAGAAACATTTGCTGACTTGC
PRAM8-2999Rev	CTAAAACACACATTGCAAAACAC

PRAM8-3642Fw	TACGAACCCATGTAAGGGGGCAG
PRAM8-3642Rev	CCTTCTGCTCCTATATCCACCATG
PRAM8-4307Fw	CGATAAAGCCTGATGCCCATCTAC
PRAM8-4307Rev	CACTCTGTCTGTCTTCCATAAATTC
PRAM8-4715Fw	GGAAGAGGTGAGTCCGGTCTC
PRAM8-4715Rev	ATGATCCTCTGCTCCCGGCGGTG
PRAM8-5802Fw	CTCACTTATTTGTGTACCGTCTTG
PRAM8-5802Rev	TGTACCTAATGCACGTTTAACACAG
PRAM8+57Fw	ACTAAGCCTAACCACAACCTGACAG
PRAM8+57Rev	CTCTCTACTTCTTTCAGCTTGTTTC
PRAM8+630Fw	ATCACCACACAATTGGGCTTTTTAC
PRAM8+630Rev	GGGTGGTAAATACAGGGGCTGAG
PRAM8+1139Fw	CACCCAGTGCGTATGCAGGAAAC
PRAM8+1139Rev	TCCTCTGGAGTACTTCTTATGCATC
PRAM8+1761Fw	GAACATCTGGCTGAACAACCTACTG
PRAM8+1761Rev	GTTACATCCCTTGTCTAAGGCAG
PRAM8+3368Fw	GAGGATTACAAAACAGGGAGGTTGAC
PRAM8+3368Rev	GCTTCCCCAGACAGCTGCAG
PRAM8+4087Fw	CAAAGAACAACCTCCAGAATTC
PRAM8+4087Rev	TCTCGACAATCAGCATCTGGTTG
PRAM8+5238Fw	GTATAGCCATGAGTGCTGGGATCAC
PRAM8+5238Rev	CCTCTGCTTTCCCATATTGGTTC
PRAM8+5831Fw	CCAACCAAGGCAAAGTCTTTTAG
PRAM8+5831Rev	GATGAATCTTTTTAAGTGAGGCAGC

Línea PA-1-PRAM15

<b>Nombre Oligonucleótido</b>	<b>Secuencia</b>
PRAM15-707Fw	GAGTGTGGGCGCCTACCATGTGTG
PRAM15-707Rev	CTGGCTGTGTGCCCGTGTATGTATG
PRAM15-1368Fw	GTCCTAGGACCCTCTGCAATCTG
PRAM15-1368Rev	CAGCTTTGTTTGGTAGCACAGAAG
PRAM15-1967Fw	CTCCCCTCAGAAATAGTCTTCAG
PRAM15-1967Rev	GATCCACTACACGCTCCCTG
PRAM15-2591Fw	CAGAGATGAGACTGGCAGGTAAC
PRAM15-2591Rev	TCTCCACCCAGGCCTGGTC
PRAM15-3121Fw	GTGAAGGGCTCAGGGTAGTATGTC
PRAM15-3121Rev	GGGTGAGTGTAGTCTGGCCTGTC
PRAM15-3913Fw	GTGAGGGTCTTGCACTGTGACTC
PRAM15-3913Rev	ATAGGTGAGTAGAGCCTCTGATAG
PRAM15-4600Fw	AAGCCTCCACATTCTCAAACAGGAC
PRAM15-4600Rev	CAGCTGCCGTCCCGAGGCCTTCAG
PRAM15+20Fw	CTCTAGGGATTTTGCCAC
PRAM15+20Fw	GCATATGTAAGTCTGAGGGAATC
PRAM15+593Fw	CAGATTCTCACTGTGCCAC

PRAM15+593Rev	CACAGAGGTGAGTGGGGTAGTTG
PRAM15+1147Fw	CAAGAAATGAGACTGGGGTCCCAG
PRAM15+1147Rev	CGCATATGGCTTTTTAATGGCTCTC
PRAM15+1849Fw	GATCCAGCTCTGTACTCTGCTTG
PRAM15+1849Rev	GTCGGTAGATGCACGTTAATAGTAC
PRAM15+2525Fw	AGTGAGTGGTGGGTAGGCGCTC
PRAM15+2525Rev	GAAACAGCTGAGGAGGAGGCATCTC
PRAM15+3148Fw	AGAGAACCACGATAAAGGAGTCCAC
PRAM15+3148Rev	GGCTCCCTTGCATCCGCTGTC
PRAM15+3817Fw	TGTTGTCGTGTGATCTCCAGCAG
PRAM15+3817Rev	GGGACGTCAATGTGCTTCTCAAC
PRAM15+4473Fw	CCCCATGGTGTGTAAGGACAG
PRAM15+4473Rev	GAGCTGCAGCCAAGGCCCATG

Línea PA-1-PRAM6

<b>Nombre Oligonucleótido</b>	<b>Secuencia</b>
PRAM6-273Fw	CCAACAGTTCAGGCTCTTGACACG
PRAM6-273Rev	TGCTTGGCTGACTATGAGGGTGATG
PRAM6-741Fw	TACAAAGCTAGATCCGAGTGTGTTT
PRAM6-741Rev	GAAACATGGGAACTTTTGGGGTGG
PRAM6-1377Fw	GGTATCCCTCAGCATATGAACCAC
PRAM6-1377Rev	GGCACTCTTAACTCTCTTAACTGC
PRAM6-3063Fw	TGCCTTGGTCACATGGTGGAGATC
PRAM6-3063Rev	ACTGCTCACCTCAAACACTTTCCTG
PRAM6-3745Fw	CCTAACCAACTAATTATGTGCCCTC
PRAM6-3745Rev	ACAGGACAAACATTAGCTCAGTTGC
PRAM6-4417Fw	CTTATCCGTTTCTGCCTAGACCAAG
PRAM6-4417Rev	AGGTGTTTATGCCTTTATGACCCAC
PRAM6-5116Fw	CCCCTTGGACGCTTGCTGTGTGC
PRAM6-5116Rev	GGGTTGGGGAGGGAACAGAAGGC
PRAM6-5666Fw	GAACTGCTCAGAAATGCCTCACAG
PRAM6-5666Rev	AAGCACTGCCACACACACCCATGC
PRAM6+46Fw	AAGACTTGGCTTGCTGCTCATAGAG
PRAM6+46Rev	GTTTAGATCAGGCCAATGCCAAAAG
PRAM6+550Fw	CGTTATTGGGTCCACTTGAGAAACC
PRAM6+550Rev	GAGCCTTGGCTAGAGAGGATTGGAG
PRAM6+1203Fw	TCCATCTGGGAGAGTCCCCTC
PRAM6+1203Rev	CAAATGTCCCCAGAGCTGGTAG
PRAM6+1792Fw	CAGACTGGCTGTCAAGATCACACAG
PRAM6+1792Rev	AACCAGTAGTTAGCCTTACAGACTC
PRAM6+2571Fw	CGTGAATCCTAGAACAAGCATGAAC
PRAM6+2571Rev	CATTCTACTACCCAGTCTCCAGTC
PRAM6+3198Fw	GCCAAACATGAGCCACGGCTCTG
PRAM6+3198Rev	GCGTTGCACTTCAATGTGGTATGC

PRAM6+3978Fw	AAGCTGAAGAATGGGGCCTAAGATC
PRAM6+3978Rev	GATGGCAGGCGTTTAGCCACCAAC
PRAM6+4613Fw	AAGCTACCAGGGCTCCAGTATCCCG
PRAM6+4613Rev	GGGTCCTTCTCTCTACTCTTCTCCT

Linea PA-1-PRAM221

Nombre Oligonucleótido	Secuencia
PRAM221-19FW	CTAGCCCCACTATGAAGCCGCTCG
PRAM221-19Rev	TTTTGAGGTTGGCTGAATTTTATGC
PRAM221-1514Fw	CGATTAGAGGGTGGGAAGGGTTTTG
PRAM221-1514Rev	ACCTTTTGGTCATAATGACTTGGAG
PRAM221-2130Fw	CCACGGTGTACTCAAACGTGAATG
PRAM221-2130Rev	ATGCTAATGGTAAACAGTCAGAGC
PRAM221-2670Fw	ACACGTAATGCACTGAAGAAGATC
PRAM221-2670Rev	ACAGAGCCAAATCTACATCTGTCTG
PRAM221-3575Rev	CGACTGAAGCGCGGTGAAAGTAC
PRAM221-3575Fw	TTTCTATAGCAGAGGACAGTAGTG
PRAM221-4134Fw	AAGAAATCCAGACTTCAGCACTGG
PRAM221-4134Rev	AGTGCATTGTGGGGCTGAACCAAGC
PRAM221+0Fw	ATCGCTGCATGGCTCTTTCAATTC
PRAM221+0Rev	TCTGCCACTGACAGCAGCAGATGTC
PRAM221+548Fw	GCAATGGGCTATGGTTGGGATGAAG
PRAM221+548Rev	GGTTGCAGAGGCATTTTGCTTTCAG
PRAM221+989Fw	CGTTGGGTGCCTTTTTGAAGGTTG
PRAM221+989Rev	ACATGATGCAAGTTCAGATGTCTGC
PRAM221+1668Fw	TACCAGGGACAGCAAAATTGGGTTG
PRAM221+1668Rev	GCCCAGCCTGTTTCCCTCATTCTTG
PRAM221++2405Fw	CCAGGAGGTGGACACCAGG
PRAM221+2405Rev	AAAACCCAAAACTCTCTTAGGAAG
PRAM221+2989Fw	AGGCACGTGCACCCAAGAAGACCTG
PRAM221+2989Rev	CTGTCCAGTCCTCAGGGGTTCCATG
PRAM221+3646Fw	AGGACAGGGGAAGGGGAGGAAAG
PRAM221+3646Rev	CCATCATTTAGAATCACCGATGCTG
PRAM221+4374Fw	GAGAGTCATGGGGAGACCTAGGG
PRAM221+4374Rev	GAGTAAGCGATTTCCCTCTGCAGC

LíneaPA-PRAM286

Nombre Oligonucleótido	Secuencia
PRAM286-3Fw	CCCACAACATGAGTAAATTCATGTC
PRAM286-3Rev	ACCCAGATGAATACAGTGCTAACTC
PRAM286-1100Fw	GTGAGAGAAACATGTTGGGATGAG
PRAM286-1100Rev	CTTCAGTTAGCTTCTCCCTATTAC

PRAM286-2072Fw	GGGACTACAAGCCATCCTCAAATAG
PRAM286-2072Rev	CTGCAAAGAAGCCACCTTAAGTAGC
PRAM286-2809Fw	AAGCCCTCCTCCACTCTTTGGC
PRAM286-2809Rev	AGCGCAAGATTGGGGAGCTGGAAAAG
PRAM286-3332Fw	CAGTCCAGAGAAGATGAACCTCCAG
PRAM286-3332Rev	CCCTGGCTCCCTTCTTTGATTAC
PRAM286-3901Fw	GTTACAACTCTTGACGGCTTCAATC
PRAM286-3901Rev	GGGGAAATGGCACTAGGGGAAAAG
PRAM286-4600Fw	AAGTTGAATACAGGTTCTCTGAGTG
PRAM286-4600Rev	AACCACATACCTGTAGCTTGTTTTC
PRAM286+6Fw	TCCCTCCATGAATGAGAACAGGGGC
PRAM286+6Rev	TGGCTCTGATTTTACCACTAGCTTG
PRAM286+607Fw	GACAAGCTAGAGACATAGAACAGGC
PRAM286+607Rev	TCCCATTTGGTGAGATGAAGCTCCTG
PRAM286+1210Fw	ACCACCAACAACCCGACAGTCTTC
PRAM286+1210Rev	AGGGAGTGCCGAGAGGGGAAGC
PRAM286+1865Fw	ACCATTGGGGGATTCTGGGGGC
PRAM286+1865Rev	GGGACCCACGACAAAGAAAGTTG
PRAM286+2922Fw	TATGACTGAGGGGGTGAAGGAG
PRAM286+2922Rev	TTGGGAGTAGTGTGAAAGAGAAGCG
PRAM286+3507Fw	GCCATGAAGCAAGGTCAAGAACAAG
PRAM286+3507Rev	ACCACCATTCGAAGCTCAGATGTC
PRAM286+4244Fw	AGAGAAGCAATGTCCCAAGACAAAC
PRAM286+4244Rev	ACAGTAACACAGTATCTGGGCATGC

Línea PA-1-PRAM272

Nombre Oligonucleótido	Secuencia
PRAM272+11Fw	GCTTTCTGATGTGTGGATTCATCTC
PRAM272+11Rev	GTT TCC ACT TTG AGA TGC TGT TAA C
PRAM272+584Fw	CTGACCATTCCCATTCACTTCTATTC
PRAM272+584Rev	GCTTCTTTCTAGTGTATCATGTG

Línea H9-pRAM3

Nombre Oligonucleótido	Secuencia
H9-PPRAM3-579Fw	AAGGAGAAAGCATTCAAACAGTAGAG
H9-PPRAM3-579Rev	CCCAAGTTTTCAGGTTTATCAGTTG
H9-PRAM3-1637Fw	GTGAAAACCAAGGGCACACCAGGC
H9-PRAM3-1637Rev	CATTTCCCAGATCTTCTTCTCTTC
H9-PRAM3-2380Fw	CTCAACAATTCCTAGGCAGGGTC
H9-PRAM3-2380Rev	TTTCCATGTAGGGGCAGTGGCAG
H9-PRAM3-3053Fw	AGTCACTCTCATCCACATCCCAAG
H9-PRAM3-3053Rev	AACATGAAGCTCCATCCACATTCCG

H9-PRAM3+1260Fw	GGTTTCATAGGGTGGATGCATCGGC
H9-PRAM3+1260Rev	GAGATTTAGGTGGGGACACAAAGTC
H9-PRAM3+1832Fw	CAAAGGTGTTACAGGGACCATGC
H9-PRAM3+1832Rev	TCTGCCACAGGGACCTCCTTC
H9-PRAM3+2511Rev	TGCCTCTAAAGTCACCTGTCTTGC
H9-PRAM3+2511Rev	CTGTCTGCAGGCTTTCTACTCTG
H9-PRAM3+3285Rev	AGGCTCACCTGTCAAACAAACAGG
H9-PRAM3+3285Fw	GAACCATCCTCCTTCTCTTGCTG

Línea H9-pPRAM4

Nombre Oligonucleótido	Secuencia
H9-PRAM4FW-118	CTCCCTCCTCCCCAGCTTAAGTTC
H9-PRAM4Rev-118	GTCTGAAGTATGGCCCTTTACAGC
H9-PRAM4FW-875	GAGATCGACCAAAGCCCTCAGTAAC
H9-PRAM4Rev-875	CTGATGGGGATAAATTGACTGTGCG
H9-PRAM4FW-1414	TTAACTCCCCACATCATCCAGGAC
H9-PRAM4Rev-1414	TTGAAAGAGAGAGACAGTGTGTTGG
H9-PRAM4FW-2617	TTAGCCCTCTCTTAGCTCTTCCACC
H9-PRAM4Rev-2617	CTATGTAGACAGTGGCAGCAGCAGC
H9-PRAM4Fw+1551	CCCCAAGCTAGACCACTAGGAG
H9-PRAM4Rev+1551	ATGCAGGAGTTAAGAACTAGGATG
H9-PRAM4Fw+2129	TGCAAGGTACTTCAGAGGTCCTCAG
H9-PRAM4Rev+2129	GGCAGTCTCACATTTTGAAAACCTTC
H9-PRAM4Fw+2583	AGTTCCGTTGCATTATCCACTGTTG
H9-PRAM4Rev+2583	GGTTACACGTGGCAATTTCCCTCTC
H9-PRAM4Fw+3120	ATACCTCATGCCCCATTTTATATG
H9-PRAM4Rev+3120	AAAAAAAAGACAACCCCATCCCCAC

Inserción1-927

Nombre Oligonucleótido	Secuencia
1-927-269Fw	CAGGCAAAGTGCTATTTGAAGAACTC
1-927-269Rev	TCACTGAGCAGTTAGGCGATTGC
1-927-739Fw	CTCTGTAAAACCTCCTGATCCTCAG
1-927-739Rev	CAGAACAGCTTGATGAGTCTTTAC
1-927-1157Fw	AAACACACACACACAGACCCCATG
1-927-1157Rev	AATAGCCTAAAATGGTGGGACAAAG
1-927+0Fw	CAGACAAAACCTTATTTCTAGGTTT
1-927+0Rev	AGATGAAGGAAGAATACAGAAGGAG
1-927+710Fw	CCACCGTGCCCGGCTGAGC
1-927+710Rev	GCCACTGCGCTCCAGCCTG
1-927+1014Fw	CAAAGTCTGGGATTACTGGCTGAC
1-927+1014Rev	GTCACCACACTCAGCCAAAGTTTT

1-927+1478Fw	GAAAGTCATAGCAGAACAAGTTAGC
1-927+1478Rev	TGTTAGCATATTGTTGTAAGCTCTGG

Inserción X-810

Nombre Oligonucleótido	Secuencia
X-810-187Fw	AGACAAAACCCCTAGTTTCTTGCTCC
X-810-187Rev	AAGGCTCAACTGAAATGTTTTAAG
X-810-534Fw	TATCCTAGCAGATGCTGTTAACCTC
X-810-534Rev	CCACCAAGTTTTATTAAGCCAGTC
X-810-889Fw	AGAGGTAAGTGAACCAAGATTGAAG
X-810-889Rev	CTCAGAAGCTCCAGTCACTAGGTG
X-810-1305Fw	GGGGTATGAACAGGGTTTTAATGC
X-810-1305Rev	ACCCATTTCAATTCAGAAATCCAG
X-810-1600Fw	TAATTGTGTTGAAAGTGCTTCTGGC
X-810-1600Rev	CTGAAGCCGAGAAAAGTTGAATGGC
X-810+890Fw	GGAGATGTTTCCATTTGACTTCTGG
X-810+890Rev	TCAGCCACTTTCATAGGCATAACTC
X-810+1618Fw	CATATGCTGAAAGGGTGAAGGACAC
X-810+1618Rev	AAGGCTTGACCAGGGAGGAATTTAC

Inserción 4-144

Nombre Oligonucleótido	Secuencia
4-144+339Fw	TGGTTGCCTAGGTATGTTCACTTTG
4-144+339Rev	GTCTTCTGGATGTTAGATTCAGGTC
4-144+598Fw	GTGAGGGACTGCTCCAGAACAAAAG
4-144+598Rev	TCTTAAATCACTTTCTTCCCCCATG
4-144+868Fw	CATATATGTACAGAGATAAACAGAGC
4-144+868Rev	TCCTTCCATCTCAGCCTCCCAAG
4-144+1201Fw	AAGGCAGGCAAGAAGCCAGGC
4-144+1201Rev	TAATGAATCCAAGGATTTTCCCATG
4-144+1557Fw	TGTTATATGTCCTTCCAGAGATATG
4-144+1557Rev	CAATGTGCATATGTAATCTTAGGAG
4-144+1925Fw	GAAGGTGTCCCCACACTTGCTTAC
4-144+1925Rev	GTGGAAAAAGGGGACAATAACTCAG

**Oligonucleótidos para genotipado/qPCR de inserciones endógenas**

Nombre Oligonucleótido	Secuencia
RiboPSubunit25Fw	CCACTCCTGGGGCCCTCTC
RiboPSubunit25Rev	CGGCAGGATCTTCCGCCAGC



1-927Fw	TCTCCAGAATGGCCACTGTAGAATG
1-927Rev	CCTATTCGGCCATCTTGGCTCCTC
X-810Fw	AGAACATGTAGAATGATAAGAGTTG
X-810Rev	CCTATTCGGCCATCTTGGCTCCTC
4-144Fw	CATACTACCGAGTTATACCCAACCTG
4-144Rev	CTGGGAGCTGTAGACCGGAG

### **Oligonucleótidos para RT-qPCR**

<b>Nombre Oligonucleótido</b>	<b>Secuencia</b>
OSBPL6EX3Fw1	CCAAAGGGACAGTAGGCAGA
OSBPL6EX4Rev1	GCTCGGTGCTAGAGGAAGC
RAD51BEX1Fw1	AAAGGGTGGGGAAACTTGAA
RAD51BEX2Rev1	CCCCTCGTTTTAGTTTCTTGC
FAM65BEX1Fw1	TGTCCTCTCCAAGATGCAGTT
FAM65BEX2Rev1	GGGAGTCTGGTCGGTAGTCC
EYA2Fw3	GATGCACGGCACAACAGGGTTC
EYA2Rev3	GGAGGGTTGTAGGATGAGCCG
COL4A4Fw1	GGAGCAGCAGGGGACAAAGG
COL4A4Rev1	GGGTGCCAGGTATGCCATC
SGCDFw9	CGAGTTTTAGGAGCGGAGGGCA
SGCDRev9	CCTTTTGGGGCCTCCATCACT
RPAP2Fw	CTTTTGCAGCAATT TTTGTTATC
RPAP2Rev	CTTTGTTCTTCTTTAGCAGTTG
NANOGFw	TGCAGTTCAGCCAAATTCTC
NANOGRRev	CCTAGTGGTCTGCTGTATTACATTAAGG
POUF51Fw	AGTGAGAGGCAACCTGGAGA
POUF51Rev	ACACTCGGACCACATCCTTC
GAPDHFw	TGCCACCAACTGCTTAGC
GAPDHRev	GGCATGGACTGTGGTCATGAG
ACTINAFw	ACCGAGCGCGCTACAG
ACTINARRev	CTTAATGTCACGCACGATTTCC
NEOFw	AGCTGCGCAAGGAACGCCCGTC
NEORRev	GCGCCCGGTTCTTTTTGTCAAG
TF L1 Fw	GAAATTAGTCTGAACAGGTGAGAGG
TF L1 Rev	TCCTCTGGTCCGGAAGGT

### **Oligonucleótidos para reversotranscripción**

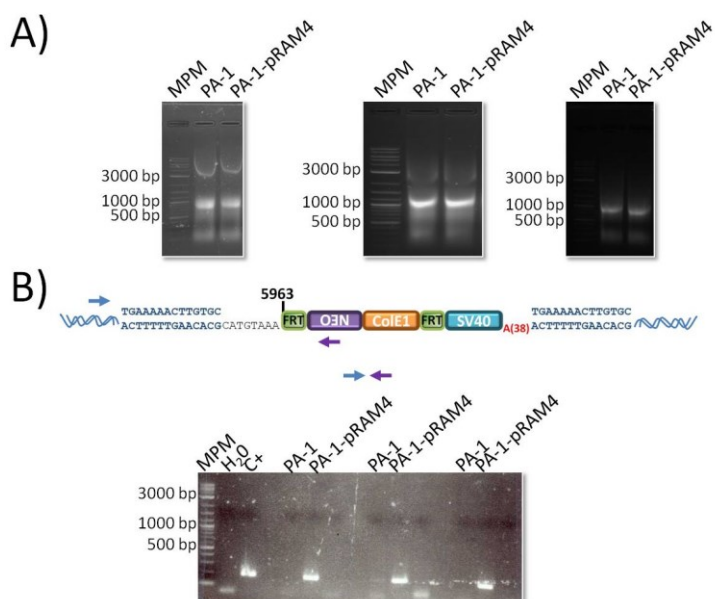
<b>Nombre Oligonucleótido</b>	<b>Secuencia</b>
Neo RT	TTCCGCTTCAGTGACAACGTCG
Gapdh RT	GGCATGGACTGTGGTCATGAG

**Oligonucleótidos para genotipado de inserciones de LINE-1**

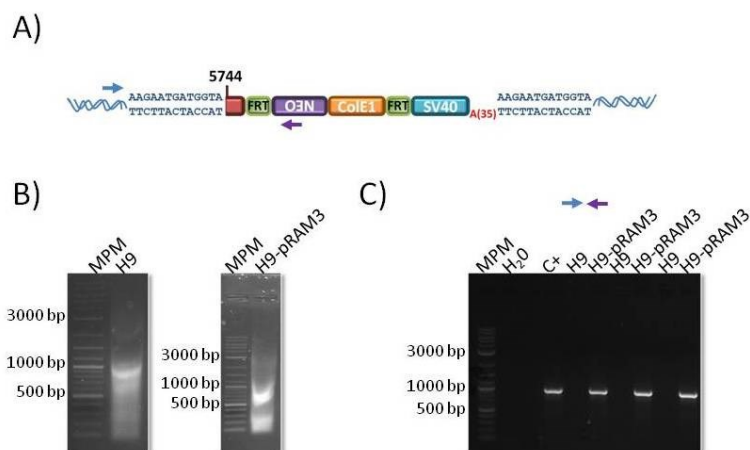
<b>Nombre Oligonucleótido</b>	<b>Secuencia</b>
PRAM4 PCR INSERT2	AGCAGGAAAGAAGACAAAAGGTGAC
PRAM6 PCR INSERT 2	ACAGCCTGGACCATACTGAATGATG
PRAM8 Chrm1 PCR	CATCTGGGGCTCTCTCTGTTCTC
PRAM15 PCR INSERT2	CTGGAAGTTCTGTCTCAAGGTCCTG
PRAM221 Chrm20 PCR	TTCTAGCCCCACTATGAAGCCGTC
PRAM 286 Chrm1 PCR	AACCCACAACATGAGTAAATTCATG
PRAM 272 PCR	AACTACAAAGGGGCATATCATAGC
H9-PRAM4 -118Fw	CTCCCTCCTTCCCAGCTTAAGTTC
H9-PRAM3 -574Fw2	AACCCAACATTAGTAGTTGGTTCTG
2F AS	CATTTTCACGATATTGATTCTTCC
2H sentido	GATATCATCTCACACCAGTTAG
SV40 REV FLP2	TAAAGCAATAGCATCACAAATTCAC
Neo Fw	TGGATACTTTCTCGGCAGGAGC
Neo Rev	CATTGAACAAGATGGATTGCACGC
Neo 6071	CGTTTGTTCGGATCGGGTTAGTTAC
5'UTRAS	CTGGCACTCCCTAGTGAGATG

***3.5.19. Anexo II. Controles de calidad:***  
***Sonicaciones y PCRs de genotipado***

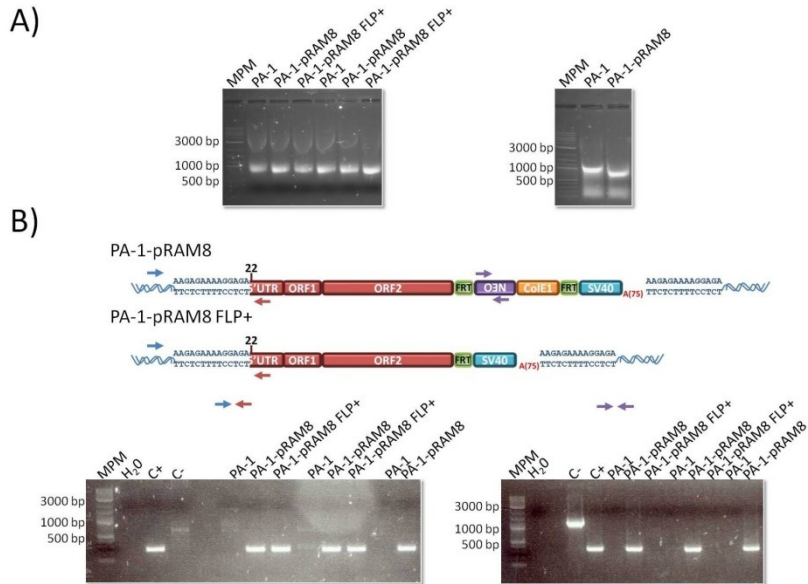




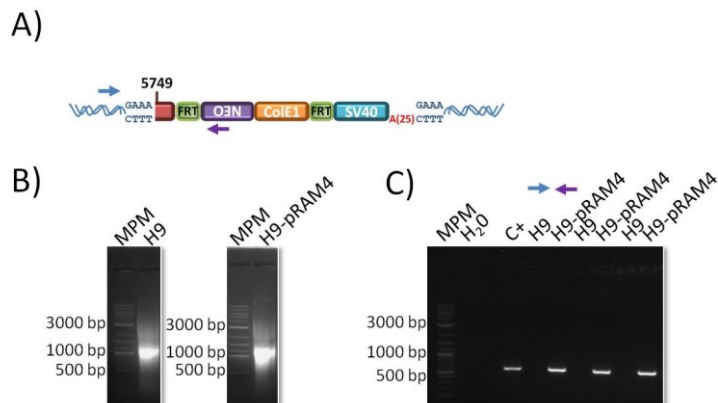
**Figura 7: Controles de calidad del experimento de ChIP-qPCR mostrado en figura 34B .A)** Sonificaciones de las tres réplicas biológicas utilizadas en el experimento. **B)** Arriba, esquema de la inserción de la línea PA-1-pRAM4. Las flechas indican el posicionamiento de los oligonucleótidos para el genotipado (PRAM4 PCR INSERT 2, flecha azul; Neo 6071, flecha morada). Abajo, PCRs de genotipado de las tres réplicas biológicas utilizadas en el experimento.



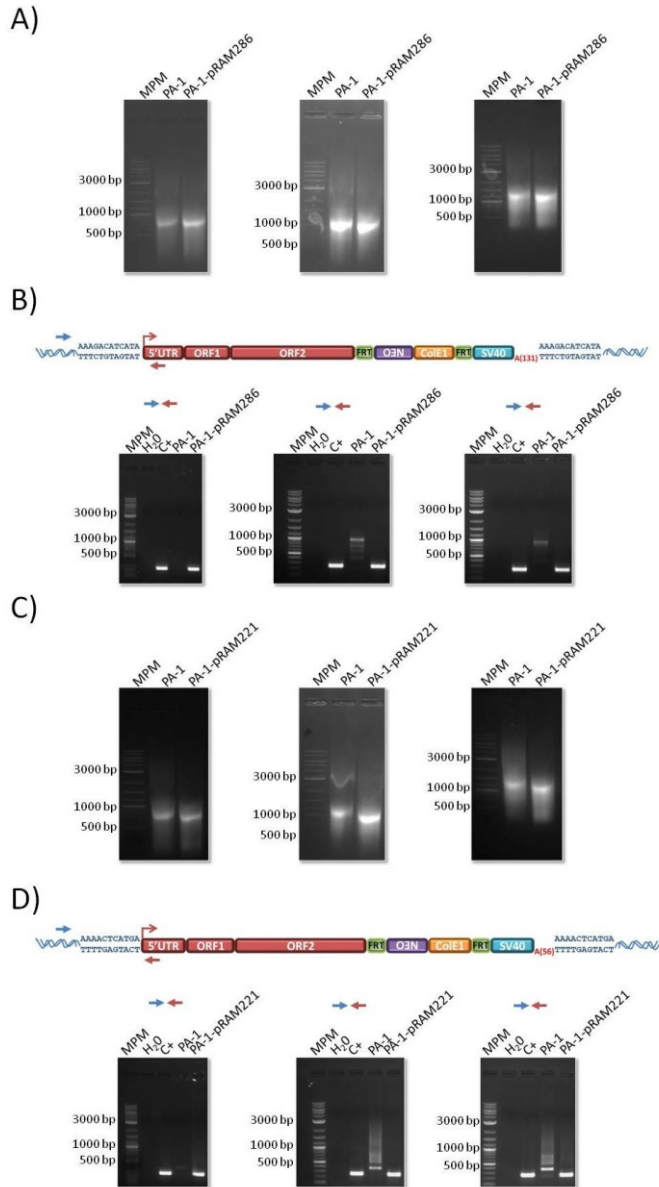
**Figura 8: Controles de calidad de los experimentos de ChIP-qPCR mostrados en figura 34E y 39I. A)** Esquema de la inserción de la línea H9-pRAM3. Las flechas indican la posición de los oligonucleótidos para el genotipado (H9-pRAM3 -547 Fw2, flecha azul; Neo 6071, Flecha morada). **B)** Sonificación representativa de tres réplicas biológicas. **C)** PCRs de genotipado de las tres réplicas biológicas utilizadas en el experimento de ChIP-qPCR.



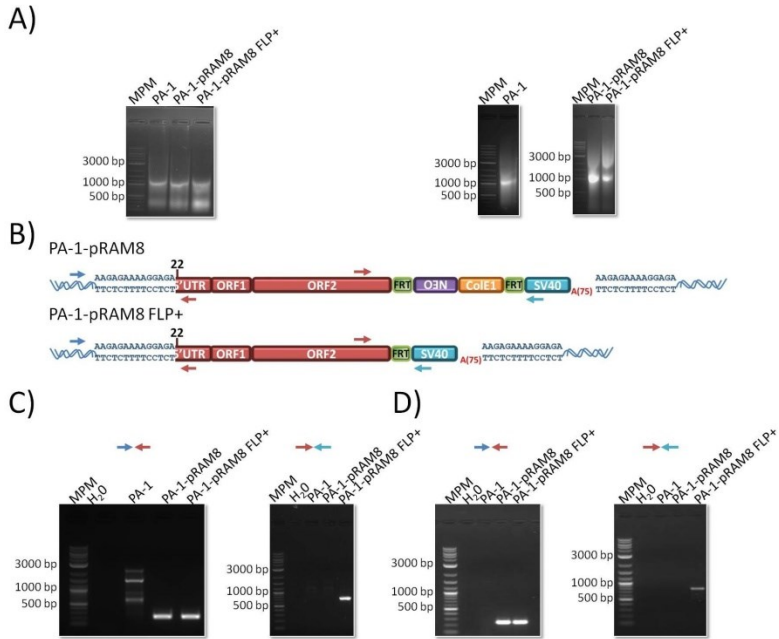
**Figura 9: Controles de calidad de los experimentos de CHIP-qPCR mostrados en figura 35E y 48B.** **A)** Sonicaciones de las réplicas biológicas utilizadas en el experimento de CHIP-qPCR. Las flechas indican la posición de los oligonucleótidos para el genotipado (de izquierda a derecha: PRAM8 Chrm1 PCR, flecha azul; 5'UTRAS, flecha roja; Neo Fw, Neo Rev; flechas moradas). **B)** Arriba, esquema de la inserción de la línea PA-1-pRAM8 antes y después del tratamiento con flipasa. Las flechas indican la posición de los oligonucleótidos para el genotipado. Abajo, PCRs de genotipado en las réplicas biológicas utilizadas en el experimento.



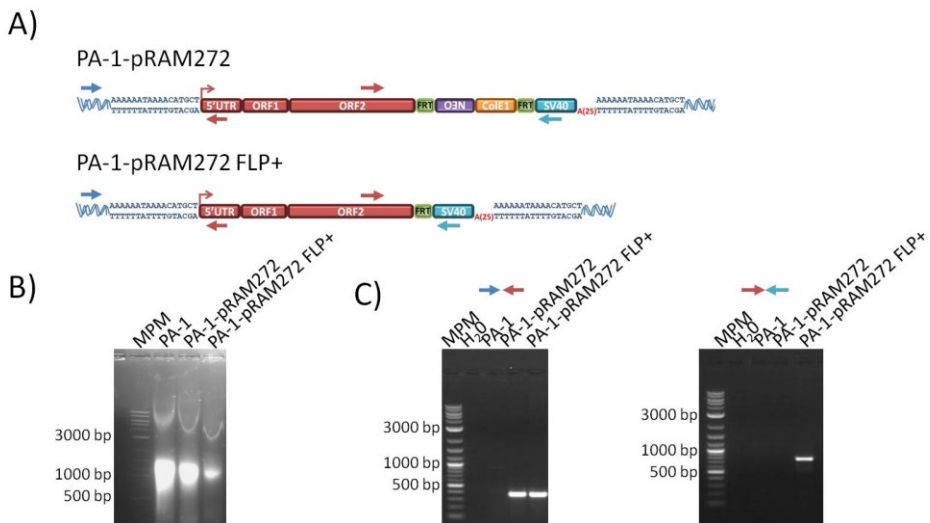
**Figura 10: Controles de calidad de los experimentos de CHIP-qPCR mostrados en figura 36C y 39G.** **A)** Esquema de la inserción en la línea H9-pRAM4. Las flechas indican la posición de los oligonucleótidos para el genotipado (H9-pRAM4 -118 Fw, flecha azul; Neo 6071, flecha morada). **B)** Sonicación representativa de las tres réplicas biológicas utilizadas en el experimento de CHIP-qPCR. **C)** PCRs de genotipado de las tres réplicas biológicas utilizadas en el experimento de CHIP-qPCR.



**Figura 11: Controles de calidad de los experimentos de ChIP-qPCR mostrados en figura 38B y 38E. A)** Sonicaciones experimento ChIP-qPCR PA-1-pRAM286. **B)** Arriba, esquema de la inserción de la línea PA-1-pRAM286. PCRs de genotipado de las tres réplicas biológicas del experimento de ChIP-qPCR en PA-1-pRAM286. **C)** Sonicaciones experimento ChIP-qPCR PA-1-pRAM221. **D)** Arriba, esquema de la inserción de la línea PA-1-pRAM221. PCRs de genotipado de las tres réplicas biológicas del experimento de ChIP-qPCR en PA-1-pRAM221. En **B)** y **D)**, las flechas indican la posición de las oligonucleótidos para el genotipado (PRAM 286 Chrm 20 PCR y PRAM 221 Chrm 1 PCR, flecha azules; 5'UTRAS, flecha roja)

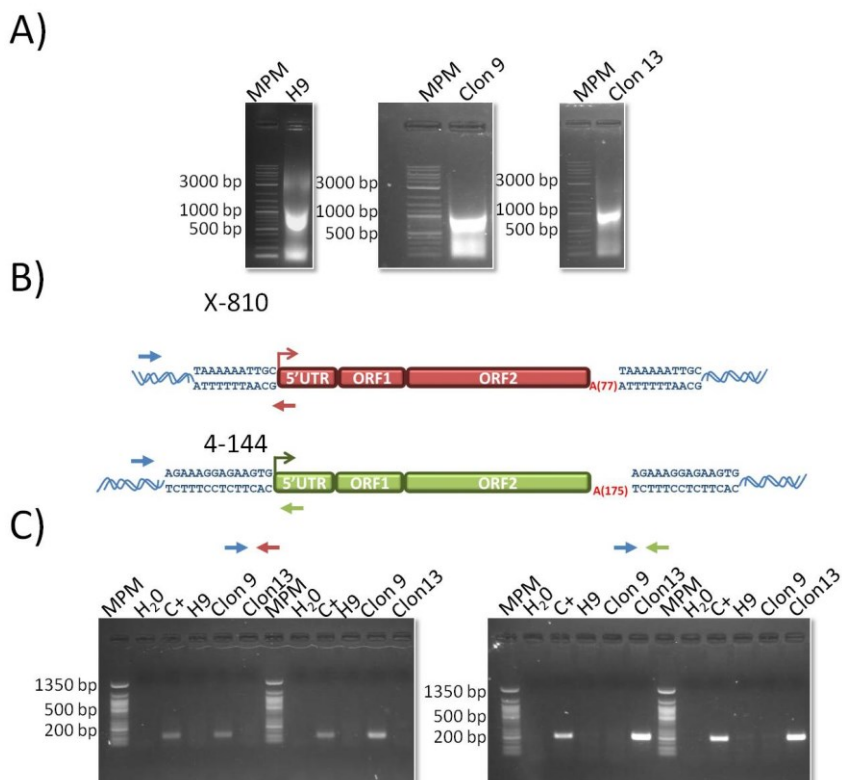


**Figura 12. Controles de calidad de los experimentos de ChIP-qPCR mostrados en figura 39A y 48E**  
**A)** Sonidaciones de las dos réplicas biológicas utilizadas en el experimento de ChIP-qPCR. **B)** Esquema de la inserción de la línea PA-1-pRAM8 antes y después del tratamiento con flipasa. Las flechas indican el posicionamiento de los oligonucleótidos utilizados para el genotipado (de izquierda a derecha: PRAM8 Chrm1 PCR, flecha azul, 5'UTRAS y 2H sentido flechas rojas; SV40 REV FLP2, flecha turquesa). **C)** Genotipado de la primera réplica biológica. **D)** Genotipado de la segunda réplica biológica.

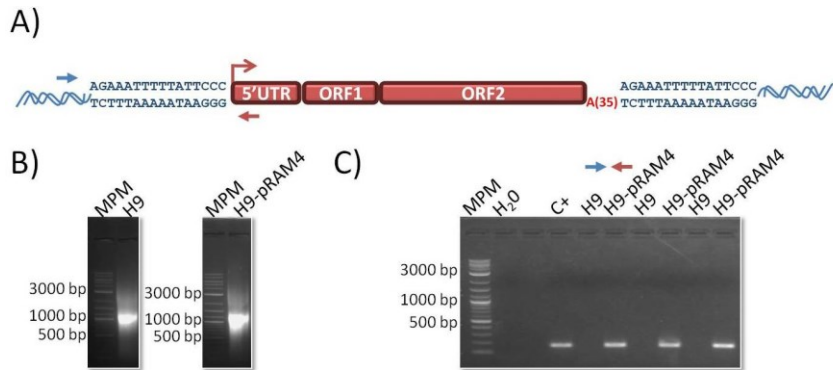




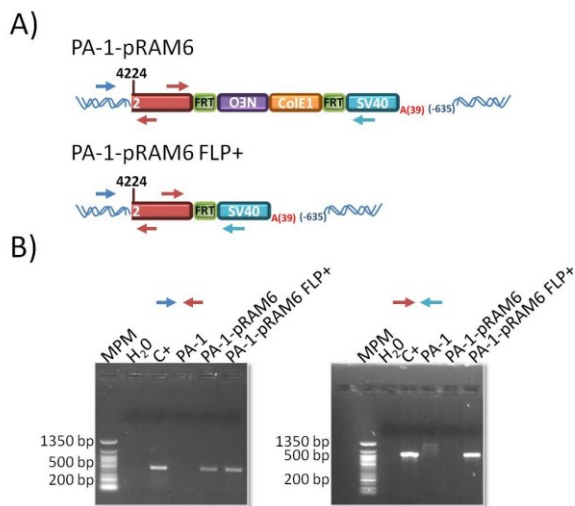
**Figura 13. Controles de calidad de los experimentos de ChIP-qPCR mostrados en figura 40B, C y 48H. A)** Esquema de la inserción de la línea PA-1-pRAM272 antes y después del tratamiento con flipasa. Las flechas indican el posicionamiento de los oligonucleótidos utilizados para el genotipado (de izquierda a derecha: PRAM272 PCR, flecha azul; 5'UTRAS y 2H sentido, flechas rojas; SV40 REV FLP2, flecha turquesa). **B)** Sonicación del experimento de ChIP-qPCR. **C)** PCRs de genotipado.



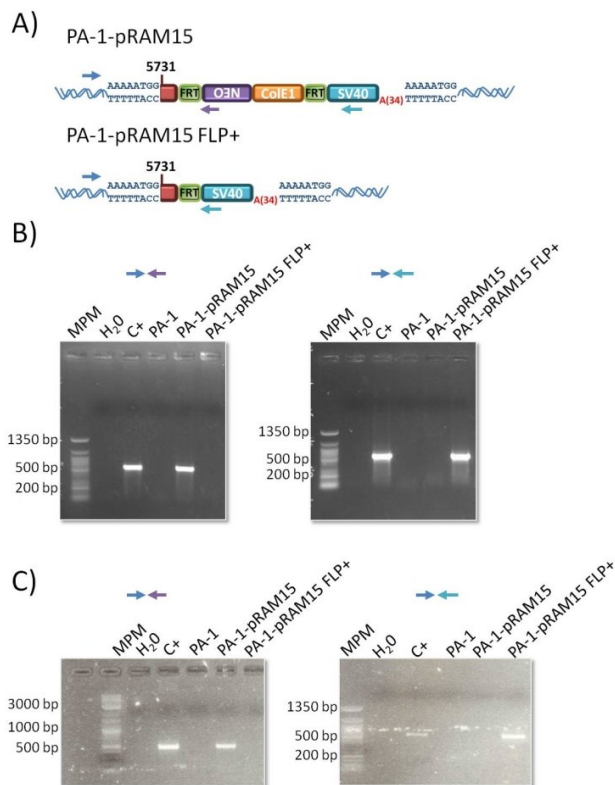
**Figura 14: Controles de calidad de los experimentos de ChIP-qPCR mostrados en las figuras 55B, C. A)** Sonicación representativa de las dos réplicas biológicas utilizadas para los experimentos de ChIP-qPCR. **B)** Esquema de las inserciones (X-810 y 4-144) de los subclones H9-pRAM4.9 y H9-pRAM4.13. Las flechas indican el posicionamiento de los oligonucleótidos utilizados para el genotipado (X-810 Fw y 4-144 Fw, Flecha azules; X-810 Rev, flecha roja; 4-144 Rev, flecha verde). **C)** PCRs de genotipado de las dos réplicas biológicas utilizadas.



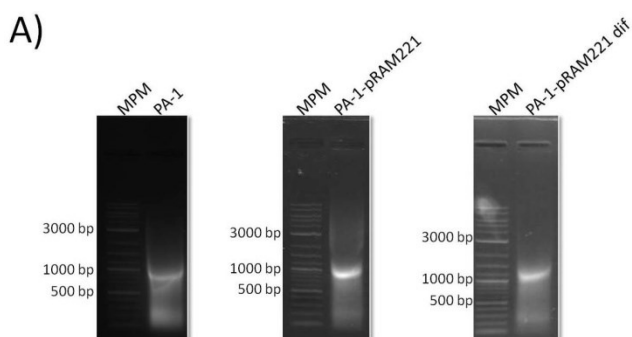
**Figura 15: Controles de calidad del experimento de ChIP-qPCR mostrados en figura 56B. A)** Esquema de la inserción 1-927. Las flechas indican el posicionamiento de los oligonucleótidos utilizados para el genotipado (1-927 Fw, Flecha azul; 1-927 Rev, flecha roja). **B)** Sonicación representativa de las tres réplicas biológicas. **C)** Genotipado de la inserción en las tres réplicas biológicas utilizadas para el experimento de ChIP-qPCR.



**Figura 16. PCRs de genotipado de los experimentos de ChIP-qPCR mostrados en figura 35B, 39C y 48C, F .A)** Esquema de la inserción de la línea PA-1-pRAM6 antes y después de la recombinación mediada por flipasa. Las flechas indican la posición de los oligonucleótidos para el genotipado (de izquierda a derecha: PRAM6 PCR INSERT 2, flecha azul; 2FAS y 2HAS, flechas rojas; SV40 REV FLP 2, flecha turquesa). **B)** PCRs de genotipado.



**Figura 17: PCRs de genotipado de los experimentos de ChIP-qPCR mostrados en figuras 36A y 48D, G. A)** Esquema de la inserción de la línea PA-1-pRAM15 antes y después de la recombinación mediada por Flipasa. Las flechas indican el posicionamiento de los oligonucleótidos antes y después de la recombinación mediada por flipasa (de izquierda a derecha: PRAM15 PCR INSERT 2, flecha azul; Neo 6071, flecha morada; SV40 REV FLP 2, flecha turquesa). **B)** Genotipado primera réplica biológica. **C)** Genotipado segunda réplica biológica.



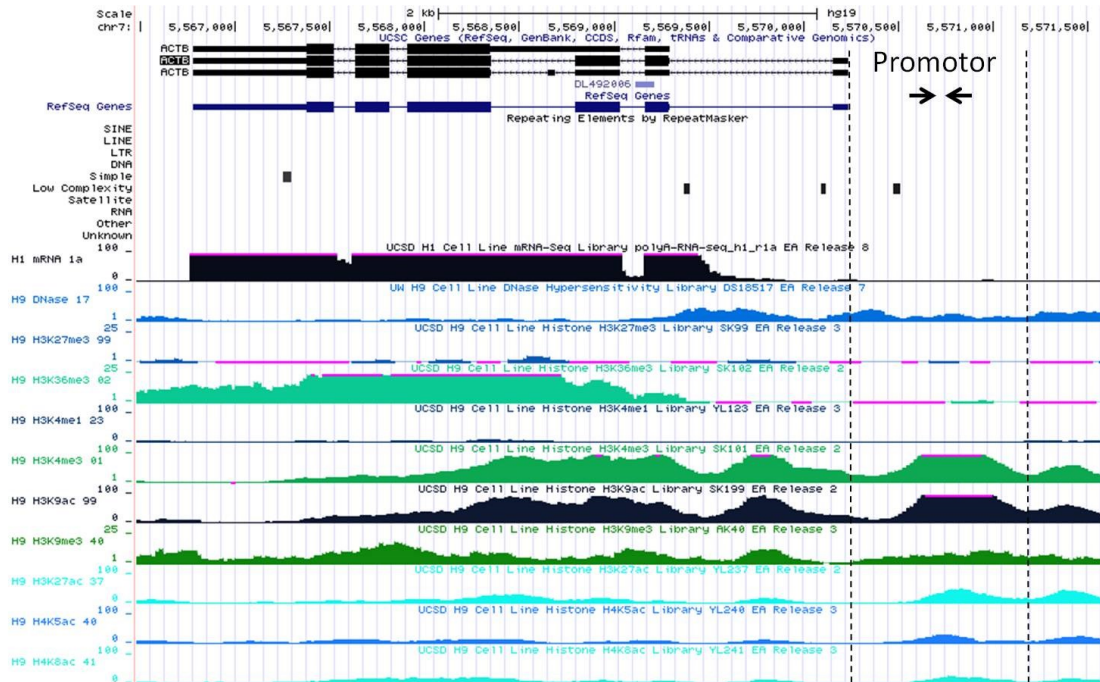
**Figura 18: Control de calidad del experimento de ChIP-qPCR mostrado en figura 50. A)** Se muestra la sonicación del experimento de ChIP-qPCR en la línea PA-1-pRAM221 en estado diferenciado junto con la línea parental PA-1 y la línea clonal PA-1-pRAM221 en estado pluripotente.



3.5.20. Anexo III: Controles biológicos utilizados en los experimentos de  
ChIP-qPCR

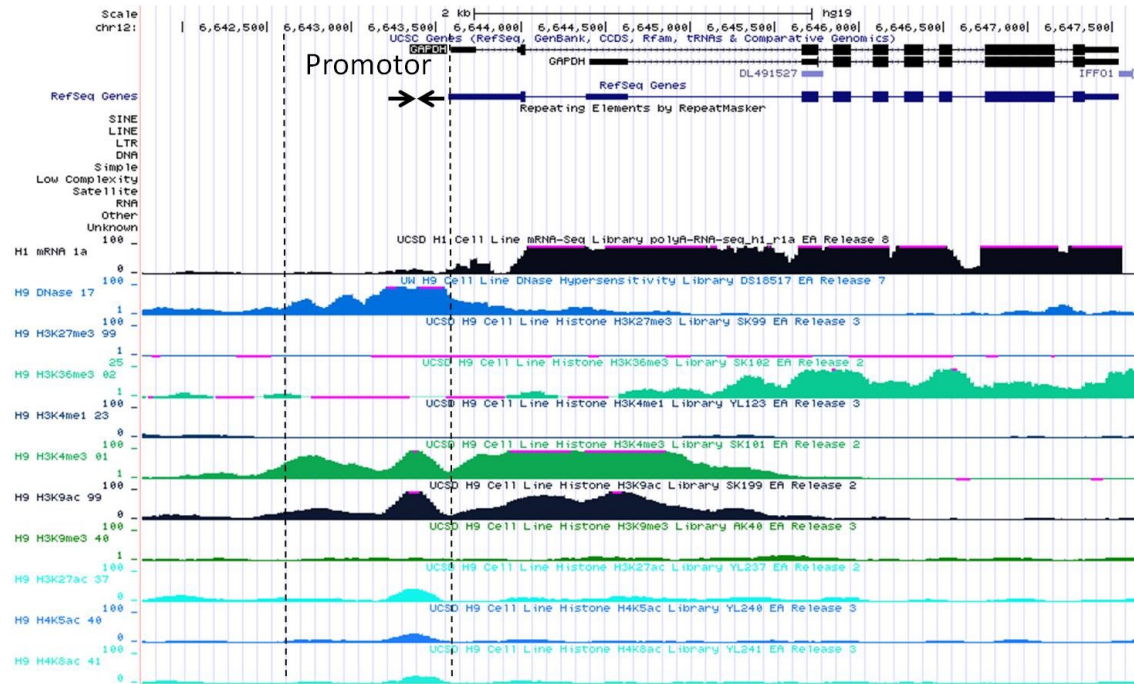


## ACTB: control positivo de enriquecimiento de H3K4me3 y H3K9ac y negativo para H3K9me3 y H3K27me3



Los oligonucleótidos para CHIP-qPCR fueron diseñados aguas arriba del primer exón, donde se suele encontrar la región promotora que viene delimitada aproximadamente por las líneas discontinuas. Las dos flechas negras marcan la región donde se diseñaron los oligonucleótidos.

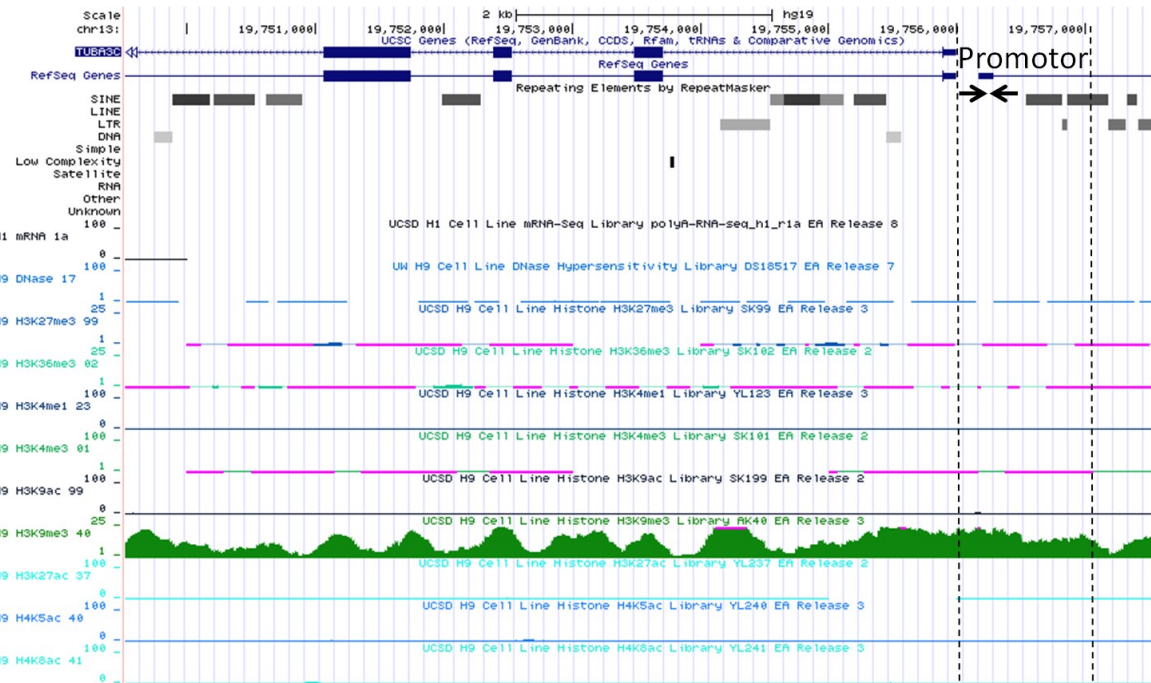
## GAPDH: Control positivo de enriquecimiento para H4ac y negativo para H3K9me3



Los oligonucleótidos para CHIP-qPCR fueron diseñados aguas arriba del primer exón, donde se suele encontrar la región promotora que viene delimitada aproximadamente por las líneas discontinuas. Las dos flechas negras marcan la región donde se diseñaron los oligonucleótidos.

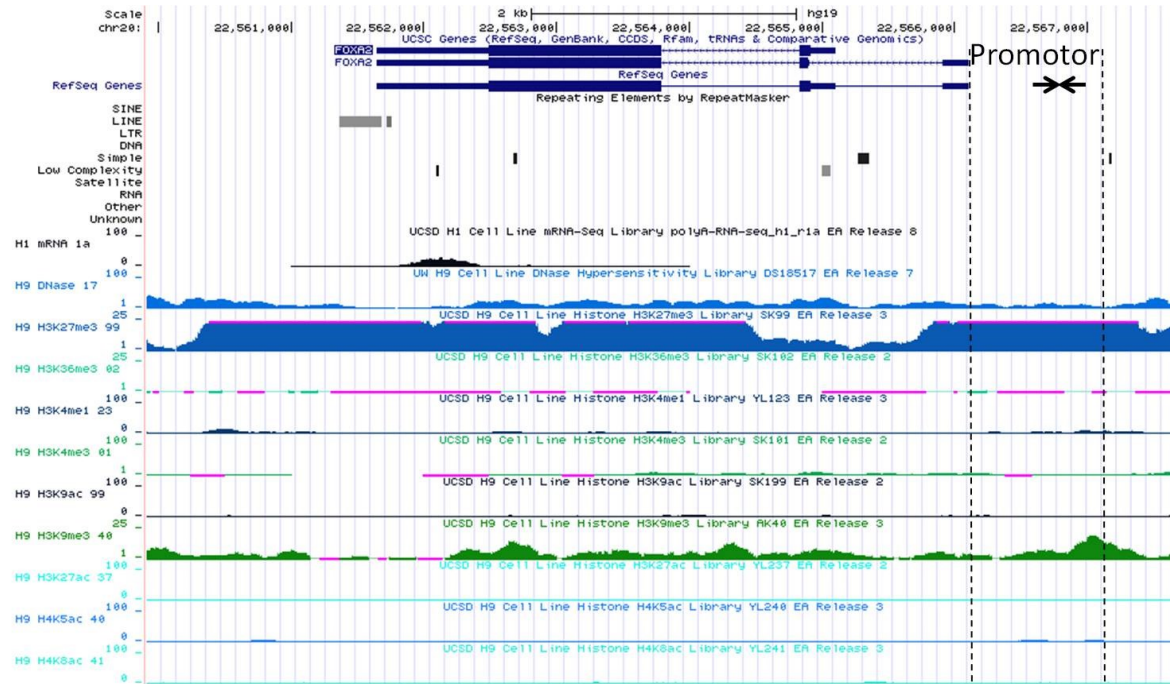


## TUBA3C: control negativo de enriquecimiento para H3K4me3, H3K9ac y H4ac



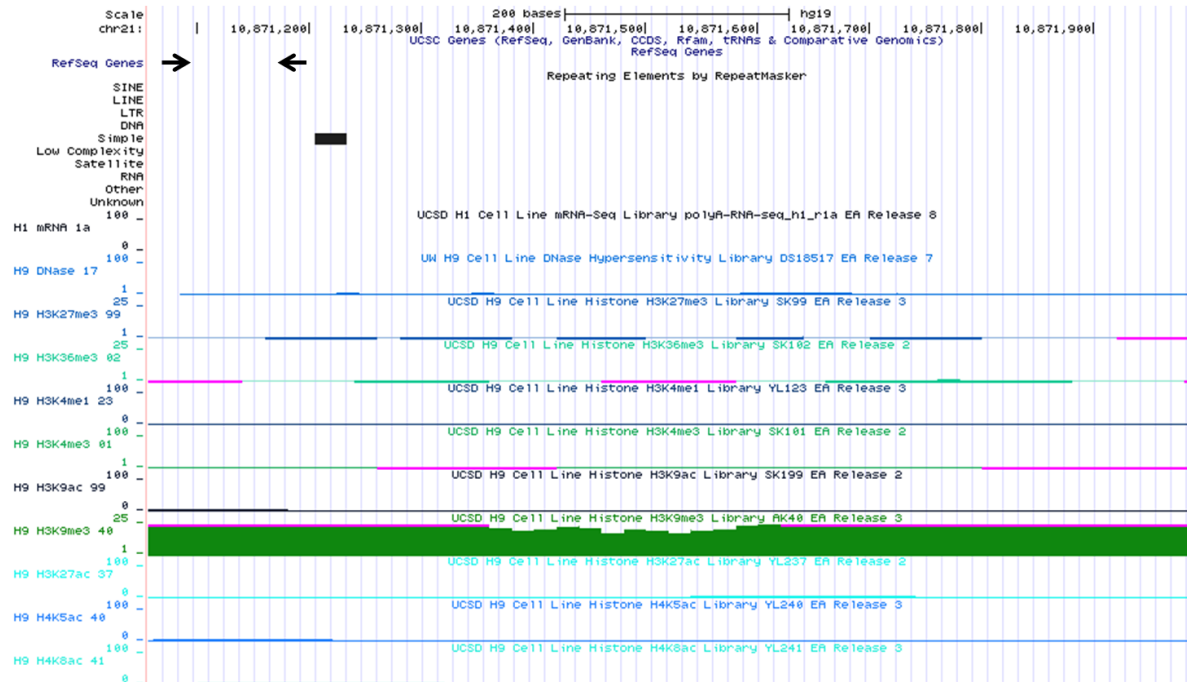
Los oligonucleótidos para CHIP-qPCR fueron diseñados aguas arriba del primer exón, donde se suele encontrar la región promotora que viene delimitada aproximadamente por las líneas discontinuas. Las dos flechas negras marcan la región donde se diseñaron los oligonucleótidos.

## FOXA2: Control positivo de enriquecimiento para H3K27me3



Los oligonucleótidos para CHIP-qPCR fueron diseñados aguas arriba del primer exón, donde se suele encontrar la región promotora que viene delimitada aproximadamente por las líneas discontinuas. Señalado con dos flechas negras se marca la región donde se diseñaron los oligonucleótidos.

## CHR21: región intergénica enriquecida en H3K9me3. Control positivo de enriquecimiento de H3K9me3 y negativo para H3K4me3



Los oligonucleótidos para ChIP-qPCR fueron diseñados justamente en la zona señalada con las flechas negras.



# **RESULTADOS**



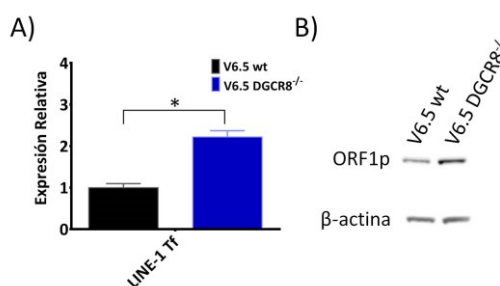
## 4. Resultados

### 4.1. Estudiar el papel del microprocesador en la regulación del retroelemento LINE-1

Este objetivo, se centrará en determinar si el complejo DROSHA/DGCR8 interviene durante el ciclo de vida del retroelemento LINE-1 afectando de alguna forma a su movilidad utilizando fundamentalmente el ensayo de retrotransposición *in vitro* en células en cultivo (ver introducción para más detalles) [12].

#### 4.1.1. Efecto del microprocesador (DROSHA/DGCR8) sobre los niveles de ARNm y proteína de LINE-1

Mediante HITS-CLIP (del inglés *High-throughput sequencing and crosslinking immunoprecipitation*) de DGCR8, Macías *et al.*, [124] demostraron que en la secuencia del ARN del retroelemento LINE-1, existen distintos lugares de unión para DGCR8. DGCR8 es la proteína compañera de DROSHA en el microprocesador que reconoce la estructura del pri-microRNA en el núcleo uniéndose al mismo y posibilitando así el corte por parte de DROSHA y generándose de esta forma un pre-microRNA que sale al citoplasma para seguir siendo procesado hasta dar lugar a un microRNA maduro [123]. Por tanto, la unión de DGCR8, podría conllevar que el ARN del retroelemento LINE-1 sea procesado por el complejo microprocesador y consecuentemente, que se reduzcan los niveles de dicho transcrito. Para comprobar esta hipótesis se determinaron los niveles de ARNm de LINE-1 de la subfamilia T<sub>F</sub> en células embrionarias de ratón que carecen de ambos alelos de DGCR8 (V6.5 DGCR8<sup>-/-</sup>) mediante RT-qPCR. Se utilizan células pluripotentes ya que éstas sobreexpresan de forma natural los elementos LINE-1 [36], [44], [59].



**Figura 19: Efecto del microprocesador (DROSHA/DGCR8) sobre los niveles endógenos de ARN y proteína de LINE-1. A)** RT-qPCR que muestra los niveles de ARN de una subfamilia de LINE-1 murino (T<sub>F</sub>) en presencia/ausencia de DGCR8 en células embrionarias de ratón V6.5. Los valores de expresión fueron normalizados respecto a β-actina. Se muestra la media de tres experimentos independientes

con su correspondiente desviación estándar. Las diferencias se contrastaron estadísticamente con un *T* de student, obteniéndose un *p*-valor=0,019. **B)** Western Blot que muestra los niveles de proteína de LINE-1 (ORF1p) en presencia/ausencia de DGCR8 en células embrionarias de ratón. Como control de carga se utilizó  $\beta$ -actina

Cómo se puede observar en la Figura 19A, la ausencia de DGCR8 conlleva un aumento significativo que equivale al doble del nivel original de ARNm de LINE-1 T<sub>F</sub>.

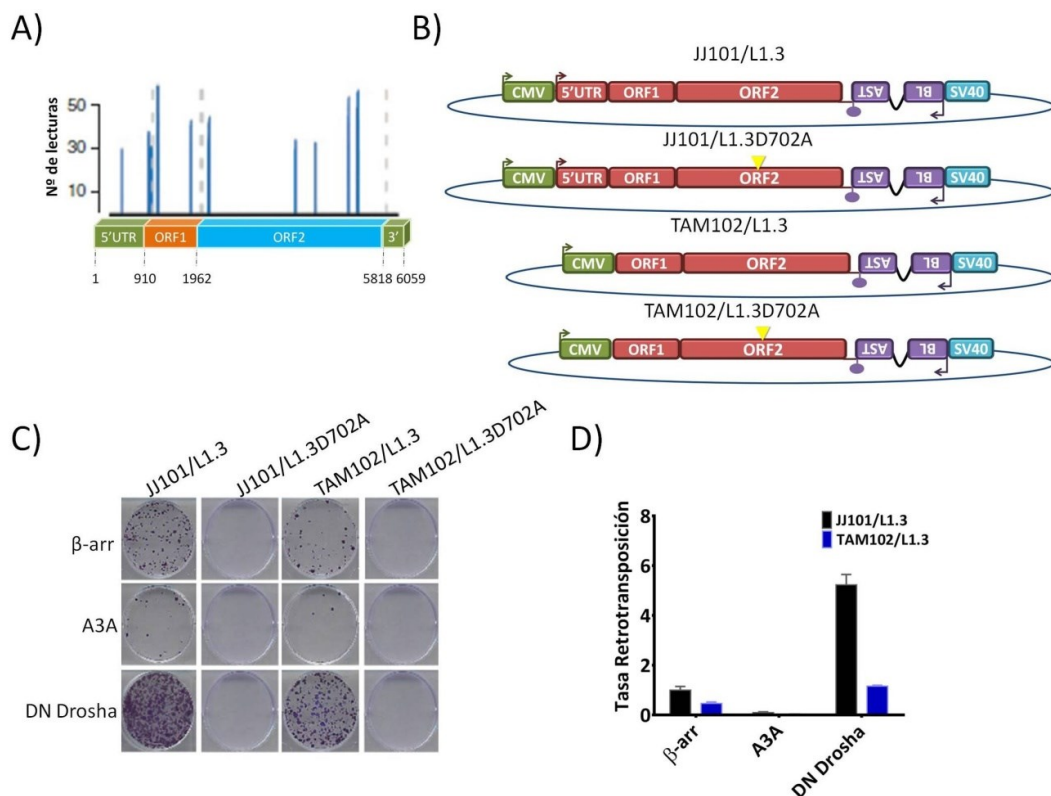
Seguidamente, se comprobó si este aumento de ARNm, se traduciría en un aumento del nivel de proteína de LINE-1. Para ello, se analizó mediante western blot los niveles de ORF1p, de nuevo, en células V6.5 wt y V6.5 DGCR8<sup>-/-</sup>. Como se podía esperar, ese aumento de ARNm se tradujo en una mayor cantidad de proteína de LINE-1 en células carentes de microprocesador, como se puede observar en Figura 19B. De esta forma, el microprocesador reduce los niveles de ARN y, por consiguiente, los niveles de proteína de LINE-1.

#### 4.1.2. Efecto del microprocesador (DROSHA/DGCR8) sobre la retrotransposición.

Para determinar si el aumento de ARNm de LINE-1 así como de proteína, mediado por la ausencia parcial de un microprocesador funcional, podría eventualmente dar lugar a un mayor número de inserciones de LINE-1, se llevó a cabo un ensayo de retrotransposición [12]. Brevemente, como se describió en la introducción, el ensayo de retrotransposición permite medir la retrotransposición a través de la actividad de un gen reportero fusionado a la inserción de LINE-1. El diseño del vector de retrotransposición asegura que el gen reportero solo sea funcional después de una ronda de retrotransposición. Idealmente, en este caso, el ensayo de retrotransposición se debería hacer en células embrionarias ya que las nuevas inserciones de LINE-1 se acumulan mayoritariamente durante el desarrollo embrionario temprano [77], [78], sin embargo, por motivos técnicos los niveles de movilidad de L1 *in vitro* en células madre embrionarias son muy bajos. Por tanto, se utilizaron células HeLa ya que se ha demostrado que albergan una elevada tasa de retrotransposición *in vitro* lo cual las hace buenas candidatas para comprobar cómo distintos elementos reguladores actúan sobre la retrotransposición [12], [113]. Según el análisis de las lecturas del HITS-CLIP de DGCR8, podrían existir varios lugares de unión para DGCR8 sobre la región 5'UTR del ARN de LINE-1 (Figura 20A). Por tanto, se decidió realizar el ensayo de retrotransposición utilizando dos vectores diferentes: el primero, JJ101/L1.3, que contiene un promotor exógeno (CMV) además de la región 5'UTR de LINE-1 con actividad promotora mientras que el otro carece de dicha región (TAM102/L1.3) y la transcripción es únicamente dirigida por un promotor exógeno (CMV) (Figura 20B). El gen reportero de ambos constructos es el gen de resistencia a blasticidina. Así, la



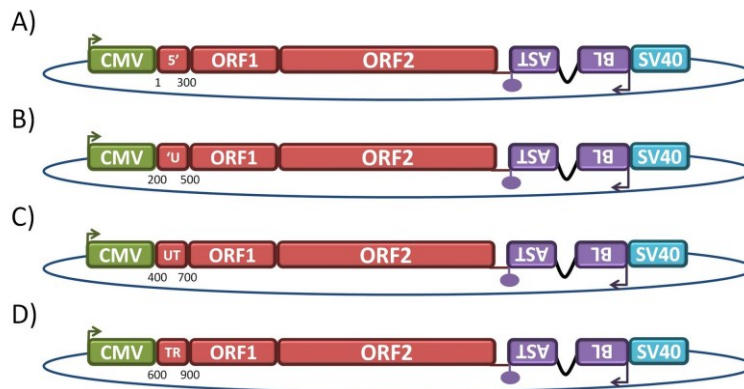
retrotransposición será medida como colonias resistentes después de un periodo de selección con dicho antibiótico. De esta forma, se intentó validar funcionalmente si el ARN de LINE-1 es una diana del microprocesador. Para ello las células fueron transfectadas con



**Figura 20: Ensayo de retrotransposición en células HeLa en presencia de un microprocesador parcialmente funcional (DN-Drosha).** **A)** Mapeo de las lecturas del HITS-CLIP sobre el ARN de LINE-1. **B)** Esquema de los vectores de retrotransposición que se han utilizado para el ensayo. Todos los constructos llevan fusionado un gen de resistencia a blasticidina. JJ101/L1.3D702A y TAM102/L1.3 son controles negativos. El triángulo amarillo señala la posición de la mutación. **C)** Ensayo de retrotransposición en células HeLa con los constructos descritos en B) co-transfectados bien con B-arrestina ( $\beta$ -arr) bien con una proteína que inhibe la retrotransposición APOBEC3A (A3A) o con una forma dominante negativa de DROSHA (DN Drosha). Se muestra un ensayo representativo de tres réplicas biológicas. Cada punto azul representa un evento de retrotransposición.  $\beta$ -Arr ( $\beta$ -arrestina), A3A (Apobec3A) y DN Drosha (Dominante negativo de Drosha). **D)** Cuantificación del ensayo de retrotransposición. Las colonias azules fueron cuantificadas manualmente para determinar la tasa de retrotransposición en cada caso. La tasa de retrotransposición de cada vector fue normalizada respecto a la tasa del vector JJ101/L1.3 co-transfectado con B-arrestina que toma valor 1. La tasa de retrotransposición de cada vector fue normalizada respecto a la eficiencia de transfección.

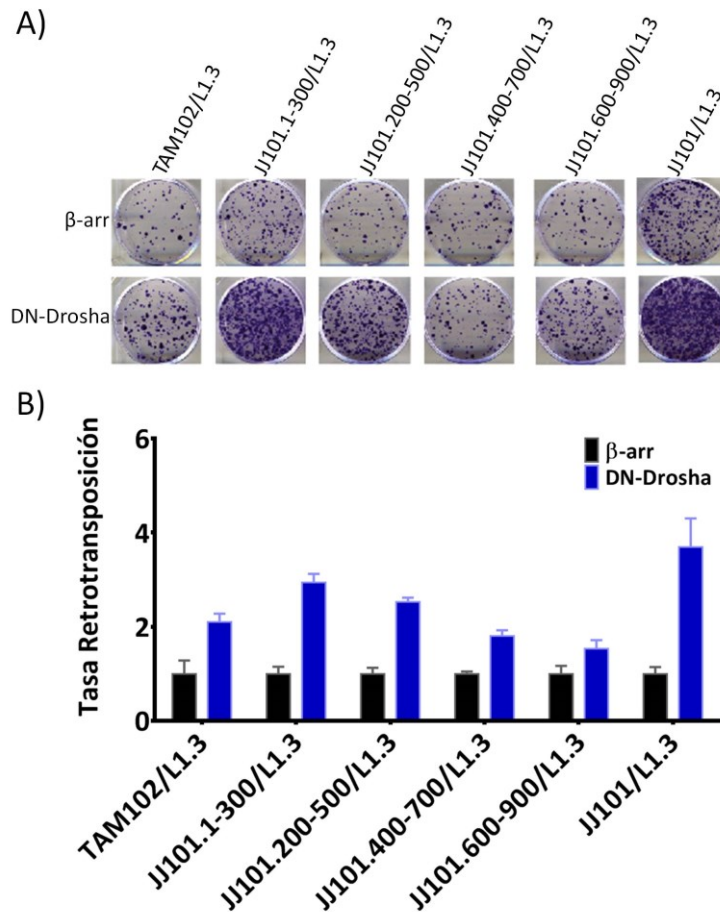
dichos constructos además de sus correspondientes mutantes en el dominio reverso transcriptasa que se incluyen en el ensayo como controles negativos de retrotransposición (JJ101/L1.3D702A; TAM102/L1.3D702A) (Figura 20B). Los vectores de retrotransposición fueron co-transfectados con un plásmido que expresa una proteína mutante que induce un efecto dominante negativo de la proteína DROSHA (DN-Drosha) o un plásmido que expresa A3A (APOBEC3A) que es una proteína inhibidora de la retrotransposición o bien un plásmido que expresa una proteína ( $\beta$ -arrestina) que no influye sobre la retrotransposición *in vitro* [217]. En estos experimentos se observó, que efectivamente, el microprocesador controla la movilidad del retroelemento LINE-1 ya que su ausencia conllevó un aumento de la retrotransposición en torno a 5 veces comparando la retrotransposición desde el vector JJ101/L1.3 en presencia de  $\beta$ -arr y en presencia del DN-Drosha (Figura 20C, D). Además, se demostró que existen sitios de unión del microprocesador en la región 5' no traducida (5'UTR) del retroelemento LINE-1 (Figura 20C, D) con importancia funcional ya que la ausencia de la región 5'UTR (TAM102/L1.3) conllevó una clara reducción en el incremento de retrotransposición mediado por un microprocesador no funcional (DN-Drosha) (Figura 20C, D).

En resumen, este ensayo demostró que el microprocesador regula negativamente la retrotransposición de LINE-1 en células en cultivo.



**Figura 21: Constructos derivados del JJ101L1.3. A)** JJ101.1-300/L1.3. Constructo cuya 5'UTR ha sido reemplazada por la porción de la misma del nucleótido 1 al 300. **B)** JJ101.200-500/L1.3. Constructo cuya 5'UTR ha sido reemplazada por la porción de la misma del nucleótido 200 al 500. **C)** JJ101.400-700/L1.3. Constructo cuya 5'UTR ha sido reemplazada por la porción de la misma del nucleótido 400 al 700. **D)** JJ101.600-900/L1.3. Constructo cuya 5'UTR ha sido reemplazada por la porción de la misma del nucleótido 600 al 900.

Estos resultados sugieren que la región 5'UTR actúa como diana directa del microprocesador. Para validar los lugares de unión para DGCR8 contenidos en esta región



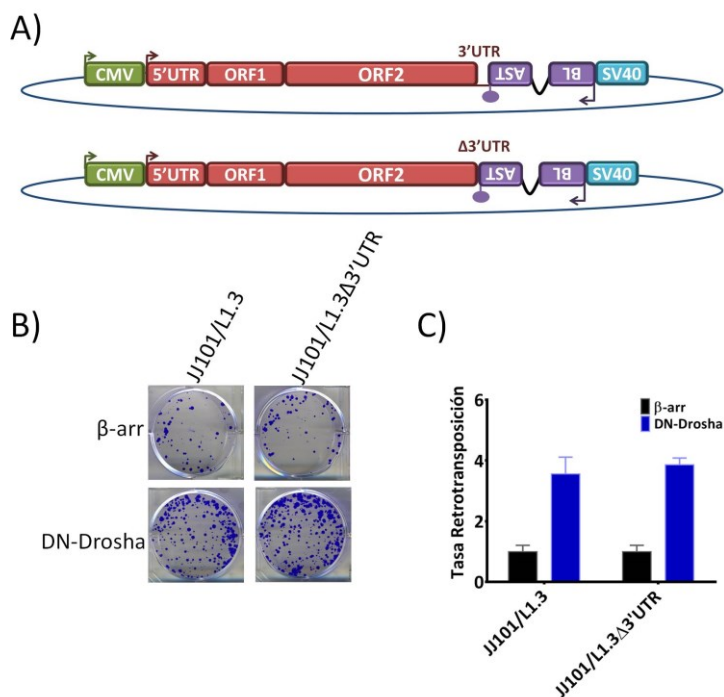
**Figura 22: Mapeo funcional de los sitios de unión de DGCR8 mediante un ensayo de retrotransposición en ausencia de un microprocesador funcional. A)** Ensayo de retrotransposición en células HeLa con los constructos descritos en figura 21 junto con el JJ101/L1.3 y TAM102/L1.3 descritos en figura 20B. Se muestra un ensayo de retrotransposición representativo de 3 réplicas biológicas. Cada punto azul representa un evento de retrotransposición.  $\beta$ -arr ( $\beta$ -arrestina) y DN-Drosha (Dominante Negativo de Drosha). **B)** Cuantificación del ensayo de retrotransposición. Las colonias azules fueron cuantificadas manualmente para determinar la tasa de retrotransposición en cada caso. La tasa de retrotransposición de cada vector co-transfectado con el DN-Drosha fue normaliza respecto a la tasa de retrotransposición de cada vector co-transfectado con B-arrestina que toma valor 1. La tasa de retrotransposición de cada vector fue normalizada respecto a su eficiencia de transfección.

se generaron 4 constructos derivados del JJ101/L1.3 en los cuales sólo estaba presente una porción de 300 pares de bases de la región 5' UTR. Así, estos plásmidos contenían las regiones comprendidas entre 0-300; 200-500; 400-700 y 600-900 (Figura 21). Al igual que

se hizo anteriormente, se llevó a cabo un ensayo de retrotransposición con los nuevos constructos más el JJ101/L1.3 y el TAM102/L1.3 como controles (Figura 20B). Los vectores de LINE-1 fueron co- transfectados nuevamente con  $\beta$ -arrestina y la forma mutante de DROSHA que induce un efecto dominante negativo. Para determinar los fragmentos que albergan lugares de unión para DGCR8, se normalizó la tasa de retrotransposición desde cada constructo co-transfectado con DN-Drosha respecto a la tasa de retrotransposición desde el mismo constructo co-transfectado con  $\beta$ -arrestina. De esta forma, se demostró que la ausencia del microprocesador produce un incremento en los niveles de retrotransposición superior a los observados con el plásmido TAM102/L1.3 (que carece de 5'UTR) cuando el constructo contiene la región 1-300 y, en menor medida cuando incluye la región 200-500 (Figura 22A, B). Sin embargo, los constructos que contenían la región 400-700 y 600-900 se comportaron de forma similar al TAM102/L1.3 (Figura 22A, B).

De esta forma, estos datos sugieren que la zona del 5'UTR donde se une el microprocesador para regular negativamente la retrotransposición se encuentra principalmente en la región 1-300 del ARN del LINE-1 (Figura 22A,B)

Existen estudios que demuestran que la ausencia de otras proteínas de la ruta de biogénesis de microARNs, como DICER, conlleva un aumento de la retrotransposición mediado por un aumento de la cantidad de ARN de LINE-1 [47]. Por ello nos propusimos determinar si el aumento de la retrotransposición que se observa en ausencia de un microprocesador funcional estaba mediado por una reducción en la cantidad de microRNAs y, por tanto, no era debido a un procesamiento directo del ARN del LINE-1. Teniendo en cuenta que está descrito que la mayoría de los microRNAs se unen a la región 3' no traducida de sus ARNs diana (3'UTR) [220], se decidió utilizar un nuevo constructo cuyo LINE-1 carece de región 3' no traducida ya que esta región no se requiere para la retrotransposición *in vitro* (Figura 23A). De esta forma, se pudo testar si la ausencia de microRNAs debido a la falta de un microprocesador funcional (DN-Drosha) tenía algún efecto sobre la retrotransposición. Así, el nuevo constructo denominado JJ101/L1.3 $\Delta$ 3'UTR se transfectó en paralelo con el constructo JJ101/L1.3 siendo ambos co-transfectados con  $\beta$ -arrestina y DN-Drosha. Igual que anteriormente, la tasa de retrotransposición desde ambos constructos en presencia de DN-Drosha fue normalizada respecto a  $\beta$ -arrestina. El constructo que carece de la región 3'UTR de LINE-1 (JJ101/L1.3 $\Delta$ 3'UTR) dio lugar a una tasa de retrotransposición similar al constructo que la contenía (JJ101/L1.3) cotransfectado en ambos casos con DN-Drosha (Figura 23B, C). Estos datos sugieren que el control del microprocesador sobre la retrotransposición del elemento LINE-1 no está mediado por microRNAs, y por tanto, apoya la hipótesis de que dicho control es un efecto directo mediado por procesamiento del ARN mensajero de LINE-1.



**Figura 23: Ensayo de retrotransposición en ausencia parcial de microARNs por falta de un microprocesador funcional.** **A)** Constructos utilizados para el ensayo de retrotransposición. En este caso se utilizó un derivado del JJ101/L1.3 (descrito en Figura 20B) al que le falta la región 3'UTR (JJ101/L1.3Δ3'UTR). **B)** Ensayo de retrotransposición en células HeLa con los constructos descritos en A). Dichos constructos fueron co-transfectados bien con B-arrestina ( $\beta$ -arr) o bien con la forma dominante negativa (DN-Drosha). Cada punto azul representa un evento de retrotransposición. Se muestra un ensayo representativo de tres réplicas biológicas. **C)** Cuantificación del ensayo de retrotransposición. Las colonias azules fueron cuantificadas manualmente para determinar la tasa de retrotransposición en cada caso. La tasa de retrotransposición de cada vector co-transfectado con el DN-Drosha fue normalizada respecto a la tasa de retrotransposición de cada vector co-transfectado con  $\beta$ -arrestina que toma valor 1. A su vez la tasa de retrotransposición de cada vector fue normalizada respecto a su eficiencia de transfección.

En resumen, el microprocesador (DROSHA/DGCR8) regula negativamente la movilidad del retroelemento LINE-1 probablemente uniéndose principalmente al ARN de éste a través de los primeros 300 nucleótidos de la región 5'UTR. Además, este efecto es directo ya que el efecto del microprocesador sobre la tasa de retrotransposición del elemento L1 mostró ser independiente de la región 3'UTR, y por tanto, de los posibles microRNAs que pudieran regular a LINE-1.

## **4.2. Estudio de la regulación epigenética de las inserciones *de novo* de LINE-1 y del impacto epigenético de las mismas en el genoma adyacente**

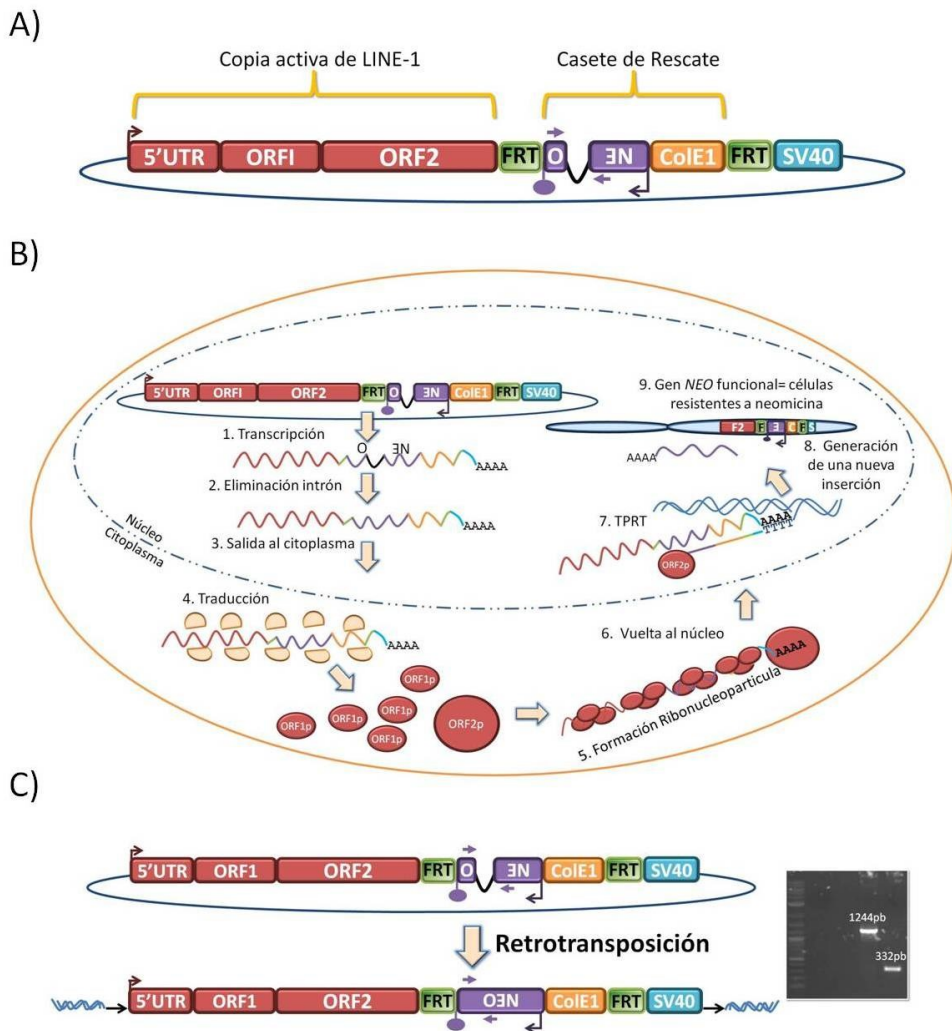
### **4.2.1. Generación de líneas clonales en células PA1 y H9**

Los tipos celulares usados en este estudio son células de teratocarcinoma humano embrionario (PA-1) y células madre embrionarias humanas (H9) ya que las nuevas inserciones endógenas de LINE-1 heredables se generan mayoritariamente en etapas tempranas del desarrollo embrionario [77], [78]. A pesar de su procedencia tumoral, las células PA-1 mantienen características de célula pluripotente, poseen un genoma estable y pueden ser manipuladas genéticamente con sencillez [221]. De hecho, líneas celulares de carcinoma embrionario se utilizan para estudiar procesos biológicos asociados al desarrollo embrionario de forma rutinaria [221]. Además, ambos tipos celulares soportan la retrotransposición *in vitro* de LINE-1 y sobreexpresan los elementos LINE-1 endógenos [44], [60]. Igualmente, según datos de nuestro laboratorio, células PA-1 y H9 son capaces de acumular nuevas inserciones de LINE-1 endógenas a lo largo de distintos pases cuando son mantenidas en cultivo (Muñoz-López et al., en preparación). Por todo ello, estas líneas celulares representan un buen modelo para el estudio propuesto.

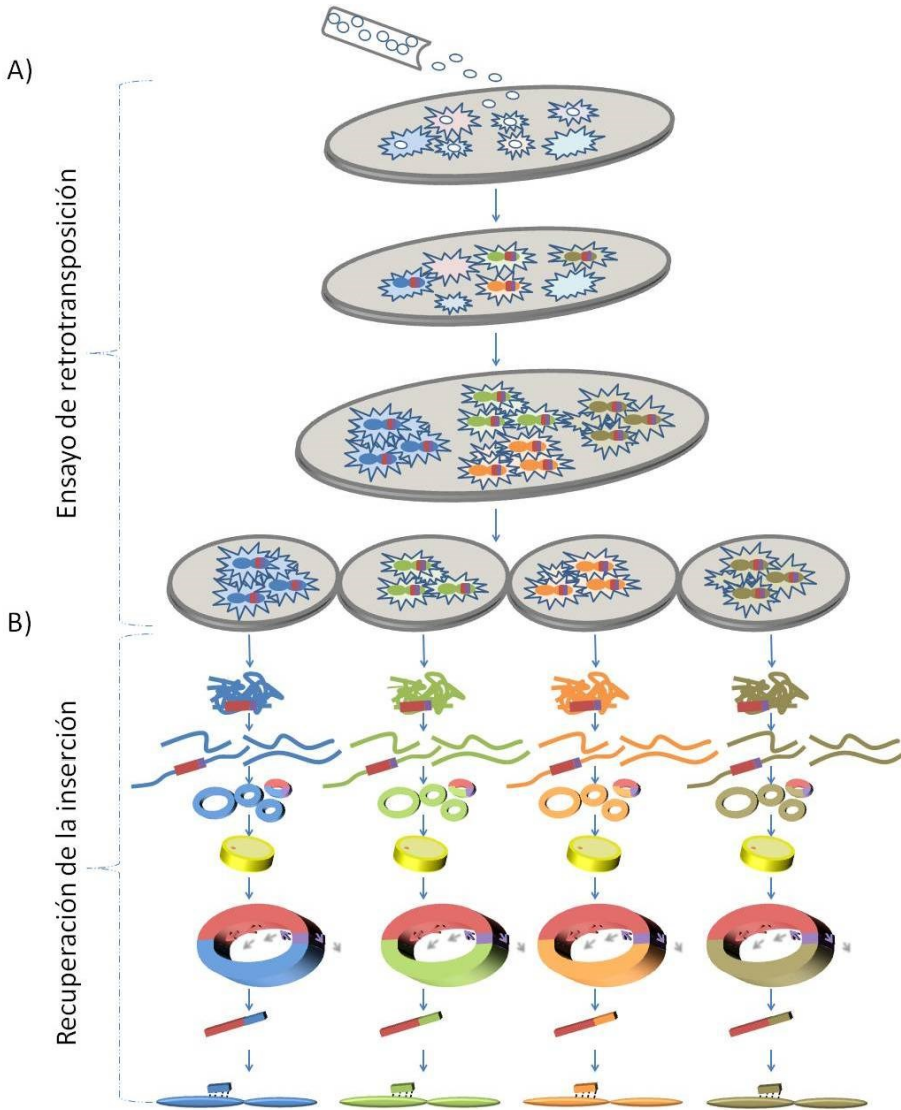
Como primer paso, se generó una serie de líneas clonales que contenían al menos un nuevo evento de retrotransposición artificial de LINE-1. El vector utilizado (pRAM-LRE3) (Figura 24A) es una evolución de un vector anterior ya descrito, que permite el rescate o la recuperación de la inserción y, por tanto, conocer su localización genómica [92]. Según los datos de García-Pérez et al., [60], las células PA-1 soportan una elevada tasa de retrotransposición *in vitro*, sin embargo, ésta es difícilmente detectable mediante la expresión del gen reportero (EGFP) debido al silenciamiento epigenético de las inserciones *de novo*. De esta forma, se intentó reproducir este mismo escenario con el vector pRAM-LRE3 cuyo gen reportero es un gen de resistencia a neomicina. Para ello, se generaron tanto líneas clonales resistentes como líneas clonales sensibles a neomicina, todas ellas con al menos una inserción de *de novo* de LINE-1. A partir de aquí, las líneas clonales generadas serán referidas como resistentes y sensibles.

En primer lugar se comenzó con la generación de líneas clonales resistentes. Para ello, se llevó a cabo, un ensayo de retrotransposición con el vector pRAM-LRE3 como se ha descrito previamente [12] (Figura 24B, detallado en introducción). No obstante, el ensayo se modificó ligeramente ya que después del proceso de selección con neomicina, las colonias resistentes resultantes fueron aisladas y crecidas individualmente hasta generar líneas clonales (Figura 25A). Posteriormente, se procede con el rescate de la inserción

según ha sido previamente descrito [92]. Brevemente, el ADN genómico de cada línea clonal es digerido con una enzima de restricción que corta cada 2000-3000 pares de



**Figura 24: Herramientas para la generación de líneas clonales con inserciones artificiales de novo de LINE-1.** **A)** Esquema del vector pRAM-LRE3. De izda. a dcha., en rojo, copia retrocompetente de LINE-1; en verde, secuencia FRT; casete de rescate formado por el gen de resistencia a neomicina (morado) interrumpido por intrón IV5 de la  $\gamma$ -globina y el origen de replicación bacteriano CoIE1 (amarillo); en verde, secuencia FRT; en azul, señal de poliadenilación SV40. **B)** Ensayo de retrotransposición descrito en Moran et al., [12] utilizando el vector pRAM-LRE3 (más detalles en introducción). **C)** Esquema de la PCR diseñada para detectar nuevos eventos de retrotransposición (PCR de NEO). Las flechas moradas señalan el sitio de hibridación de la pareja de oligonucleótidos antes y después de la retrotransposición. Los productos de PCR generados antes y después de la retrotransposición se observan en el gel de la derecha.



**Figura 25: Esquema del proceso de obtención de líneas clonales resistentes y caracterización de sus inserciones de novo de LINE-1. A)** Ensayo de retrotransposición. Para más detalles ver introducción Las formas estrelladas representan células, los círculos representan el plásmido pRAM-LRE3, las barras rojas/moradas representan la inserción y las formas ovaladas de color azul, verde, naranja y marrón representan los cromosomas. **B)** Recuperación de la inserción. El ADN genómico es digerido con HindIII y SSPI. Los fragmentos resultantes se ligan en condiciones diluidas lo que favorece la ligación intramolecular formándose plásmidos de ADN genómico. La ligación es transformada en bacterias ultracompetentes. La transformación es sembrada en placas de LB con kanamicina. Estos plásmidos de ADN genómico son extraídos desde colonias resistentes. Por último, los plásmidos son secuenciados con oligonucleótidos que hibridan en regiones únicas de la inserción (Casete de Rescate) para encontrar las regiones genómicas ubicadas en los extremos 5' y 3'. En



función de la truncación, se pueden requerir más rondas de secuenciación hasta identificar la región genómica ubicada en el extremo 5' de la inserción. Finalmente, se mapean en el genoma de referencia las regiones genómicas encontradas en el extremo 5' y en el extremo 3'

Línea Clonal	<sup>1</sup> Punto Truncación	<sup>2</sup> DSI	Longitud Poli A	Delección	Localización <sup>3</sup>	Intrón/ <sup>4</sup> Exón	<sup>5</sup> Sentido /Antisentido
PA-1-pRAM4	5963	16	38	No	Crom 2/Gen OSBLP6	Intrón 3 (1,3,4,5)	Antisentido
PA-1pRAM6	4224	No	39	635	Crom 6/Gen FAM65B	Intrón 1 (3,4)	Antisentido
PA-1-pRAM8	22	14	75	No	Crom 1/Gen LPNH2	Intrón 1/8 (1)	Sentido
PA-1-pRAM11	6020	9	73	No	Intergénica		
PA-1-pRAM15	5731	3	39	No	Crom 14/ Gen RAD51B	Intrón 10 (2/3)	Antisentido
PA-1-pRAM26	3738	No	74	No	Crom 5/Intergénica		
PA-1-pRAM27	Copia completa	12	54	No	Crom X/Intergénica		
PA-1-pRAM31	5811	15	61	No	Crom 3/Intergénica		
PA-1-pRAM33	2382	15	45	No	Crom 16/Gen ANKRD26P1	Intrón 8/16	Antisentido
PA-1-pRAM36	4872	No	34	35	Crom 15/Gen THSD4	Intrón 5/16	Sentido
PA-1-pRAM37	6009	9	44	No	Crom 11/Gen ARRB1	Intrón 1 (1,2)	Antisentido
PA-1-pRAM41	Copia completa	17	61	No	Crom 4/Gen MAPK10	Intrón 1/6	Antisentido
PA-1-pRAM45	22	17	40	No	Crom4/Gen FRAS1	Intrón1 (1,2)	Antisentido
PA-1-pRAM29	Copia completa	14	71	No	Crom X/Intergénica		
H9-pRAM3	5744	13	35	No	Crom 2/Gen COL4A4	Intrón 1/47	Antisentido
H9-pRAM4	5749	4	25	No	Crom 5/Gen SGCD	Intrón 4/8 (1)	Antisentido

**Tabla 6:** Inserciones de líneas resistentes en PA-1 y H9. <sup>1</sup>Punto de truncación, calculado alineando la inserción respecto a la secuencia del vector pRAM-LRE3. <sup>2</sup>DSI, duplicaciones del sitio de inserción. <sup>3</sup>Crom significa cromosoma. <sup>4</sup> () Isoformas del gen donde se localiza la inserción. <sup>5</sup>Sentido/Antisentido, se refiere a su orientación respecto al sentido transcripcional del gen.

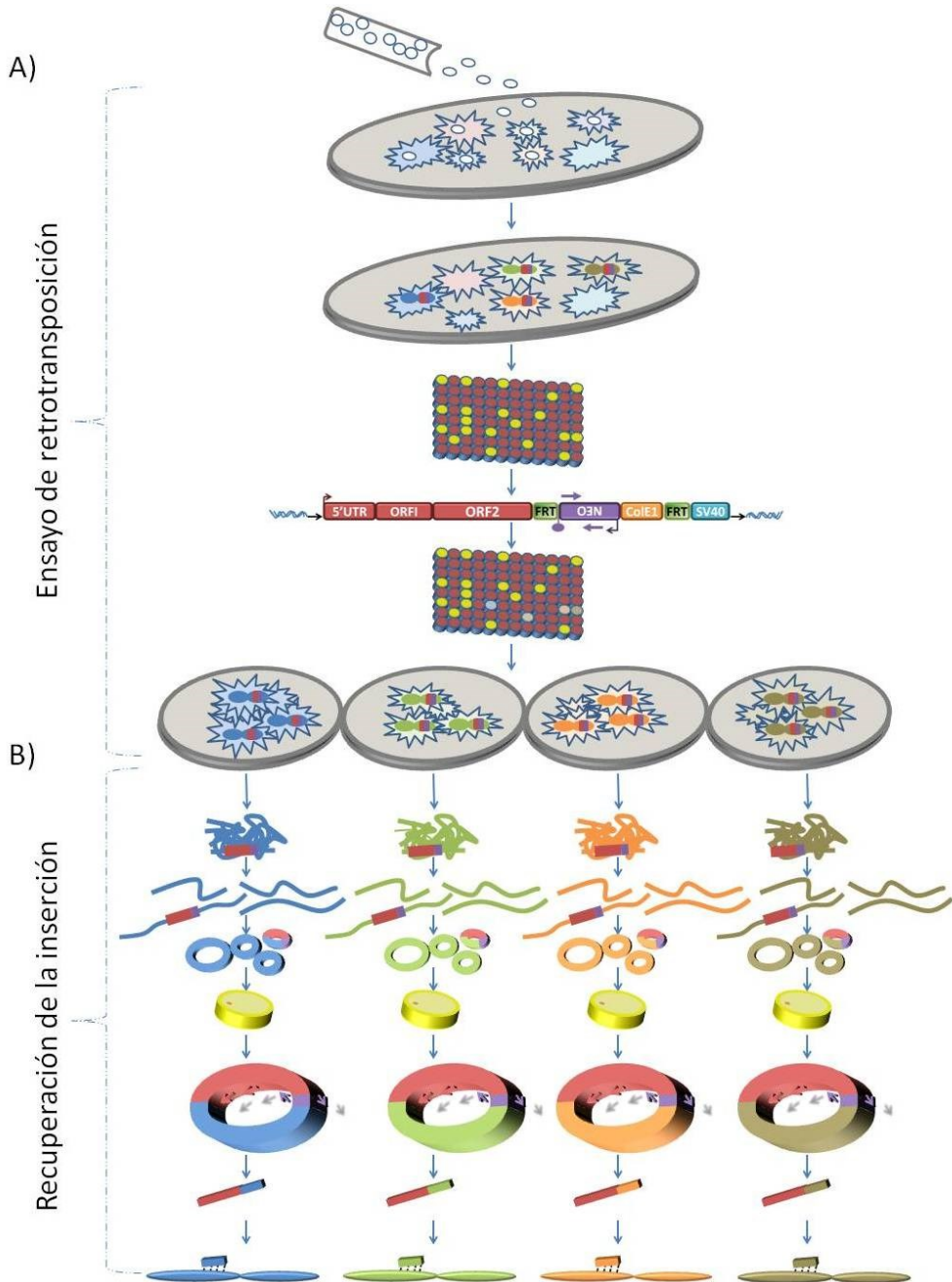
bases aproximadamente en el genoma humano. Seguidamente, una ligación en condiciones diluidas favorece la ligación intramolecular de los fragmentos resultantes generando plásmidos de ADN genómico. Posteriormente, la transformación de la ligación en bacterias dará como resultado colonias resistentes a kanamicina. Esto se debe a que alguno de estos plásmidos de ADN genómico contiene la inserción. Este proceso es posible

debido al casete de rescate ya que el gen de resistencia a neomicina confiere resistencia a kanamicina en bacterias así como el origen de replicación bacteriano (ColE1) permite la replicación del plásmido por dichas bacterias. Por último, los plásmidos resultantes son secuenciados utilizando oligonucleótidos que hibridan en regiones únicas de la inserción. En general, rondas sucesivas de secuenciación son necesarias para alcanzar las regiones genómicas que flanquean la inserción en el extremo 5' y 3' de la misma, obteniendo el sitio de inserción a resolución de nucleótido (Figura 25B).

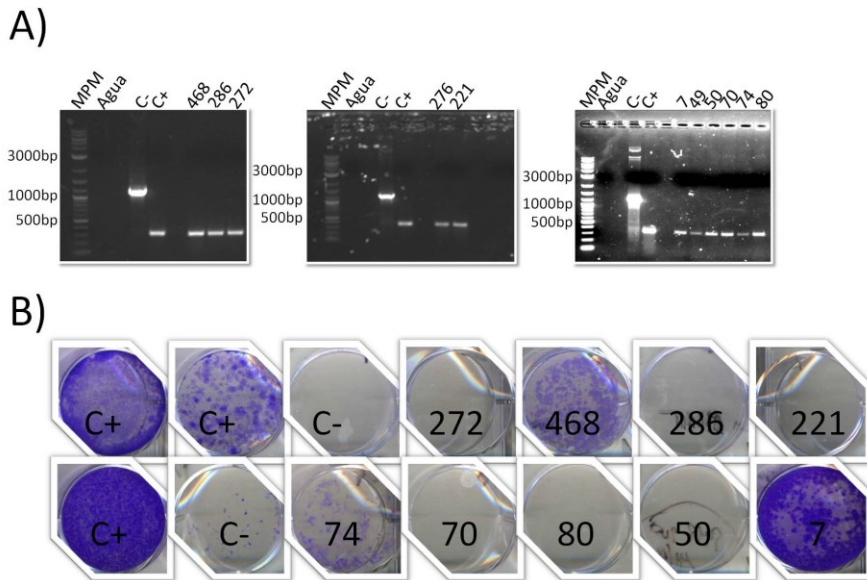
De esta forma, se identificó la localización genómica y la secuencia completa de 14 inserciones que confieren resistencia a neomicina en PA1 y 2 en H9 (Tabla 6).

En segundo lugar, se llevó a cabo la generación de las líneas sensibles. Con el fin de no restringir el análisis a inserciones con una elevada expresión del gen *NEO* debido al proceso de selección, se diseñó un protocolo para obtener líneas clonales con inserciones que no confieren resistencia a neomicina. Por esta razón, los nuevos eventos de retrotransposición fueron identificados por PCR evitando así la selección con neomicina. Esta PCR se basa en utilizar una pareja de oligonucleótidos que se posicionan a ambos lados del intrón que interrumpe el gen de resistencia a neomicina. De esta forma, el tamaño del producto de PCR es diferente antes y después de la retrotransposición. Por tanto, el producto de PCR generado una vez que se haya eliminado el intrón y se haya producido la retrotransposición será indicativo de un nuevo evento de retrotransposición. Esta PCR será referida como PCR de *NEO* (Figura 24C). Este tipo de líneas clonales sólo pudieron ser generadas en células PA-1 debido a la imposibilidad de las células H9 para crecer desde una célula individual debido a su muerte por apoptosis [222].

Brevemente, las células PA-1 fueron transfectadas con el vector pRAM-LRE3. Después de 5 días, se realizó una dilución clonal de las células transfectadas, generándose finalmente 600 líneas clonales. Después, se identificaron los nuevos eventos de retrotransposición por PCR (Figura 26A). La búsqueda por PCR identificó un total de 11 líneas clonales positivas que mostraron el producto de PCR de 332 pares de bases indicativo de un nuevo evento de retrotransposición (Figura 27A). Estas líneas clonales fueron sembradas y crecidas en presencia de neomicina para encontrar las líneas clonales que contenían un evento de retrotransposición que no confería resistencia al antibiótico. De las 11 líneas clonales, 6 eran sensibles a neomicina (PA-1-pRAM221, PA-1-pRAM286, PA-1-pRAM272; PA-1-pRAM50, PA-1-pRAM70, PA-1-pRAM80) mientras que 3 líneas clonales eran resistentes a neomicina (PA-1-pRAM468, PA-1-pRAM 7 y PA-1-pRAM74) (Figura 27B). Este ensayo no pudo ser llevado a cabo en los 2 clones restantes (PA-1-pRAM276 y PA-1-pRAM49). Finalmente, se procedió con la recuperación y caracterización



**Figura 26: Esquema del proceso de obtención de línea clonales sensibles en células PA-1 y caracterización de sus inserciones de novo de LINE-1. A) Ensayo de retrotransposición sin selección con neomicina e identificación de nuevos eventos de retrotransposición por PCR. Los círculos amarillos representan pocillos con crecimiento celular. Los círculos de color azul, verde, naranja y marrón representan clones positivos por PCR. B) Recuperación de la inserción según lo descrito en figura 25B**



**Figura 27: Obtención de líneas clonales sensibles en células PA-1. A)** PCR de NEO que identifica 11 posibles líneas clonales sensibles. Se muestran tres geles de agarosa. En todos ellos, MPM (marcador de peso molecular), C- (plásmido pRAM-LRE) y C+ (línea clonal con inserción de LINE-1). Los números sobre los pocillos indican la línea clonal. **B)** Ensayo de resistencia a neomicina en las líneas clonales identificados por PCR. C+ (línea clonal resistente), C- (células PA-1 wt). Los números indican la línea clonal.

Línea Clonal	<sup>1</sup> Punto Truncación	<sup>2</sup> DSI	Longitud Poli A	Delección	<sup>3</sup> Localización	Intrón/ <sup>4</sup> Exón	<sup>5</sup> Sentido/Antisentido
PA-1-pRAM221	Copia completa	11	56	No	Crom 20/Gen EYA2	Intrón 1 (1,5)	Sentido
PA-1-pRAM286	Copia completa	12	131	No	Crom 1/ Intergénica		
PA-1-pRAM272	Copia completa	17	25	No	Crom 16/Intergénica		

**Tabla 7:** Inserciones de líneas sensibles en PA-1. <sup>1</sup>Punto de truncación: calculado alineando la inserción respecto a la secuencia del vector pRAM-LRE3. <sup>2</sup>DSI: duplicaciones del sitio de inserción. <sup>3</sup>Crom significa cromosoma. <sup>4</sup> () Isoformas del gen donde se encuentra la inserción. <sup>5</sup>Sentido/Antisentido: se refiere a su orientación respecto al sentido transcripcional del gen.

de las inserciones como ha sido previamente descrito para las inserciones de las líneas resistentes (Figura 26B).

En este caso, se consiguió identificar la localización genómica así como las características principales de 3 inserciones pertenecientes a líneas sensibles (Tabla 7). Por otro lado, una de las inserciones restantes (PA-1-pRAM468) se encuentra en medio del proceso de caracterización. Sin embargo, 7 de las 11 inserciones restantes identificadas

por PCR no pudieron ser caracterizadas debido a problemas técnicos propios del proceso de rescate de la inserción.

– *Análisis de las inserciones caracterizadas*

Todas las inserciones generadas en las líneas resistentes y sensibles, presentan las características de una inserción canónica de LINE-1. La mayoría 16/17 presentan duplicaciones del sitio de inserción con una longitud de entre 2-17 nucleótidos lo cual está dentro del rango habitual (2-20) para las duplicaciones del sitio de inserción [67]. Por otro lado, 1/17 inserciones generó una deleción de 635 pares de bases, algo previamente ya descrito [92], [93]. La longitud de las colas de poli A variaron entre 25 y 131 nucleótidos, presentando una longitud superior a las colas de poli A de las inserciones endógenas de LINE-1, algo también ya descrito anteriormente [92]. A pesar del bajo número de inserciones, se cumplen las proporciones para determinados fenómenos asociados a la retrotransposición. Primero, la mayoría (14/19) de las inserciones se encuentran truncadas en el extremo 5' mientras que 5/19 eran de copia completa [93]. Segundo, las inserciones se integraron a lo largo de todo el genoma en distintos cromosomas [92], [93]. Por tanto, las inserciones obtenidas muestran las características típicas de una inserción de LINE-1 endógena.

4.2.2. Regulación epigenética de las inserciones artificiales de novo de LINE-1

Para determinar las modificaciones de histonas presentes sobre las inserciones artificiales *de novo* de LINE-1 se llevaron a cabo distintos experimentos de inmunoprecipitación de cromatina utilizando anticuerpos que reconocen modificaciones de histonas asociadas a cromatina transcripcionalmente activa tales como H3K4me3, H4ac (acetilación de la histona 4 que engloba la acetilación en cuatro residuos de lisina (5,8,12,16)), y H3K9ac así como anticuerpos que reconocen modificaciones de histonas asociadas a cromatina transcripcionalmente inactiva como H3K9me3 y H3K27me3. El enriquecimiento de cada modificación de histona sobre la inserción de LINE-1 fue determinado mediante qPCR sobre el ADN inmunoprecipitado. Por otro lado, la única secuencia presente en las inserciones artificiales *de novo* y ausente en el resto de inserciones endógenas de LINE-1 es el gen de resistencia a neomicina. Por tanto, se utilizaron oligonucleótidos que hibridan sobre el gen de resistencia a neomicina ya que éste constituye una secuencia única que puede ser utilizada como referencia evitando así amplificaciones inespecíficas desde los LINE-1 endógenos. Los niveles de enriquecimiento obtenidos sobre la inserción de LINE-1 fueron comparados con los valores de referencia obtenidos para zonas del genoma muy enriquecidas o poco enriquecidas para cada

modificación en cuestión utilizando datos de [www.roadmapepigenomic.org](http://www.roadmapepigenomic.org) (Ver Anexo III) [165]. Por otro lado, la significancia estadística del enriquecimiento obtenido respecto al control negativo de enriquecimiento se contrastó estadísticamente mediante un T de student siempre que se dispuso del número adecuado de réplicas biológicas. Finalmente, el estudio se realizó en 5 de las 14 líneas clonales resistentes de PA-1 las cuales fueron elegidas en función del punto de truncación de las inserciones cubriendo todas las posibilidades desde una inserción casi de copia completa hasta una inserción extremadamente truncada en 5'. Además se incluyeron las 2 líneas clonales resistentes de H9 ya que son las únicas líneas generadas en células madre embrionarias humanas. Por último, las 3 líneas clonales sensibles de PA-1 fueron todas incluidas como únicas representantes de esta categoría. En las figuras donde se muestran estos resultados, las líneas resistentes se encuentran resaltadas en morado mientras que las líneas sensibles se encuentran resaltadas en gris.

En esta sección, en el Anexo IV, se ha incluido un esquema con las características principales así como su localización respecto a los demás elementos cercanos del genoma para cada una de las inserciones. Además, se ha incluido información acerca del estado del promotor del gen donde las inserciones génicas se encuentran integradas. Asimismo, también se incorpora información del sitio de inserción respecto a las modificaciones de histonas allí presentes. Estos datos fueron obtenidos de [www.roadmapepigenomic.org](http://www.roadmapepigenomic.org). Debido a la ausencia de estudios de ChIP-seq en células PA-1, se han utilizado como referencia células H9. Además, también se han consultado datos procedentes del proyecto ENCODE para células NT2-D1, las cuáles son semejantes a células PA-1 ya que son células de carcinoma humano embrionario.

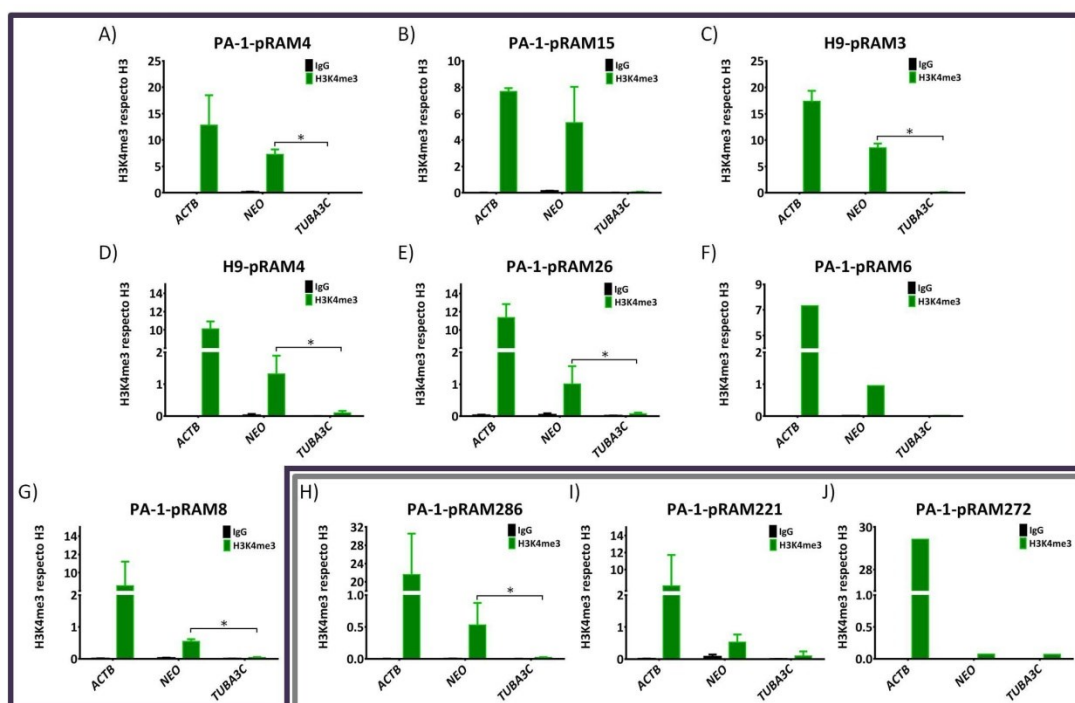
#### 4.2.2.1. *Modificaciones de histonas asociadas con actividad transcripcional*

##### – *Trimetilación de la lisina 4 de la histona 3 (H3K4me3)*

H3K4me3 es una modificación generalmente localizada sobre los promotores, cerca del lugar de inicio de la transcripción. Su presencia correlaciona positivamente con la actividad transcripcional [137], [177], [178]. Por otro lado, elementos LINE-1 transcripcionalmente activos en el genoma de células de cáncer de mama presentan dicha modificación inmediatamente aguas arriba de su promotor [71]. Además, elementos LINE-1 reactivados transcripcionalmente del genoma de ratón presentan esta modificación en su promotor [198]. Por ello, se determinó el enriquecimiento de H3K4me3 sobre el gen *NEO* en todas las líneas clonales incluyendo resistentes y sensibles mediante ChIP-qPCR.

Cómo se observa en la figura 28, 7/7 inserciones pertenecientes a líneas resistentes se encuentran enriquecidas en H3K4me3 en distintos grados (Figura 28A-G). 5 de estas 7 inserciones se encuentran significativamente enriquecidas en H3K4me3 (Figura 28A, C, D, E, G) mientras que las 2 restantes muestran la misma tendencia (Figura 28B, F). De hecho, las inserciones que se encuentran más enriquecidas en H3K4me3 pertenecen a líneas resistentes (Figura 28A-C). Por otro lado, tan solo 1 de 3 inserciones pertenecientes a líneas sensibles está enriquecida en H3K4me3 (Figura 28H), siendo este enriquecimiento marginal. Las inserciones de líneas sensibles restantes no muestran enriquecimiento de H3K4me3 sobre sus inserciones (Figura 28I, J).

Resumiendo, H3K4me3 está mayoritariamente enriquecida y en mayor grado sobre inserciones de líneas resistentes, consistente con su selección con neomicina.



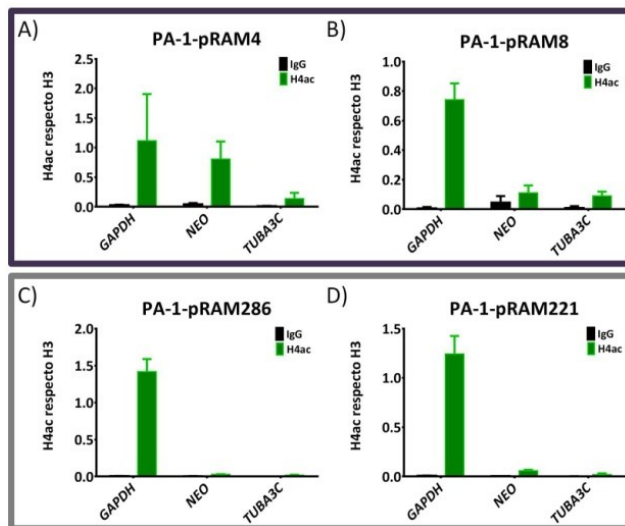
**Figura 28: Enriquecimiento de H3K4me3 sobre las inserciones artificiales de novo de LINE-1. A-G)** Líneas clonales resistentes resaltadas en morado. **H-J)** Líneas clonales sensibles resaltadas en gris. **A), C), D), E), G), H), I)** Se muestra la media de tres ChIPs-qPCR independientes con su correspondiente desviación estándar. Los p-valoros obtenidos fueron los siguientes: para A) p-valor=0,00018165, para C) p-valor=5,44197E-05, para D) p-valor=0,021353461, para E) p-valor=0,044941295, para G) p-valor=0,00025613 y para H) p-valor=0,049116592. **B)** Se muestra la media de dos ChIPs-qPCR independientes. Las barras de error muestran el rango de valores. **F), J)** Se muestra un experimento de ChIP-qPCR. **A-J)** ACTB es un control positivo y TUBA3C es un control negativo.

– Acetilación de la histona 4 (H4ac)

H4ac se refiere a un conjunto de acetilaciones que ocurren en una serie de residuos de lisina de la histona 4. De hecho, las acetilaciones reconocidas por el anticuerpo que se usa en el presente estudio son las siguientes: H4K5ac, H4K8ac, H4K12ac y H4K16ac. Estas acetilaciones se encuentran fundamentalmente sobre promotores y su presencia correlaciona con actividad transcripcional [140], [182]. Por otro lado, es una modificación que aparece enriquecida sobre las inserciones artificiales *de novo* de LINE-1 que son reactivadas después del tratamiento con TSA en células PA-1 [60]. Por tanto, H4ac puede ser importante en la regulación de las inserciones de LINE-1.

El enriquecimiento de H4ac fue determinado en 2 de las 5 líneas clonales resistentes de PA-1 así como en 2 de las 3 líneas clonales sensibles de PA-1. Sin embargo, tan solo una inserción de una línea resistente (PA-1-pRAM4) mostraba un enriquecimiento que se podía considerar biológicamente relevante (Figura 29A). Además, esta inserción muestra un enriquecimiento elevado de H3K4me3 (Figura 28A). Por otro lado, el resto de inserciones 3/4 mostraron un enriquecimiento de H4ac similar al control negativo (Figura 29B-D).

Por tanto, a pesar del bajo número de inserciones analizadas, existe una tendencia hacia la ausencia de acetilación de la histona 4 sobre las inserciones *de novo* LINE-1. De hecho, la presencia de H4ac parece estar ligada a un alto enriquecimiento de H3K4me3.



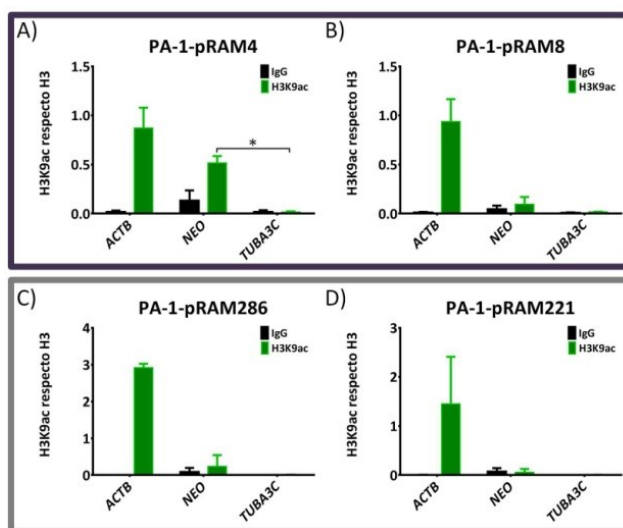
**Figura 29: Enriquecimiento de H4ac sobre las inserciones artificiales de novo de LINE-1. A), B)** Líneas clonales resistentes resaltadas en morado. **C), D)** Líneas clonales sensibles resaltadas en gris. **A-D)** Se muestra la media de dos ChIPs-qPCR. Las barras de error muestran el rango de valores. GAPDH es un control positivo y TUBA3C es un control negativo.



– Acetilación de la lisina 9 de la histona 3 (H3K9ac)

H3K9ac es una modificación asociada con actividad transcripcional, es decir, se encuentra en promotores que son activos. De hecho, su presencia está relacionada con una mayor expresión génica [136], [151].

En el presente estudio, se determinó el enriquecimiento de esta modificación mediante ChIP-qPCR sobre la inserción de 2 de las 5 líneas resistentes y de 2 de las 3 inserciones sensibles en células PA-1. Solo una inserción se encuentra significativamente enriquecida en H3K9ac (Figura 30A). Dicha inserción perteneciente a una línea clonal resistente, es la misma que presenta un elevado enriquecimiento de H4ac y H3K4me3 (Figura 29A, 28A). Sin embargo, el resto de las inserciones presentan valores de enriquecimiento similares al control negativo (Figura 30B-D).



**Figura 30: Enriquecimiento de H3K9ac sobre las inserciones artificiales de novo de LINE-1. A), B)** Líneas clonales resistentes resaltadas en morado. **C), D)** Línea clonales sensibles resaltadas en gris. **A), B), D)** Se muestra la media de tres ChIPs-qPCR con su correspondiente desviación estándar. El p-valor obtenido en A) fue 0,000244177. **C)** Se muestra la media de dos ChIPs-qPCR independientes. Las barras de error muestran el rango de valores. **A-D)** ACTB es un control positivo mientras que TUBA3C es un control negativo.

A pesar también del bajo número de inserciones estudiadas, los resultados de H3K9ac también sugieren que existe una tendencia hacia la ausencia de acetilación de la histona H3 sobre las inserciones *de novo* de LINE-1. Al igual que ocurría con H4ac, la presencia de H3K9ac parece estar ligada a un alto enriquecimiento de H3K4me3.

#### 4.2.2.2. Modificaciones de histonas asociadas con represión transcripcional

##### – Trimetilación de la lisina 9 de la histona 3 (H3K9me3)

La trimetilación de la lisina 9 de la histona 3 es una marca asociada con represión génica y con la formación de la heterocromatina [170], [183]–[185]. Además está ligada a la represión de los elementos LINE-1 existentes en el genoma de ratón [117]. Por tanto, es una candidata ideal para participar en la regulación de las inserciones artificiales *de novo* de LINE-1. La presencia de H3K9me3 fue determinada mediante ChIP-qPCR en todas las líneas clonales generadas tanto en PA-1 como en H9.

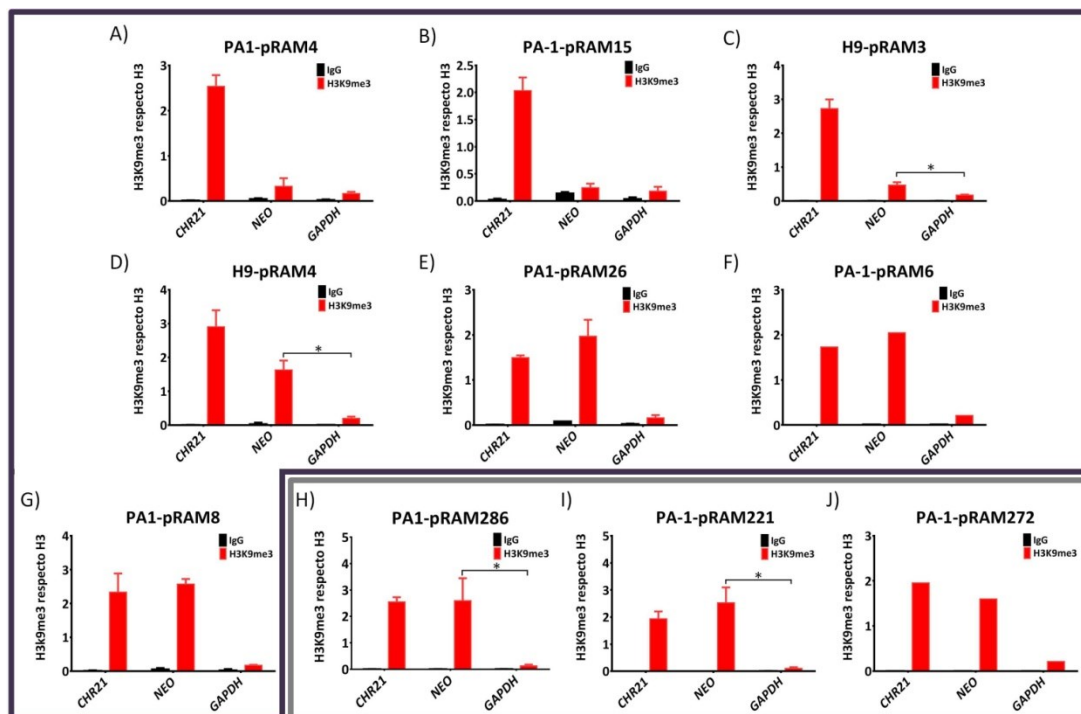
Como se observa en la Figura 31, 5 de las 7 inserciones de líneas resistentes se encuentran enriquecidas en H3K9me3 en distintos grados (Figura 31C-G). 2 de estas 5 inserciones de líneas resistentes se encuentran significativamente enriquecidas en H3K9me3 (Figura 31C, D) mientras que las 3 restantes muestran la misma tendencia incluso con valores de enriquecimiento mayores (Figura 31E-G). En cambio, 2/7 inserciones de líneas resistentes muestran un enriquecimiento de H3K9me3 similar al control negativo (Figura 31A, B). Por otro lado, todas las inserciones de líneas sensibles se encuentran enriquecidas en H3K9me3 (Figura 31H-J). 2 de 3 inserciones de líneas sensibles están significativamente enriquecidas (Figura 31H, I) mientras que la inserción de la línea sensible restante muestra la misma tendencia con un enriquecimiento de H3K9me3 semejante (Figura 31J). Las inserciones que presentan un enriquecimiento elevado de H3K9me3 similar al control positivo pertenecen tanto a líneas resistentes (Figura 31E-G) como a líneas sensibles (Figura 31H-J).

Curiosamente, las 5 inserciones de líneas resistentes enriquecidas en H3K9me3 (Figura 31C-G) se encuentran simultáneamente enriquecidas en H3K4me3 (Figura 28C-G). Además, las inserciones no enriquecidas en H3K9me3 (Figura 31A, B) se corresponde con dos de las inserciones más enriquecidas en H3K4me3 (Figura 28A, B).

En resumen, la inserciones *de novo* de LINE-1 tanto de líneas sensibles como de líneas resistentes se encuentran enriquecidas en H3K9me3 en un grado similar. Además, las inserciones de las líneas resistentes enriquecidas en H3K9me3 presentan al mismo tiempo H3K4me3.

##### – Trimetilación de lisina 27 de la histonas 3 (H3K27me3)

La trimetilación de la lisina 27 es una modificación asociada a represión génica. Localizada en gran medida sobre promotores bivalentes junto con H3K4me3 donde



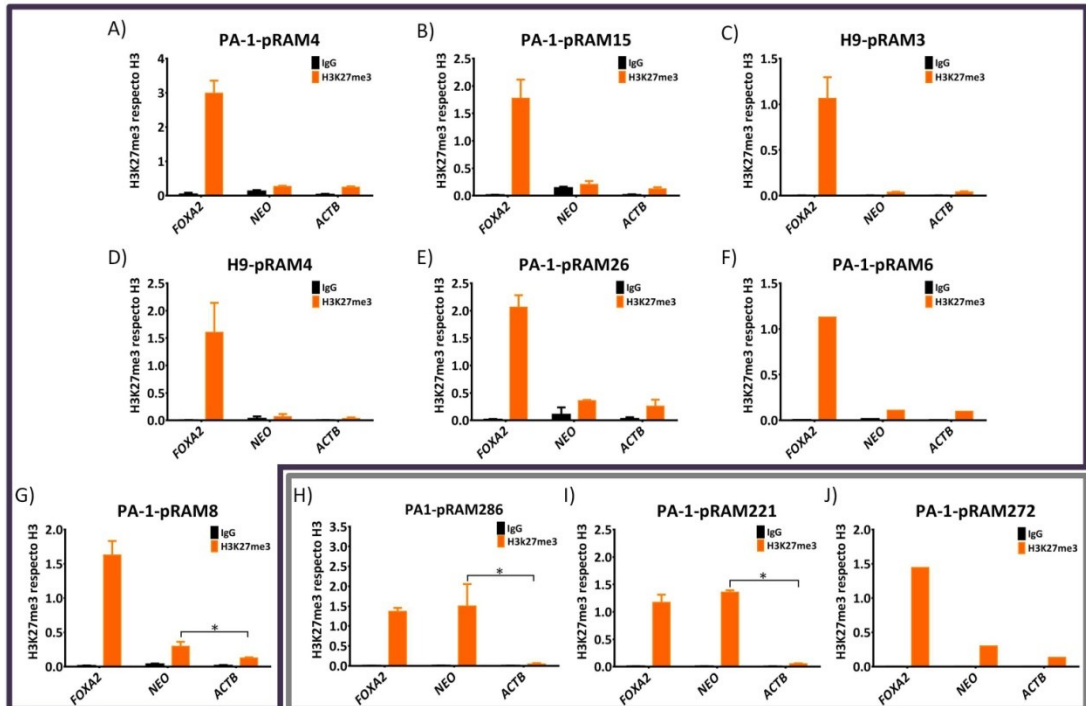
**Figura 31: Enriquecimiento de H3K9me3 sobre las inserciones artificiales de novo de LINE-1. A-G)** Líneas clonales resistentes resaltadas en morado. **H-J)** Líneas clonales sensibles resaltadas en gris. **C), D), H), I)** Se muestra la media de tres ChIPs-qPCR con su correspondiente desviación estándar. Los p-valores obtenidos fueron los siguientes: para C)  $p\text{-valor}=0,027555981$ , para D)  $p\text{-valor}=0,000401467$ , para H)  $p\text{-valor}=0,007261971$ , para I)  $p\text{-valor}=0,001632241$ . **A), B), E), G)** Se muestra la media de dos ChIPs-qPCR. Las barras de error muestran el rango de valores. **F), J)** Se muestra un experimento de ChIP-qPCR. **A-J)** CHR21 es un control positivo y GAPDH es un control negativo.

correlaciona inversamente con la actividad transcripcional [134], [159], [160]. Por otro lado, dominios en el genoma humano y murino enriquecidos en esta modificación contienen elementos LINE-1 [154], [164]. Por tanto, se trata también de una buena candidata para el presente estudio. La presencia de dicha modificación sobre la nueva inserción de LINE-1 fue determinada mediante ChIP-qPCR en todas las líneas clonales tanto resistentes como sensibles.

Como se observa en la Figura 32, tan solo 1/7 inserciones pertenecientes a líneas resistentes se encuentran significativamente enriquecidas en H3K27me3 (Figura 32G). Por el contrario, el resto de inserciones de líneas resistentes presentan valores de enriquecimiento similares al control negativo (Figura 32A-F). Por otro lado, 2/3 inserciones de las líneas sensibles están significativamente enriquecidas en H3K27me3 (Figura 32H, I). Además, la tercera inserción procedente de una línea sensible parece estar modestamente

enriquecida en H3K27me3 (Figura 32J). Sólo las dos inserciones pertenecientes a líneas sensibles muestran valores de enriquecimiento elevados (Figura 32H, I).

Sorprendentemente, las inserciones procedentes de líneas sensibles que presentan un enriquecimiento de H3K27me3 similar al control positivo (Figura 32H, I), además se encuentran altamente enriquecidas en H3K9me3 (Figura 31H, I).



**Figura 32: Enriquecimiento de H3K27me3 sobre las inserciones artificiales de novo de LINE-1.** A-G) Inserciones de líneas clonales resistentes resaltadas en morado. H-J) Inserciones de líneas clonales sensibles resaltadas en gris. A), C), D), E), G), H), I) Se muestra la media de tres ChIPs-qPCR con su correspondiente desviación estándar. B) Se muestra la media de dos ChIPs-qPCR. Las barras de error indican el rango de valores. F), J) Se muestra un solo experimento de ChIP-qPCR. G-I) Los p-valor obtenidos fueron los siguientes: en G) el p-valor obtenido fue 0,017, en H) el p-valor obtenido fue 0,010242328 y en I) el p-valor obtenido fue 4,38969E-07. A-J) FOXA2 es un control positivo y ACTB es un control negativo.

En resumen, H3K27me3 se encuentra enriquecida fundamentalmente y en mayor grado sobre las inserciones artificiales *de novo* de LINE-1 de líneas sensibles. Además, un elevado enriquecimiento de H3K27me3 coincide con un elevado enriquecimiento de H3K9me3.

Línea Clonal/Inserción	Marcas Activas			Marcas represivas	
	H3K4me3	H4ac	H3K9ac	H3K9me3	H3K27me3
PA-1-pRAM4	x	x	x		
PA-1-pRAM15	x				
H9-pRAM3	x			x	
H9-pRAM4	x			x	
PA-1-pRAM26	x			x	
PA-1-pRAM6	x			x	
PA-1-pRAM8	x			x	x
PA-1-pRAM286	x			x	x
PA-1-pRAM221				x	x
PA-1-pRAM272				x	x

**Tabla 8:** Resumen de las modificaciones de histonas enriquecidas sobre las inserciones artificiales de novo de LINE-1. En morado se resaltan las inserciones de líneas resistentes mientras que en gris se resaltan las inserciones de líneas sensibles.

En resumen, como se puede observar en la tabla 8, el análisis realizado da lugar a que las inserciones generadas se puedan agrupar en 3 categorías: una primera categoría de inserciones que están enriquecidas únicamente en modificaciones de actividad (H3K4me3, H3K9ac y H4ac), una segunda categoría de inserciones las cuales están enriquecidas en modificaciones activas y represivas (H3K4me3 y H3K9me3) y una última categoría donde las inserciones se encuentran enriquecidas fundamentalmente en modificaciones represivas (H3K9me3 y H3K27me3).

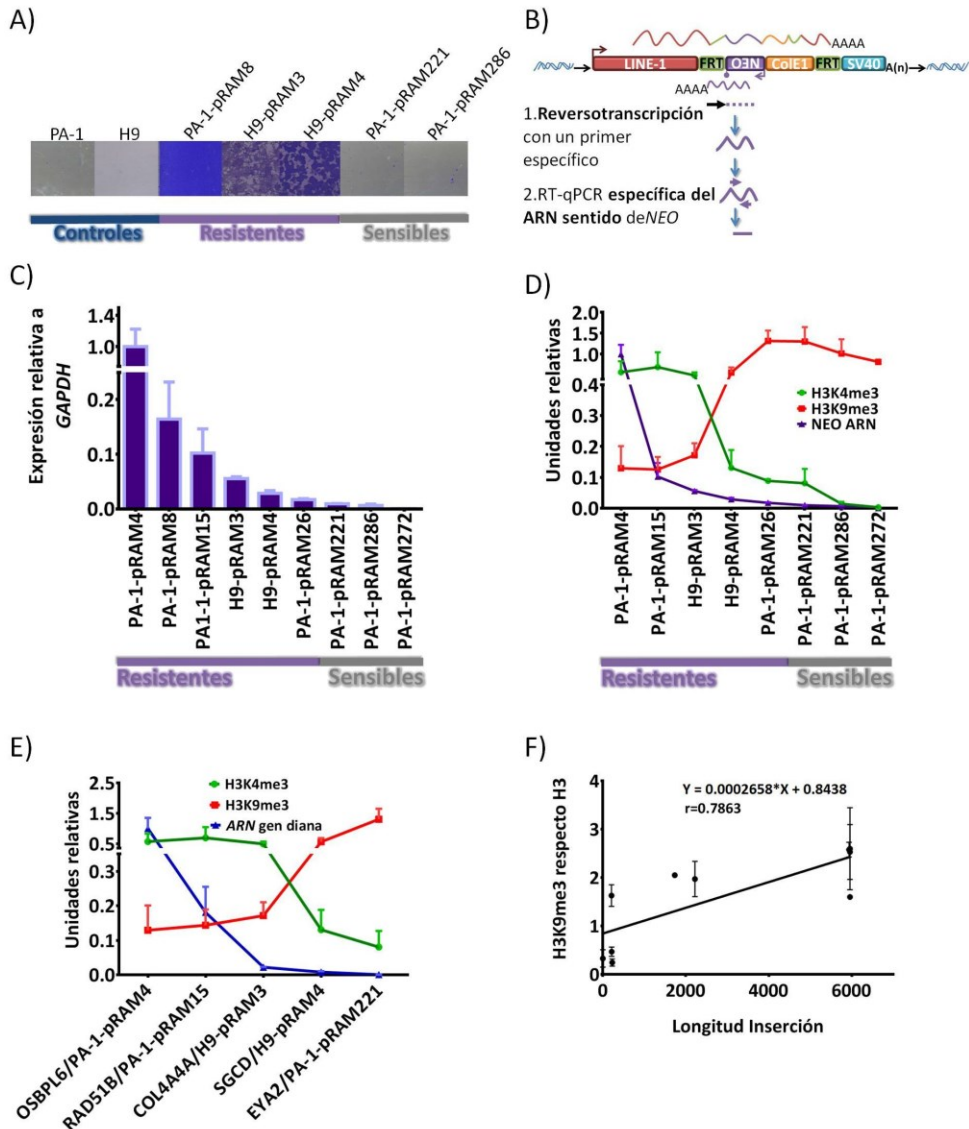
Finalmente, el enriquecimiento de H3K9me3 sobre inserciones provenientes de líneas resistentes y líneas sensibles indica que probablemente existe una tendencia hacia la represión de las nuevas inserciones de LINE-1. Sin embargo, en un determinado contexto genómico/transcripcional favorable estas inserciones pueden ganar H3K4me3 permitiendo que las líneas sean resistentes mientras que en otro contexto genómico/transcripcional más represivo se produce la ganancia de H3K27me3 dando lugar a líneas sensibles.

#### 4.2.3. Relación entre la expresión y el estado de la cromatina en inserciones de novo de LINE-1

El sistema de generación de las líneas clonales da lugar a una clasificación binaria de líneas clonales resistentes y líneas clonales sensibles en función de si son capaces de crecer en presencia de neomicina (Figura 33A). Sin embargo, según las modificaciones de histonas presentes sobre las inserciones en estudio, cabe la posibilidad de que esta clasificación no sea tan rígida. Por ello, se analizaron los niveles de transcripción del gen reportero *NEO* mediante RT-qPCR. Para amplificar específicamente el ARN producido desde el propio promotor del gen *NEO* se llevó a cabo una reacción de reverso transcripción con un oligonucleótido específico (Figura 33B). Como se podía esperar, las líneas resistentes presentan una mayor expresión del gen *NEO* que las líneas sensibles. Por otro lado, la expresión del gen *NEO* varía dentro de la categoría de líneas resistentes así como dentro de la categoría de líneas sensibles. Sin embargo, la variabilidad es mayor dentro de la categoría de resistentes. (Figura 33C). Además, algunas de las líneas resistentes presentan un nivel de expresión del gen *NEO* similar al que muestran las líneas sensibles (Figura 33C).

A la vista de los resultados de expresión del gen *NEO*, cabe preguntarse si existe una relación entre la expresión del mismo y los enriquecimientos de H3K4me3 y H3K9me3 existentes sobre las inserciones de LINE-1. Para responder a esta pregunta, se comparó entre las distintas líneas la relación entre la expresión del gen *NEO* junto con los enriquecimientos relativos de H3K4me3 y H3K9me3. De hecho, en la figura 33D se observa que líneas clonales, con un mayor enriquecimiento relativo de H3K4me3 sobre la inserción de LINE-1, son las que presentan una mayor expresión del gen *NEO* mientras que líneas clonales, cuyas inserciones muestran un mayor enriquecimiento relativo de H3K9me3, son las que presentan una menor expresión del gen *NEO*. Sorprendentemente, algunas de las inserciones de líneas clonales clasificadas como resistentes son más parecidas epigenéticamente a las inserciones de las líneas sensibles. De hecho, en la parte media del gráfico se observa el tránsito entre las líneas resistentes y las líneas sensibles a nivel epigenético. Estos datos sugieren que el balance H3K4me3/H3K9me3 sobre la inserción *de novo* de LINE-1 es importante para el control de la expresión del gen *NEO*. Además, líneas clonales, cuyo balance  $H3K4me3/H3K9me3 < 1$ , siempre presentan una menor expresión del gen *NEO*. Por tanto, como cabría esperar, el depósito H3K9me3 parece estar asociado con la represión transcripcional de las inserciones *de novo* de LINE-1.

Por otra parte, el enriquecimiento de H3K4me3 o H3K9me3 sobre una inserción podría depender del contexto transcripcional en el que se integre. Para analizar esta posible relación se utilizaron 6 de las 7 inserciones que estaban localizadas dentro de



**Figura 33: Relación entre la expresión y el estado de la cromatina en las inserciones de novo de LINE-1. A)** Ensayo de resistencia a neomicina. PA-1 y H9 son los controles negativos de resistencia. Se muestran 5 de las 10 líneas clonales estudiadas, 3 resistentes y 2 sensibles. **B)** Esquema que muestra la amplificación específica del ARN sentido del gen NEO para su cuantificación por RT-qPCR. **C)** RT-qPCR que muestra la expresión del gen NEO en las líneas resistentes y en las líneas sensibles. Se muestra la media de dos réplicas biológicas independientes. Las barras de error muestran el rango de valores. Posteriormente, los niveles de expresión obtenidos fueron normalizados respecto al nivel de expresión de NEO en la línea PA-1-pRAM4. **D)** Representación de la expresión del gen NEO junto con el enriquecimiento relativo de H3K4me3 y H3K9me3 para cada una de las inserciones. **E)** Gráfico

de tendencias que muestra la relación entre la expresión de los genes diana en la línea PA-1wt y H9wt antes de la integración y los enriquecimientos relativos de H3K4me3/H3K9me3 existentes para cada inserción de novo integrada en cada gen. Se muestra la expresión relativa de los genes respecto a GAPDH. Se normaliza respecto a la expresión de OSBPL6 que toma valor 1. Se muestra la media de 2 RT-qPCRs. Las barras de error muestran el rango de valores. **F)** Regresión lineal entre la longitud de la inserción y el enriquecimiento de H3K9me3.

genes (ver Anexo IV). Para todas ellas, se determinó, mediante RT-qPCR, el nivel de expresión de los genes diana en la línea parental (PA-1 o H9) para evitar posibles efectos de la inserción sobre la expresión génica. Se utilizaron oligonucleótidos que, como mínimo, amplificaban la isoforma donde se encontraba la inserción integrada. En la figura 33E, se observa que esta relación se cumple para 5 de las 6 inserciones analizadas. De hecho, inserciones que presentan un balance H3K4me3/H3K9me3 > 1, se encuentran localizadas en genes que presentan una mayor expresión mientras que inserciones de LINE-1 donde H3K4me3/H3K9me3 < 1 son las que se encuentran localizadas en genes menos expresados. Estos datos sugieren que el contexto transcripcional puede influir el reclutamiento de H3K4me3 y H3K9me3 sobre cada inserción y, por tanto, su regulación epigenética.

Por último, según Bulut-Karsioglu et al., [117] la longitud de las inserciones de LINE-1 endógenas en el genoma de ratón determina el enriquecimiento de H3K9me3. Por tanto, características propias de la inserción son también importantes para el reclutamiento de modificaciones de histonas. Teniendo en cuenta estos datos, se comprobó dicha relación sobre las inserciones artificiales *de novo* de LINE-1. Para ello, se analizó la relación lineal entre la longitud de la inserción y el nivel de enriquecimiento de H3K9me3 mediante una correlación de Pearson. Como se muestra en la figura 33F, se obtuvo una correlación directa entre dichas variables con un coeficiente de correlación de Pearson  $r=0,7863$  y un p-valor de 0.0070 siendo la relación lineal entre dichas variables estadísticamente significativa. Por tanto, una mayor longitud de la inserción de LINE-1 conlleva un mayor reclutamiento de H3K9me3. Este análisis sugiere que cuanto más larga es la inserción en 5', más posibilidades existen de que la maquinaria enzimática que deposita H3K9me3 reconozca la inserción *de novo* de LINE-1.

Resumiendo, la regulación epigenética de las inserciones de novo de LINE-1 podría estar influenciada por el contexto transcripcional así como por secuencias propias de la inserción localizadas en el extremo 5' de la misma.

#### 4.2.4. Impacto epigenético de las inserciones de LINE-1 sobre el genoma adyacente

Una vez determinadas las modificaciones de histonas enriquecidas sobre las inserciones *de novo* de LINE-1, cabe preguntarse si las inserciones son capaces



de inducir cambios en el enriquecimiento de dichas modificaciones en el genoma adyacente. Para ello, mediante ChIP-qPCR se determinó el enriquecimiento de las modificaciones de histonas estudiadas en el apartado 4.2.2. en el genoma adyacente a las inserciones de LINE-1 de las líneas clonales incluidas en este estudio. Estos experimentos fueron llevados a cabo únicamente en aquellas líneas cuyas inserciones estaban enriquecidas en la modificación en cuestión. Para analizar los posibles cambios de enriquecimiento en el genoma adyacente así como la distancia alcanzada por estos posibles cambios, se diseñaron parejas de oligonucleótidos aguas arriba y aguas abajo de la inserción. Estas parejas de oligonucleótidos diseñadas cada 500 pares de bases cubrían una distancia aproximada de 4-5kb aguas arriba y aguas abajo de la inserción. Además, para el diseño de estas parejas de oligonucleótidos se evitaron las secuencias repetidas presentes en el genoma humano. Por otro lado, los valores de enriquecimiento de las distintas modificaciones en las distintas posiciones genómicas fueron normalizadas respecto a la cantidad total de H3. Estos enriquecimientos fueron comparados a aquellos obtenidos en las mismas posiciones en las líneas parentales PA-1 y H9. Para ser considerados como válidos, los cambios de enriquecimiento generados debían reunir dos condiciones: i) por un lado, presentar un valor superior al control negativo y ii), por otro lado, producirse al menos 500 pares de bases desde el último nucleótido aguas arriba o aguas abajo de la inserción, ya éste es el límite de resolución de la técnica utilizada. IgG fue utilizada como control negativo pero se ha excluido de las gráficas para facilitar su comprensión ya que los valores de ésta fueron siempre menores que el control biológico negativo.

En las figuras donde se muestran los cambios en el genoma adyacente, se incluye un esquema de localización de la inserción en el cual se muestra siempre la inserción de LINE-1 en sentido 5'-3'. En rojo se muestra la inserción de LINE-1, mientras en morado aparece el gen *NEO* que para simplificar sustituye a todo el casete de rescate. Los números situados encima de las líneas discontinuas (rojo, morado, negro) representan el último nucleótido de la inserción tanto en 5' como en 3' y el punto 0. El punto 0 corresponde a la secuencia del gen *NEO* amplificada en los experimentos de ChIP-qPCR. Las distancias se han calculado respecto al punto 0, siendo la posición -1 y +1 el primer nucleótido situado inmediatamente aguas arriba y aguas abajo del amplicón *NEO* respectivamente. Las regiones situadas entre las líneas negras discontinuas resaltadas en el color correspondiente delimitan la región genómica analizada en el estudio. De hecho, los números situados debajo de las líneas indican la primera y última posición medida en el genoma aguas arriba y aguas abajo.

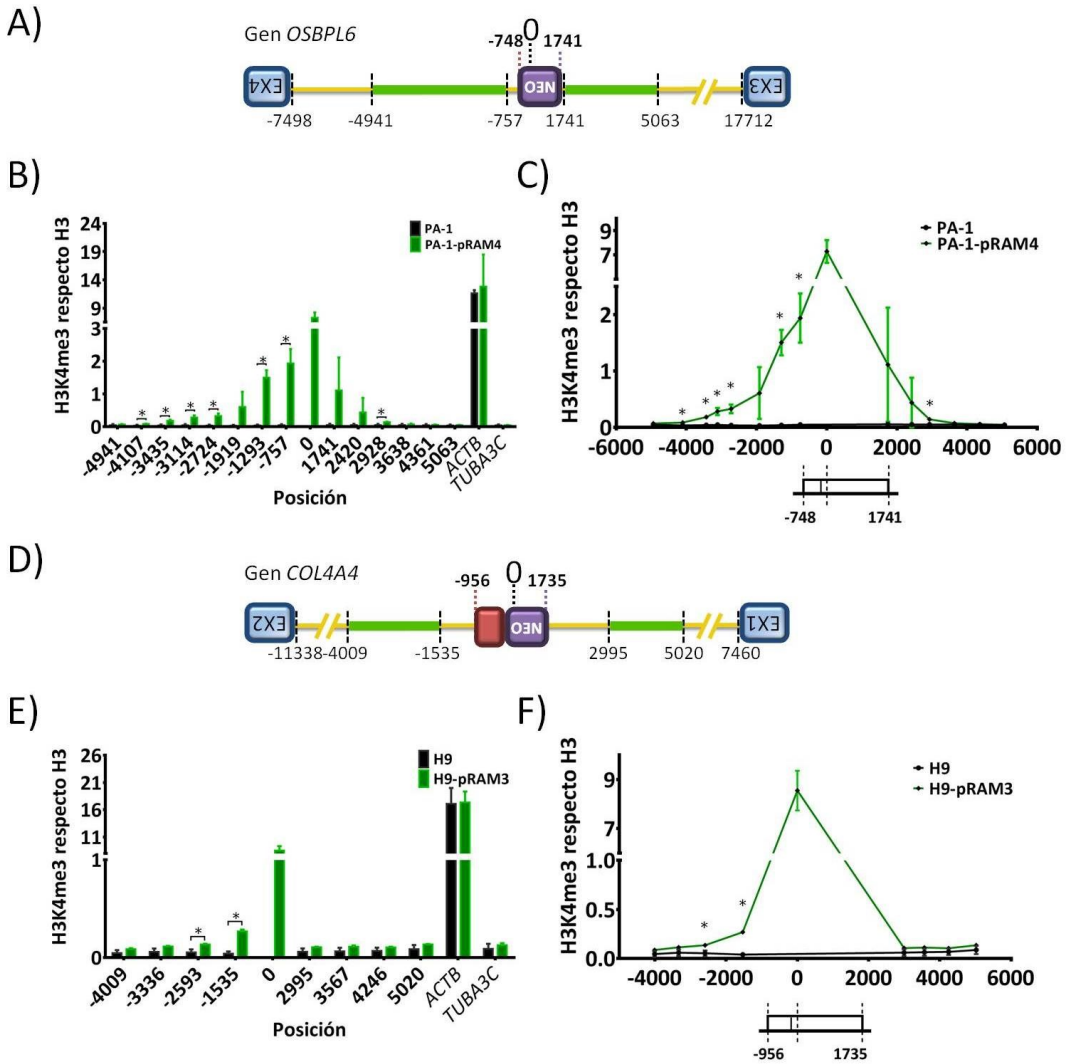
#### 4.2.4.1. Análisis de H3K4me3

En este apartado se analizarán los cambios de H3K4me3 en el genoma adyacente de 7 líneas resistentes y de 1 línea sensible cuyas inserciones estaban enriquecidas en H3K4me3 en algún grado. En primer lugar, se muestran los resultados para cuatro líneas cuyas inserciones indujeron cambios de enriquecimiento de H3K4me3 en el genoma adyacente.

A continuación, se muestra el análisis de los cambios en la línea resistente PA-1-pRAM4 cuya inserción presentaba un valor intermedio/alto de enriquecimiento de H3K4me3 (Figura 28A). En esta línea se detectan cambios significativos en el estatus de H3K4me3 tanto aguas arriba como aguas abajo de la inserción. Dichos cambios significativos alcanzan una distancia de 3359 pares de bases aguas arriba desde el punto -748 hasta el punto -4107 (Figura 34B, C). Por otro lado, los cambios significativos aguas abajo de la inserción, se restringen a una posición, concretamente al punto 2928, lo cual significa que la inserción ha inducido un cambio en el estatus de H3K4me3 a 1187 pares de bases desde el último nucleótido en el extremo 3' de la inserción (Figura 34B, C). Además los valores de enriquecimiento en estas posiciones se encuentran siempre dentro del rango delimitado por los controles biológicos y por encima del valor medio del control negativo (*TUBA3C*). Debido a la elevada variabilidad de las tres réplicas analizadas, no existen diferencias significativas en las posiciones 1741 y 2420 aguas abajo de la inserción. Sin embargo, individualmente, en cada experimento de ChIP-qPCR, estas posiciones muestran un enriquecimiento mayor en la línea clonal comparada con la línea parental. Es importante mencionar, que si se tiene en cuenta solo los datos de enriquecimiento en la región analizada de la línea control PA-1 wt se observa que es una zona pobre en H3K4me3 (Figura 34B, barras negras).

Seguidamente, se muestran los resultados para la línea resistente H9-pRAM3 cuya inserción también presenta un enriquecimiento intermedio de H3K4me3 (Figura 28C). En la línea H9-pRAM3 se observan cambios significativos en el enriquecimiento de H3K4me3 en el genoma próximo aguas arriba de la inserción. Estos cambios alcanzan una distancia de unas 1637 pares de bases desde el punto -956 hasta el punto -2593 (Figura 34E, F). Además se observa, como la región analizada en la línea parental no presenta enriquecimientos relevantes respecto al control negativo lo que significa que es una región pobre en H3K4me3 (Figura 34E, barras negras). Cabe destacar que aguas abajo de la inserción, tan sólo se pudieron analizar los posibles cambios a partir de 1200 pares de bases desde la inserción debido a las dificultades para diseñar oligonucleótidos.

Por último, se muestra el análisis realizado en las líneas resistentes PA-1-pRAM6 y PA-1-pRAM8 cuyas inserciones presentan un enriquecimiento bajo de H3K4me3 (Figura 28F,

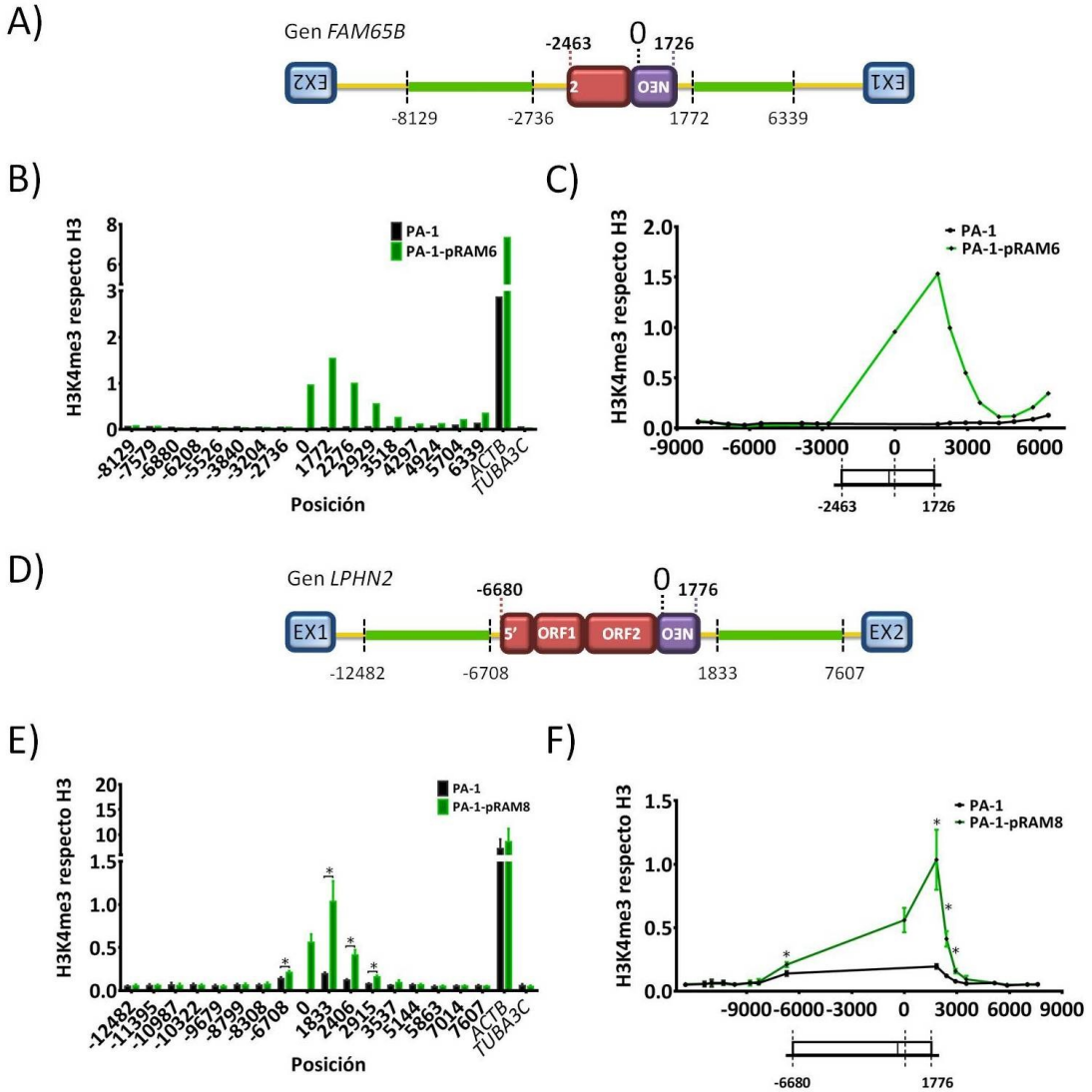


**Figura 34: Cambios de enriquecimiento de H3K4me3 en el genoma adyacente de las líneas PA-1-pRAM4 y H9-pRAM3.** **A)** Esquema que muestra la inserción de LINE-1 y la región genómica analizada en la línea PA-1-pRAM4. **B)** Gráfico de barras que muestra el CHIP-qPCR de H3K4me3 en el genoma adyacente en la línea parental PA-1 y en la línea clonal PA-1-pRAM4. Se muestra la media de tres experimentos independientes con su correspondiente desviación estándar. Los p-valores obtenidos en las posiciones marcadas con un \* fueron los siguientes: posición -4107 p-valor = 0,0234622; posición -3435 p-valor= 0,00448622, posición -3114 p-valor: 0,00442331, posición -2724 p-valor 0,0032916, posición -1293 p-valor= 0,000357508, posición -757 p-valor= 0,00170264, posición 2928 p-valor= 0,0126601. **C)** Gráfico de series que muestra las distancias reales respecto al punto 0 en el experimento mostrado en B). **D)** Esquema que muestra la inserción de LINE-1 y la región genómica analizada en la línea H9-pRAM3. **E)** Gráfico de barras que muestra el CHIP-qPCR de H3K4me3 en el genoma adyacente en la línea parental H9 y en la línea clonal H9-pRAM3. Se muestra la media de tres experimentos independientes con su correspondiente desviación estándar.

Los p-valores obtenidos en las posiciones marcadas con un \* fueron los siguientes: posición -2593 p-valor=0.00894929 y posición -1535 p-valor=0.000163545. **F)** Gráfico de series que muestra las distancias reales respecto al punto 0 en el experimento mostrado en E). En **B)** y **E)** ACTB es un control positivo mientras que TUBA3C es un control negativo. En **C)** y **F)** se muestra un esquema simplificado de la inserción donde se resaltan el último nucleótido en el extremo 5', la posición del punto 0 y el último nucleótido en el extremo 3'.

G). En ambos casos, se detectaron cambios en la composición de H3K4me3 fundamentalmente aguas abajo de la inserción. Curiosamente, el cambio inducido en la primera posición (posición 1772 para PA-1-pRAM6 y posición 1833 para PA-1-pRAM8) en el genoma adyacente en ambas líneas, es de mayor magnitud que el propio enriquecimiento presente sobre la propia inserción (Figura 35B, E).

Analizando las líneas de forma separada se observa como la inserción perteneciente a la línea PA-1-pRAM6 generó cambios en el estatus de esta modificación aguas abajo de la inserción llegando los cambios hasta la posición 3518 lo cual significa que el cambio se extiende 1792 pares de bases desde el punto 1726 (Figura 35B, C). Aunque, no se pudo determinar la significancia estadística debido a la falta de réplicas biológicas, estos cambios parecen bastante sólidos por la diferencia existente respecto a la línea parental en las mismas posiciones así como por el valor de estos enriquecimientos respecto al control negativo (Figura 35B). Por otro lado, la región analizada del genoma adyacente en la línea PA-1-pRAM6 es una región del genoma desprovista de H3K4me3 ya que los valores de enriquecimiento para esta modificación son similares al control negativo en la línea control PA-1 wt (Figura 35B, barras negras). Igualmente, la inserción de LINE-1 generó cambios fundamentalmente aguas abajo de la inserción en la línea PA-1-pRAM8. De hecho, se producen cambios significativos hasta el punto 2915 lo cual significa que los cambios se extienden 1139 pares de bases aguas abajo desde el punto 1776 (Figura 35E, F). Por otro lado, solo se observan cambios significativos en un punto aguas arriba de la inserción el cual se encuentra a 28 pares de bases de la inserción y se corresponde con el punto -6708 (Figura 35E,F). Sin embargo, según el criterio establecido no se considera un cambio válido ya que se encuentra a una distancia <500 pares de bases de la inserción. Además, los valores de enriquecimiento en estas posiciones se encuentran siempre dentro del rango establecido por los controles biológicos y por encima del valor medio del control negativo (TUBA3C). Es importante destacar que según los valores de enriquecimiento de H3K4me3 en la región analizada en la línea control PA-1 wt, se trata de una región donde existe un sutil enriquecimiento de H3K4me3 en las posiciones cercanas al sitio de inserción (Figura 35E, barras negras).

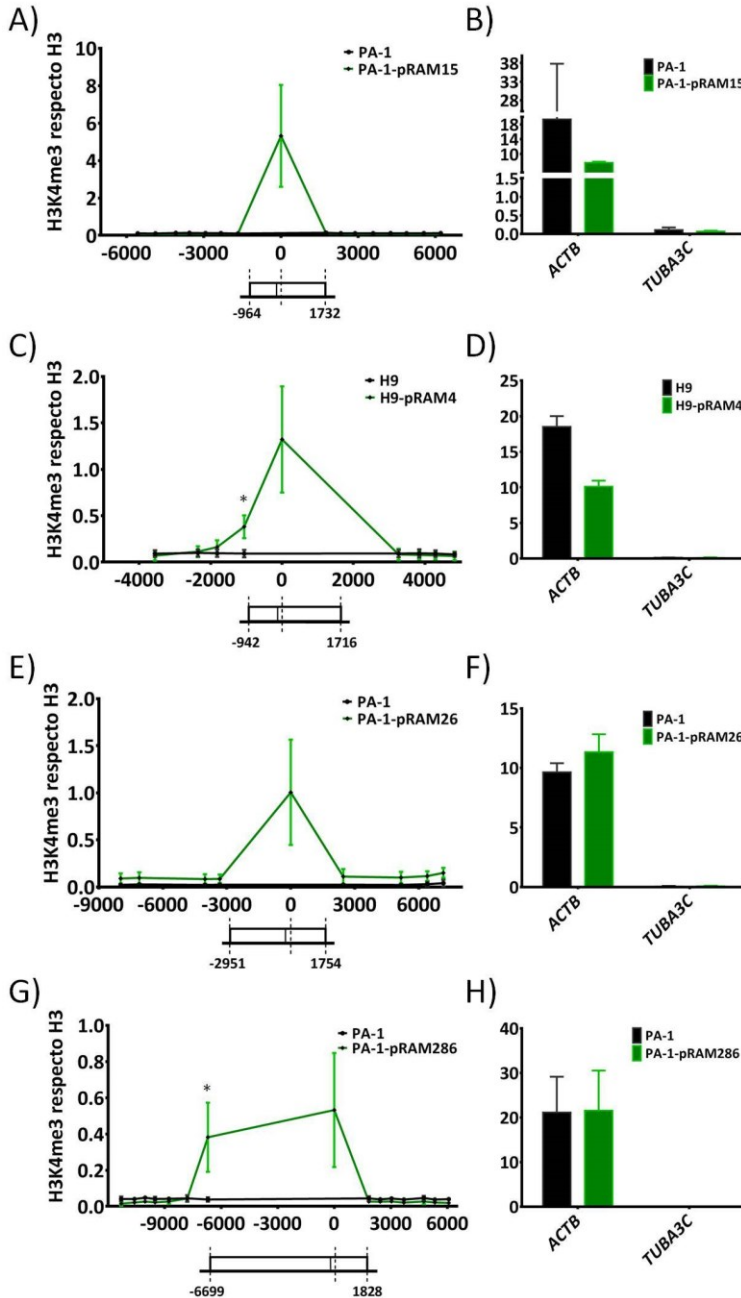


**Figura 35: Cambios de enriquecimiento de H3K4me3 en el genoma adyacente en las líneas PA-1-pRAM6 y PA-1-pRAM8.** A) Esquema que muestra la inserción de LINE-1 y la región genómica analizada en la línea PA-1-pRAM6. B) Gráfico de barras que muestra el CHIP-qPCR de H3K4me3 en el genoma adyacente en la línea parental PA-1 y en la línea clonal PA-1-pRAM6. Se muestra un único experimento. C) Gráfico de series que muestra las distancias reales respecto al punto 0 en el experimento mostrado en B). D) Esquema que muestra la inserción de LINE-1 y la región genómica analizada en la línea PA-1-pRAM8. E) Gráfico de barras que muestra el CHIP-qPCR de H3K4me3 en el genoma adyacente en la línea parental PA-1 y en la línea clonal PA-1-pRAM8. Se muestra la media de tres experimentos independientes con su correspondiente desviación estándar. Los p-valores obtenidos en las posiciones marcada con un \* fueron los siguientes: posición -6708 p-valor=0,0114049, posición 1833 p-valor=0,00354675, posición 2406 p-valor=0,00123494 y posición

2915  $p$ -valor= 0,0022143. **F)** Gráfico de series que muestra las distancias reales respecto al punto 0 en el experimento mostrado en E). En **B)** y **E)** ACTB es un control positivo mientras que TUBA3C es un control negativo. En **C)** y **F)** se muestra una representación simplificada de la inserción donde se resaltan el último nucleótido en el extremo 5', la posición del punto 0 y el último nucleótido en el extremo 3'.

Resumiendo, inserciones que presentan un enriquecimiento bajo o alto de H3K4me3 pueden inducir cambios en la composición de H3K4me3 en el genoma adyacente. Además, 3 de las 4 inserciones (H9-pRAM3, PA-1-pRAM6, PA-1-pRAM8) de LINE-1 que han producido cambios de H3K4me3 en el genoma adyacente se encuentran también enriquecidas en H3K9me3 (Figura 31C, F, G, Tabla 8). Los cambios generados son producidos fundamentalmente por inserciones truncadas en 5', excepto PA-1-pRAM8 que presenta una pequeña truncación careciendo de los primeros 22 nucleótidos (Ver Anexo IV). Estos cambios pueden alcanzar una distancia media entre 1kb y 3,3 kb, detectándose la mayor distancia en la línea PA-1-pRAM4 con la inserción más truncada y además con un enriquecimiento intermedio/alto de H3K4me3 (Figura 28A, Figura 34B).

En segundo lugar, se muestran los resultados del análisis del genoma adyacente para las cuatro líneas restantes cuyas inserciones no generaron cambios en el genoma adyacente. Estas inserciones se encontraban enriquecidas en H3K4me3 en algún grado (Figura 28B, D, E, H). El análisis de la línea resistente PA-1-pRAM15, cuya inserción presenta un enriquecimiento intermedio/alto de H3K4me3 (Figura 28B), reveló la ausencia de cambios en el estatus de H3K4me3 (Figura 36A). Por otro lado, el análisis de la línea resistente H9-pRAM4, cuya inserción presenta un enriquecimiento bajo (Figura 28D), demostró la presencia de un pequeño cambio significativo de H3K4me3 118 pares de bases aguas arriba de la inserción en la posición -1060 (Figura 36C). Aún estando dentro del rango de valores establecido por los controles biológicos, este cambio se encuentra demasiado cerca de la inserción con lo cual no puede ser considerado como válido según el criterio establecido. Cabe destacar que los posibles cambios generados aguas abajo no se pudieron analizar antes de las primeras 1500 pares de bases debido a problemas para el diseño de oligonucleótidos. De la misma forma, tampoco se encontraron cambios en la línea resistente PA-1-pRAM26 (Figura 36E), cuya inserción presenta un enriquecimiento bajo de H3K4me3 (Figura 28E). Por último, el análisis de la línea sensible PA-1-pRAM286, cuya inserción presentaba un enriquecimiento marginal de H3K4me3 (Figura 28H), mostró un mínimo cambio significativo en la posición -6702 que se encuentra 3 pares de bases aguas arriba de la inserción (Figura 36G). Al igual que ocurría con la línea H9-pRAM4, el cambio detectado no reúne las condiciones para ser considerado válido según el criterio de distancia establecido. Por tanto, este análisis demuestra que no todas las inserciones



**Figura 36: Ausencia de cambios de enriquecimiento de H3K4me3 en el genoma adyacente en las líneas clonales PA-1-pRAM15, H9-pRAM4, PA-1-pRAM26 y PA-1-pRAM286. A), C), E), G)** Gráfico de series que muestra los resultados del experimento de ChIP-qPCR de H3K4me3 en el genoma adyacente en las líneas PA-1-pRAM15, H9-pRAM4, PA-1-pRAM26 y PA-1-pRAM286 respectivamente. Abajo, se muestra una representación simplificada de la inserción donde se resalta el último nucleótido en el extremo 5', la posición del punto 0 y el último nucleótido en el extremo 3'. En **A)** se muestran los valores medios de 2 réplicas biológicas. Las barras de error muestran el rango de valores. En **C), E), G)** se muestran los valores medios de tres réplicas biológicas con su correspondiente desviación estándar. En **C)** y **G)** los p-valores obtenidos en las posiciones señaladas con un \* fueron los siguientes: en **C)** p-valor=0,0178477 y en **G)** p-valor=0,0357738. **B), D), F), H)** Gráfico de barras que muestra los controles correspondientes al experimento mostrado en **A), C), E), G)** respectivamente. ACTB es un control positivo mientras que TUBA3C es un control negativo.

truncadas (Anexo IV), enriquecidas en H3K4me3 pertenecientes a líneas resistentes generan cambios de H3K4me3 en el genoma adyacente. Por otro lado, las inserciones de líneas sensibles no generan cambios de H3K4me3.

Para concluir, los cambios de H3K4me3 son inducidos fundamentalmente por algunas inserciones truncadas pertenecientes a líneas resistentes independientemente del grado de enriquecimiento de H3K4me3 existente sobre la inserción. Estos cambios pueden alcanzar una distancia media de 1kb-3,3kb. Además estos cambios son producidos a pesar de que algunas de estas inserciones se encuentran enriquecidas simultáneamente en H3K4me3 y H3K9me3.

#### *4.2.4.2. Análisis de la acetilación de histonas*

En este apartado, se analizan los posibles cambios en los niveles de acetilación en las proximidades de la inserción de la línea PA-1-pRAM4. Esta inserción es la única que estaba enriquecida en H4ac y H3K9ac (Figura 29A, 30A). En primer lugar, se muestra el análisis de H4ac. A pesar del enriquecimiento de H4ac sobre la inserción, no se detectaron cambios en el estatus de esta modificación en la región genómica adyacente a la inserción (Figura 37B). Por otra parte, se detectó un pequeño cambio significativo en el enriquecimiento de H3K9ac en el genoma adyacente en la posición -757 (Figura 37C). El valor de enriquecimiento que se alcanza en dicha posición está comprendido dentro del rango establecido por los controles biológicos del experimento pero dicho cambio se produce únicamente 9 pares de bases aguas arriba de la inserción. Por esta razón no se considera un cambio válido.

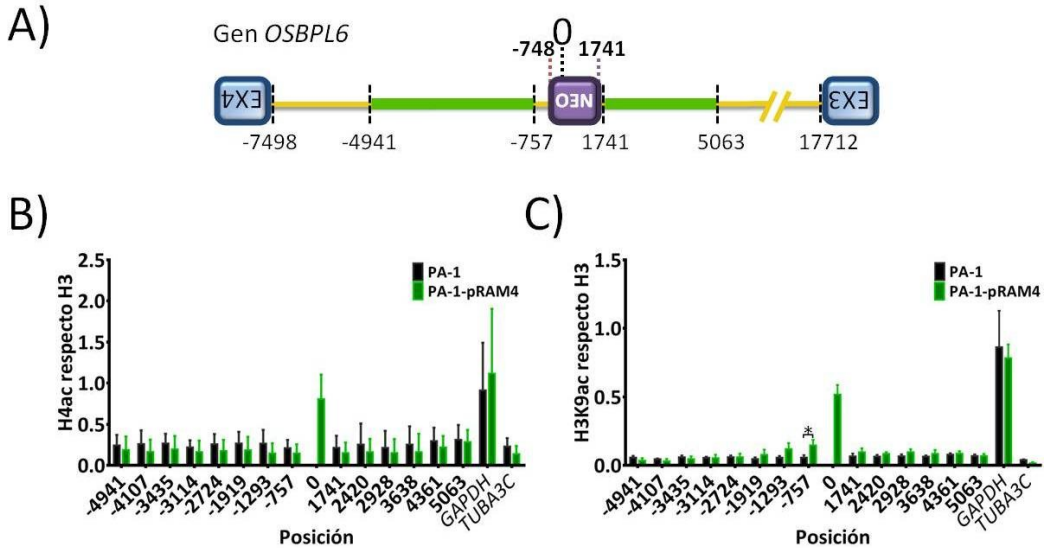
La inserción analizada perteneciente a la línea PA-1-pRAM4 presenta un enriquecimiento intermedio/alto de H3K4me3 (Figura 28A) y producía la mayor expansión de cambios de H3K4me3 en el genoma adyacente (Figura 34B). Además esta línea presentaba la mayor expresión del gen *NEO* (Figura 33C). Este dato sugiere que es bastante probable que las inserciones *de novo* de LINE-1 no induzcan cambios en los niveles de acetilación en el genoma adyacente ya que no son ni siquiera detectados en un contexto de actividad transcripcional tan favorable. Análisis realizados en el genoma adyacente de líneas cuyas inserciones no estaban enriquecidas en H4ac, tampoco mostraron la existencia de cambios de enriquecimiento de H4ac o H3K9ac (datos nos mostrados).

#### *4.2.4.3. Análisis de H3K9me3*

En este apartado se analizan los cambios de H3K9me3 en 5 líneas resistentes y en 3 líneas sensibles cuyas inserciones estaban enriquecidas en H3K9me3 en algún grado. En

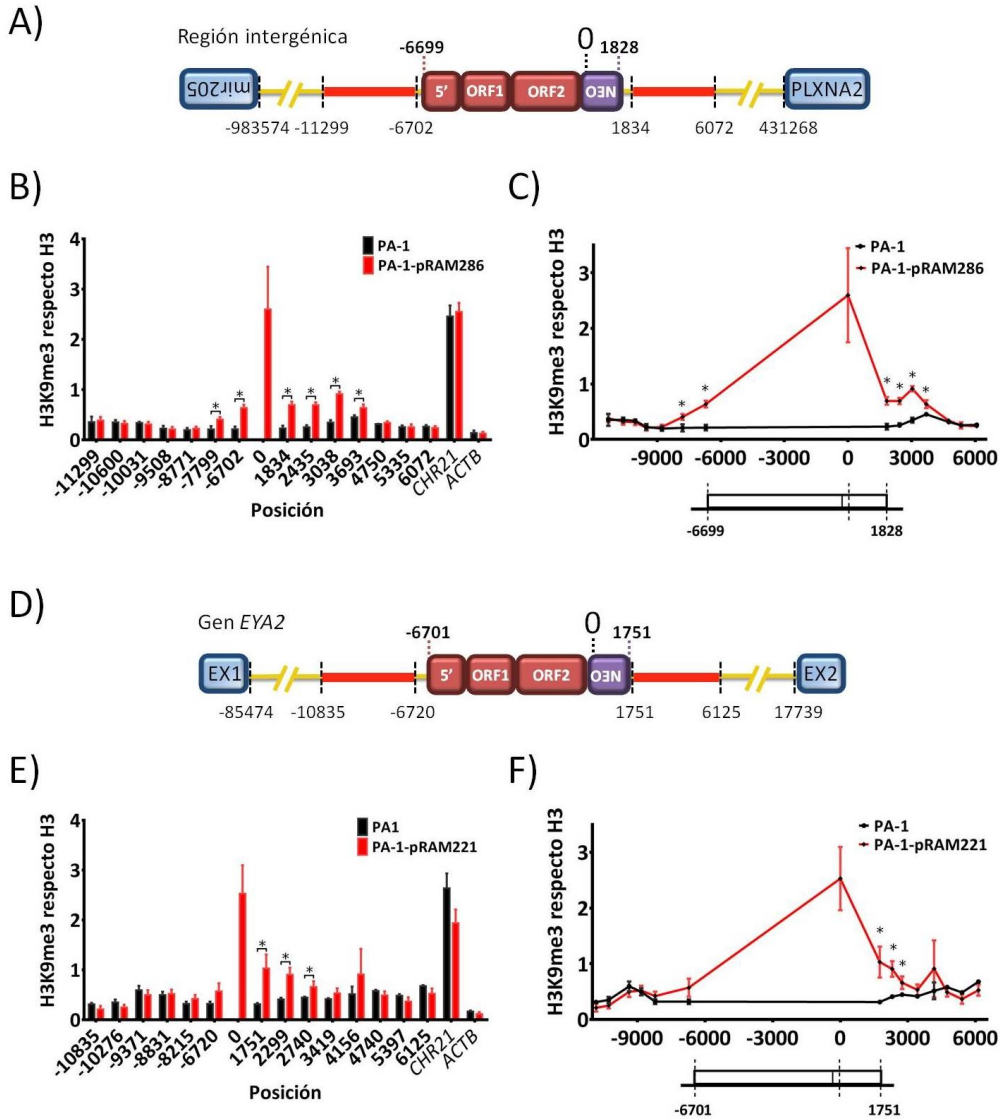


primer lugar, se muestran los resultados para dos líneas clonales sensibles (PA-1-pRAM286 y PA-1-pRAM221) cuyas inserciones indujeron cambios de enriquecimiento de H3K9me3 en el genoma adyacente. Ambas inserciones se encontraban altamente enriquecidas en H3K9me3, H3K27me3 y nada o marginalmente enriquecidas en H3K4me3 (Figura 31H, I;



**Figura 37: Ausencia de cambios de enriquecimiento de H3K9ac y H4ac en el genoma adyacente en la línea PA-1-pRAM4.** **A)** Esquema que muestra la inserción de LINE-1 y la región genómica analizada en la línea PA-1-pRAM4. **B)** Gráfico de barras que muestra el experimento de ChIP-qPCR de H4ac en el genoma adyacente en la línea parental PA-1 y en la línea clonal PA-1-pRAM4. Se muestran los valores medios de dos experimentos independientes. Las barras de error muestran el rango de valores. **C)** Gráfico de barras que muestra el experimento de ChIP-qPCR de H3K9ac en el genoma adyacente en la línea parental PA1 y en la línea clonal PA-1-pRAM4. Se muestran los valores medios de tres experimentos independientes con su desviación estándar. El p-valor en la posición -757 fue 0,0238382. En **B)** y **C)** GAPDH es un control positivo mientras que TUBA3C es un control negativo.

Figura 32H, I; Figura 28H, I; tabla 8). La inserción presente en la línea PA-1-pRAM286 es capaz de inducir cambios significativos en el enriquecimiento de H3K9me3 aguas abajo y aguas arriba de la inserción. En este caso, los cambios se expanden hasta 1865 pares de bases aguas abajo desde el punto 1828 hasta la posición 3693 (Figura 38B, C). Además, esta inserción también promueve cambios en el estatus de la modificación aguas arriba de la inserción. De hecho, se generan cambios significativos que alcanzan una distancia de 1100 pares de bases desde el punto -6699 hasta la posición -7799 (Figura 38B, C). Igualmente los valores que se alcanzan en estos puntos donde se han producido los



**Figura 38: Cambios de enriquecimiento de H3K9me3 en el genoma adyacente de las líneas PA-1-pRAM286 y PA-1-pRAM221. A)** Esquema que muestra la inserción de LINE-1 y la región genómica analizada en la línea PA-1-pRAM286. **B)** Gráfico de barras que muestra el CHIP-qPCR de H3K9me3 en el genoma adyacente en la línea parental PA-1 y en la línea clonal PA-1-pRAM286. Los p-valor obtenidos en las posiciones marcadas con un \* fueron los siguientes: posición -7799 p-valor=0,0151263, posición -6702 p-valor=0,000767813, posición 1834 p-valor=0,000775231, posición 2435 p-valor=0,000380536, posición 3038 p-valor=0,000114439 y posición 3693 p-valor=0,0166808. **C)** Gráfico de series que muestra las distancias reales respecto al punto 0 en el experimento mostrado en B). **D)** Esquema que muestra la inserción de LINE-1 y la región genómica analizada en la línea PA-1-pRAM221. **E)** Gráfico de barras que muestra el CHIP-qPCR de H3K9me3 en el genoma adyacente en la línea parental PA-1 y en la línea clonal PA-1-pRAM221. Los p-valor

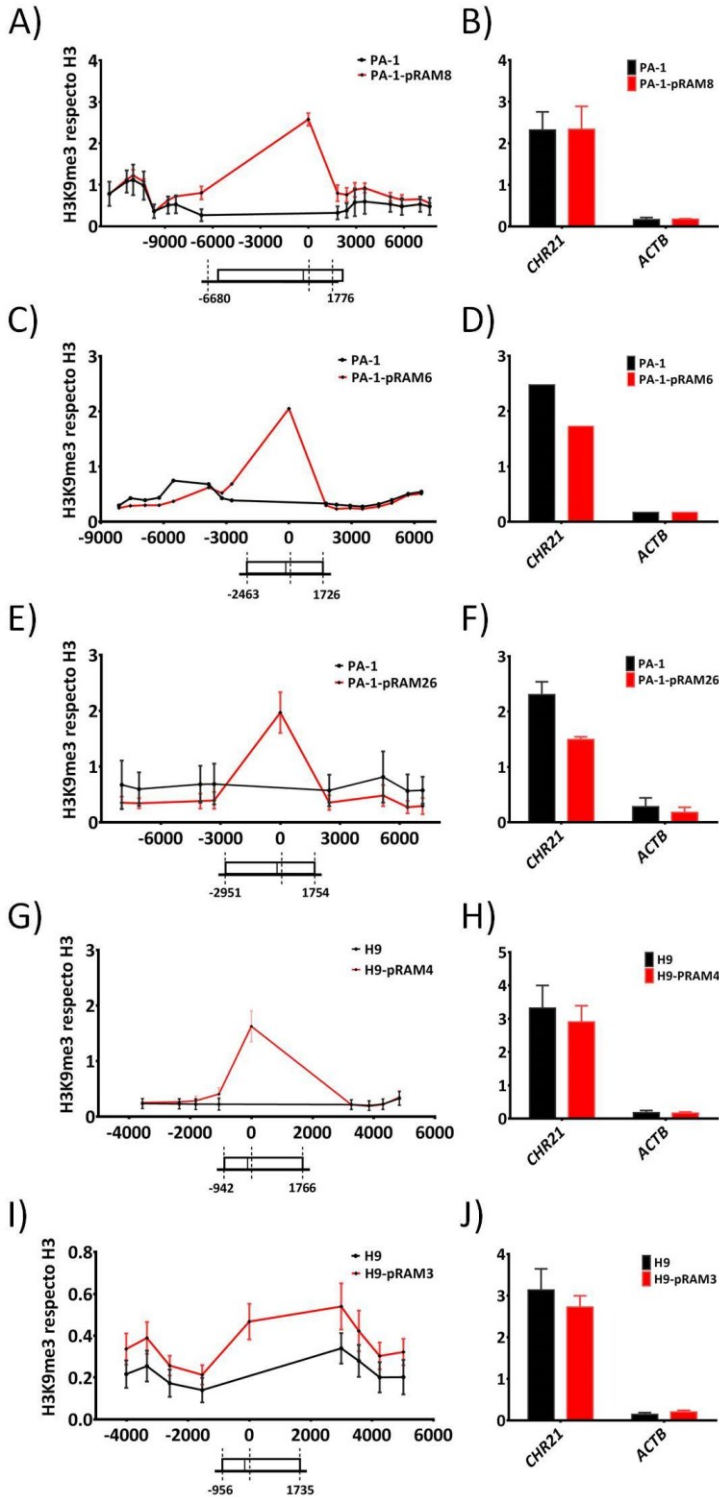
obtenidos en las posiciones marcadas con un \* fueron los siguientes: posición 1751  $p$ -valor=0,0110201, posición 2299  $p$ -valor=0,00385384 y posición 2740  $p$ -valor=0,0315478. **F)** Gráfico de series que muestra las distancias reales respecto al punto 0 en el experimento mostrado en E). En **B)** y **E)** se muestra la media de tres experimentos independientes con su desviación estándar. En **B)** y **E)** CHR21 es un control positivo mientras que ACTB es un control negativo. En **C)** y **F)** se muestra una representación simplificada de la inserción donde se resaltan el último nucleótido en el extremo 5', la posición del punto 0 y el último nucleótido en el extremo 3'.

cambios se encuentran dentro del rango delimitado por los controles del experimento. De hecho, el valor mínimo de enriquecimiento que se alcanza en estas 6 posiciones con cambios significativos representa el triple del control negativo del experimento. Por otro lado, la región genómica analizada es una región que podría estar sutilmente heterocromatinizada ya que los valores medios de enriquecimiento de H3K9me3 en las posiciones analizadas en la línea control PA-1 wt representan como mínimo el doble del valor medio del control negativo (Figura 38B, barras negras).

Igualmente, la inserción perteneciente a la línea PA-1-pRAM221 genera cambios significativos en el estatus de H3K9me3 que se extienden 989 pares de bases aguas abajo desde la posición 1751 hasta llegar a la posición 2740 (Figura 38E, F). Los valores de enriquecimiento en las posiciones donde se han encontrado los cambios se encuentran dentro del rango delimitado por los controles del experimento. En cuanto a la composición de H3K9me3 en la región analizada, se observa que puede existir una cierta heterocromatinización ya que los valores medios de enriquecimiento de H3K9me3 son ligeramente superiores al control negativo (Figura 38E, barras negras).

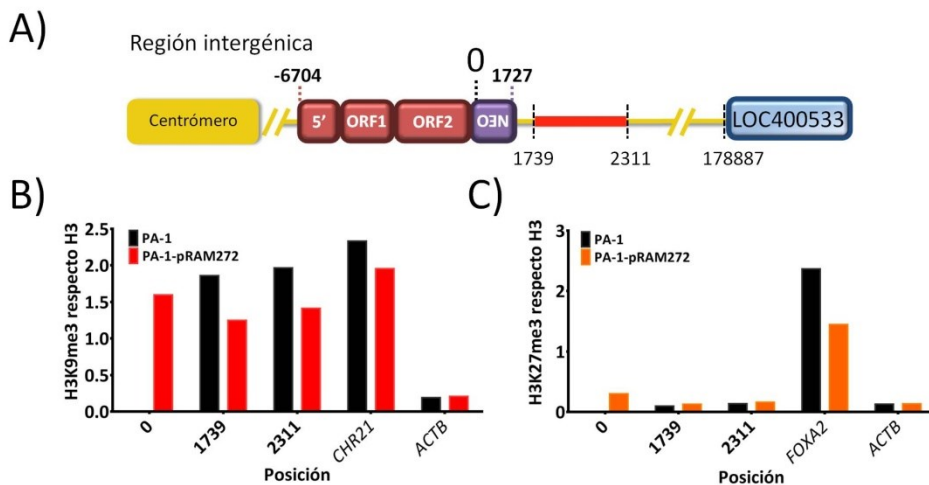
En resumen, las inserciones de las líneas sensibles PA-1-pRAM221 y PA-1-pRAM286 altamente enriquecidas en H3K9me3, H3K27me3 y nada o marginalmente enriquecidas en H3K4me3 producen cambios en el enriquecimiento de H3K9me3 que se extienden entre 1kb y 1,5kb desde la inserción. Además, estos cambios se generan tanto aguas arriba como aguas abajo. Cabe destacar que ambas inserciones son de copia completa (Ver Anexo IV).

En segundo lugar, se muestra el análisis para un grupo de 6 líneas clonales cuyas inserciones no generaron cambios de H3K9me3. En este grupo, se encuadran 5 líneas clonales resistentes (PA-1-pRAM8, PA-1-pRAM6, PA-1-pRAM26, H9-pRAM4 y H9-pRAM3) y 1 línea clonal sensible (PA-1-pRAM272). El análisis de la líneas resistentes PA-1-pRAM8, PA-1-pRAM6 y PA-1-pRAM26, cuyas inserciones presentan un enriquecimiento elevado de H3K9me3 (Figura 31G, F, E), demostró la ausencia de cambios en las tres líneas (Figura



**Figura 39: Ausencia de cambios en el enriquecimiento de H3K9me3 en el genoma adyacente en las líneas PA-1-pRAM8, PA-1-pRAM6, PA1-pRAM26, H9-RAM4 y H9-RAM3. A), C), E), G), I)** Gráfico de series que muestra los resultados del experimento de ChIP-qPCR de H3K9me3 en el genoma adyacente en las líneas PA-1-pRAM8, PA-1-pRAM6, PA-1-pRAM26, H9-pRAM4, H9-pRAM3 respectivamente. Abajo, se muestra una representación simplificada de la inserción donde se resalta el último nucleótido en el extremo 5', la posición del punto 0 y el último nucleótido en el extremo 3'. En **A), C) y E)** se muestran los valores medios de 2 réplicas biológicas. Las barras de error muestran el rango de valores. En **G) e I)** se muestran los valores medios de tres réplicas biológicas con su correspondiente desviación estándar. **B), D), F), H), J)** Gráfico de barras que muestra los controles correspondientes al experimento mostrado en **A), C), E), G), I)** respectivamente. CHR21 es un control positivo mientras que ACTB es un control negativo.

39A, C, E). Por otro lado, inserciones pertenecientes a las líneas resistentes H9-pRAM4 y H9-pRAM3 que presenta un enriquecimiento intermedio y bajo respectivamente de H3K9me3 (Figura 31D, C), tampoco produjeron cambios de H3K9me3 (Figura 39G, I). Cabe destacar que en la línea H9-pRAM4 no se pudo analizar el genoma adyacente antes de las primeras 1500 pares de bases debido a problemas para diseñar oligonucleótidos. Las inserciones de todas estas líneas resistentes analizadas además se encuentran enriquecidas en H3K4me3 (Figura 28G, F, E, D, C, tabla 8). Por último, el análisis en la línea sensible PA-1-pRAM272, cuya inserción se encuentra altamente enriquecida en H3K9me3 (Figura 31J), mostró, igualmente la ausencia de cambios (Figura 40B). Esta inserción se encuentra en la región pericentromérica del cromosoma 16 (Figura 40A), donde ya existía un elevado enriquecimiento de H3K9me3 (Figura 40B, barras negras). En resumen, las inserciones de líneas resistentes enriquecidas en H3K4me3 y H3K9me3 no producen cambios de H3K9me3 independientemente del grado de enriquecimiento de H3K9me3 existente sobre la inserción y de la longitud de la inserción ya que en este grupo hay dos inserciones muy truncadas (H9-pRAM4, H9-pRAM3), dos inserciones con puntos intermedios de truncación (PA-1-pRAM6, PA-1-pRAM26) y una inserción con un truncación mínima de 22 nucleótidos (PA-1-pRAM8) (ver Anexo IV). Además, inserciones de líneas sensibles en lugares del genoma con un elevado nivel de H3K9me3 tampoco producen cambios.



**Figura 40: Ausencia de cambios de enriquecimiento de H3K9me3 y H3K27me3 en el genoma adyacente en la línea PA-1-pRAM272.** A) Esquema que muestra la inserción de LINE-1 y la región genómica analizada en la línea PA-1-pRAM272. B) Gráfico de barras que muestra los valores de enriquecimiento de un experimento de ChIP-PCR de H3K9me3 en el genoma adyacente en la línea parental PA-1 y en la línea clonal PA-1-pRAM272. CHR21 es un control positivo y ACTB es un control

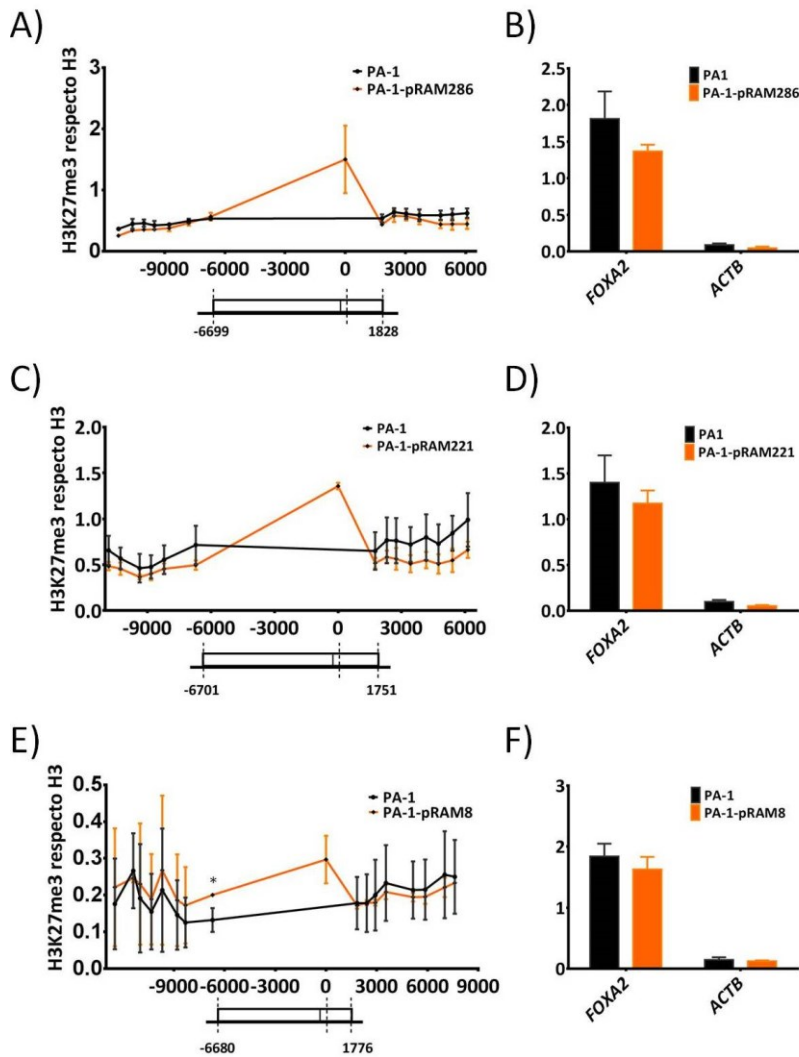
negativo. **C)** Grafico de barras que muestra los valores de enriquecimiento de un experimento de ChIP-PCR de H3K27me3 en el genoma adyacente en la línea parental PA-1 y en la línea clonal PA-1-pRAM 272. FOXA2 es un control positivo mientras que ACTB es un control negativo.

Conjuntamente, estos análisis sugieren que únicamente las inserciones de líneas sensibles producen cambios de enriquecimiento de H3K9me3, que pueden extenderse hasta 1kb-1,5kb desde la inserción. Además, estas inserciones son de copia completa, están altamente enriquecidas en H3K9me3 y H3K27me3 y marginalmente o nada enriquecidas en H3K4me3.

#### **4.2.4.4. Análisis de H3K27me3**

En este caso, se realizó el análisis sobre las 4 líneas cuyas inserciones se encuentran enriquecidas en H3K27me3. En primer lugar, se muestra el análisis para las dos líneas sensibles cuyo enriquecimiento sobre la inserción de LINE-1 fue elevado (Figura 32H, I). Como se observa en la figura 41A, C tanto en la línea PA-1-pRAM286 como en la línea PA-1-pRAM221 no se detectan cambios en el enriquecimiento de H3K27me3 en el genoma adyacente. Cabe destacar que las inserciones presentes en las líneas PA-1-pRAM286 y PA-1-pRAM221, se encuentran en lugares del genoma con enriquecimientos bajos/intermedios de H3k27me3 según los valores de enriquecimiento de H3K27me3 en la línea PA-1 wt en la región analizada (Figura 41A, C; línea negra). De hecho, para la región analizada en la línea PA-1-pRAM286 los valores de enriquecimiento de H3K27me3 se pueden catalogar como bajos/intermedios (Figura 41A, línea negra) mientras que los valores de enriquecimiento de H3K27me3 en el lugar de inserción de la línea PA-1-pRAM221 se pueden catalogar como intermedios (Figura 41C, línea negra) comparado con los valores máximos de los controles positivos. Cabe mencionar que estas líneas son las que presentaban cambios de enriquecimiento de H3K9me3 en el genoma adyacente (Figura 38B, E) cuyas inserciones estaban altamente enriquecidas en H3K9me3 (Figura 31 H, I).

En segundo lugar, se muestra el análisis de la línea resistente PA-1-pRAM8 cuya inserción presenta un enriquecimiento bajo de H3K27me3 (Figura 32G). Esta inserción no genera cambios en la composición de H3K27me3 salvo un pequeño cambio significativo 28 pares de bases aguas arriba de la posición -6680, justamente en la primera posición que se mide dentro del genoma adyacente (Figura 41E). Aún siendo significativo y quedando dentro del rango delimitado por los controles del experimento, este cambio no es considerado un cambio válido ya que se produce sólo 28 pares de bases aguas arriba de la inserción.



**Figura 41: Ausencia de cambios de enriquecimiento de H3K27me3 en el genoma adyacente en las líneas PA1-pRAM286; PA-1-pRAM221 y PA-1-pRAM8. A), C), E)** Gráfico de series que muestra los resultados del experimento de ChIP-qPCR de H3K27me3 en el genoma adyacente de las líneas PA-1-pRAM286, PA-1-pRAM221 y PA-1-pRAM8 respectivamente. Se muestran los valores medios de 3 réplicas biológicas con su desviación estándar. Abajo, se muestra una representación simplificada de la inserción donde se resaltan el último nucleótido en el extremo 5', la posición en el punto 0 y el último nucleótido en el extremo 3'. **B), D), F)** Gráfico de barras que muestra los controles correspondientes al experimento mostrado en A), C), E). FOXA2 es un control positivo mientras que ACTB es un control negativo.

Por último, se analizan los resultados para la línea sensible PA-1-pRAM272 cuya inserción parece presentar un enriquecimiento bajo de H3K27me3 (Figura 32J). Al igual

que anteriormente, la línea PA-1-pRAM272 no muestra cambios en el estatus de esta modificación en las únicas dos posiciones genómicas analizadas en este caso (Figura 40C). El nivel de enriquecimiento de H3K27me3 en la región analizada en esta línea es bastante similar a los controles negativos (Figura 40C, barras negras).

Según los análisis realizados, las inserciones artificiales *de novo* de LINE-1 no generan cambios en el enriquecimiento de H3K27me3 en el genoma adyacente independientemente del grado de enriquecimiento de esta modificación sobre la inserción de LINE-1 y del tipo de línea analizada. Además, las cuatro líneas analizadas (PA-1-pRAM8, PA-1-pRAM286, PA-1-pRAM221 y PA-1-pRAM272) presentaban simultáneamente un elevado enriquecimiento de H3K9me3 (Figura 31G, H, I, J, Tabla 8).

Línea /Inserción	Marcas Activas			Marcas represivas		Longitud		Expansión	
	K4me3	H4ac	K9ac	K9me3	K27me3	Com <sup>1</sup>	Trun <sup>2</sup>	K4 <sup>3</sup>	K9 <sup>4</sup>
PA-1-pRAM4	x	x	x				x	x	
H9-pRAM3	x			x			x	x	
PA-1-pRAM6	x			x			x	x	
PA-1-pRAM8	x			x	x		x <sup>5</sup>	x	
PA-1-pRAM15	x						x		
H9-pRAM4	x			x			x		
PA-1-pRAM26	x			x			x		
PA-1-pRAM272				x	x	x			
PA-1-pRAM286	x			x	x	x			x
PA-1-pRAM221				x	x	x			x

**Tabla 9:** Resumen regulación epigenética e impacto epigenético de las inserciones artificiales *de novo* de LINE-1. <sup>1</sup>Com significa Copia completa; <sup>2</sup>Trun significa truncada; <sup>3</sup>K4 significa H3K4me3; <sup>4</sup>K9 significa H3K9me3; <sup>5</sup>Esta línea se incluye como truncada pero sólo le faltan 22 nucleótidos. En morado se resaltan las inserciones de líneas resistentes y en gris las inserciones de las líneas sensibles.

Finalmente, como se resume en la tabla 9, todos estos análisis sugieren conjuntamente, que los cambios de H3K4me3 son inducidos por inserciones enriquecidas en H3K4me3, localizadas en un contexto transcripcional favorable ya que pertenecen a líneas resistentes y además están truncadas en 5'. A pesar del enriquecimiento de H3K9me3 en 3/4 de estas inserciones, solo se inducen cambios en el estatus de H3K4me3. Por otro lado, los cambios de H3K9me3 son inducidos por inserciones que están muy enriquecidas en H3K9me3, localizadas en un contexto transcripcional represivo ya que pertenecen a líneas sensibles y además, son de copia completa. A pesar de estar también muy enriquecidas en H3K27me3, solo generan cambios en el enriquecimiento de H3K9me3. Estos datos sugieren que un contexto represivo favorece la generación de



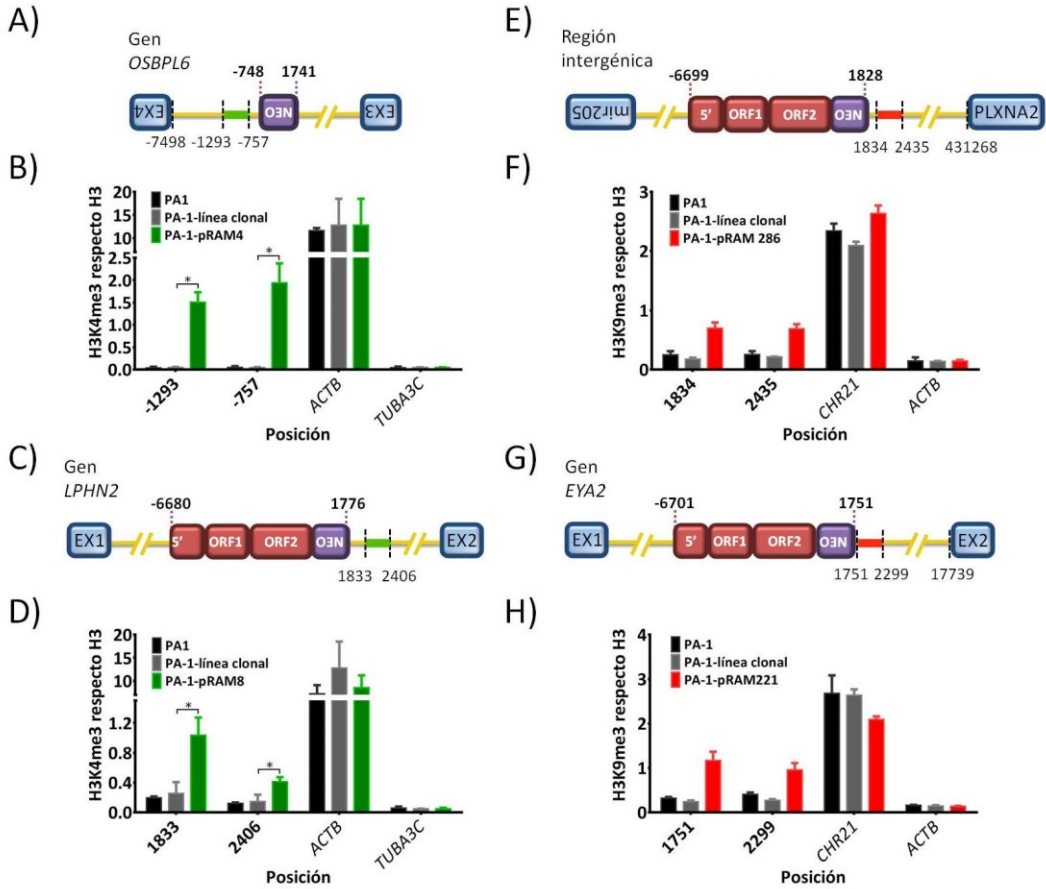
cambios en H3K9me3 en el genoma adyacente para reforzar el silenciamiento de las inserciones de copia completa.

#### 4.2.5. Estabilidad epigenética de las líneas clonales de PA-1

Teniendo en cuenta la variabilidad existente dentro del cultivo celular, es posible que las líneas clonales generadas para este estudio presenten diferencias de enriquecimiento de las modificaciones estudiadas a lo largo del genoma. Por esta razón, los cambios de enriquecimiento de H3K4me3 y H3K9me3 inducidos por la inserción de LINE-1 en el genoma adyacente de algunas líneas podrían, en realidad, representar el nivel de esa modificación en esa línea clonal. Para controlar esta limitación propia del diseño experimental, se analizó el nivel de enriquecimiento de H3K4me3 o H3K9me3 en otra línea clonal de PA-1 en dos posiciones genómicas en las cuales se habían detectado cambios en el estatus de estas marcas. Como se observa en la figura 42 B, D, la línea clonal control presenta los mismos niveles de H3K4me3 que la línea parental PA-1. De hecho, las posiciones genómicas analizadas se encuentran significativamente enriquecidas en las líneas PA-1-pRAM4 y PA-1-pRAM8 respecto a la línea clonal control (Figura 42B, D). Del mismo modo, en la figura 42F, H se observa que los niveles de H3K9me3 en las posiciones analizadas son exactamente iguales en la línea clonal control que en la línea parental. Además, las líneas PA-1-pRAM286 y PA-1-pRAM221 presentan un nivel de enriquecimiento de H3K9me3 superior a la línea clonal control en las posiciones analizadas (Figura 42F, H). Estos experimentos sugieren que las líneas clonales son epigenéticamente estables en la región analizada y, por tanto, excluyen la posibilidad de que los cambios observados sean un artefacto propio del diseño experimental. Por tanto, los cambios de H3K4me3 y H3K9me3 en el genoma adyacente son muy probablemente inducidos por la presencia de una inserción *de novo* de LINE-1.

#### 4.2.6. Eliminación de secuencias artificiales asociadas a las inserciones de novo de LINE-1

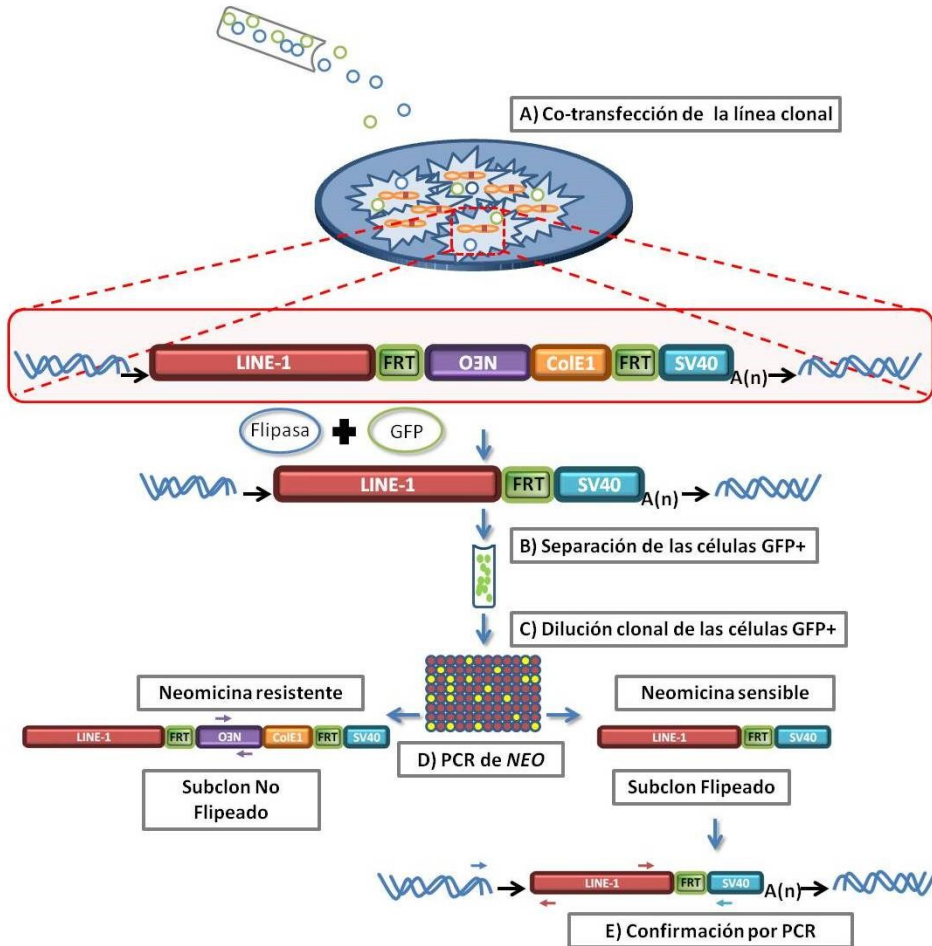
A la vista de los resultados obtenidos en los apartados anteriores, parece ser que el retroelemento LINE-1 después de su integración recluta H3K4me3 y H3K9me3 y en menor medida H3K27me3. No obstante, estos resultados han sido generados mediante el análisis de inserciones que llevan asociadas una serie de secuencias exógenas necesarias para la detección de la retrotransposición, la generación de las líneas clonales y para la caracterización de las inserciones, de ahí que se denomine a estas inserciones como artificiales. Estas secuencias se corresponden con un promotor vírico SV40, un gen de resistencia a neomicina y un origen de replicación bacteriano (ColE1) (Figura 24A). Por tanto, es posible que tales secuencias tengan un efecto sobre la regulación epigenética de



**Figura 42: Estabilidad epigenética de las líneas clonales de PA-1.** A), C), E), G) Esquema de localización de la inserción de LINE-1 en cada línea clonal donde se resaltan las posiciones analizadas por ChIP-qPCR. B), D) Gráfico de barras que muestra los ChIP-qPCRs de H3K4me3 en las posiciones señaladas en PA-1 wt, en una línea clonal de PA-1 y en PA-1-pRAM4 o PA-1-pRAM8 respectivamente. Se muestra la media de tres experimentos independientes con su desviación estándar. Los p-valores obtenidos en las posiciones marcadas con un \* son los siguientes: para B) en la posición -1293 p-valor=0,000358665, en la posición -757 p-valor= 0,00168042; para D) en la posición 1833 p-valor= 0,00852392, en la posición 2406 p-valor=0,0137055. ACTB es un control positivo y TUBA3C es un control negativo. F), H) Gráfico de barras que muestran los ChIP-qPCRs de H3K9me3 en las posiciones señaladas en PA-1 wt, en una línea clonal de PA-1 y en PA-1-pRAM286 o PA-1-pRAM221 respectivamente. Se muestra la media de dos experimentos independientes. Las barras de error muestran el rango de valores. CHR21 es un control positivo y ACTB es un control negativo.

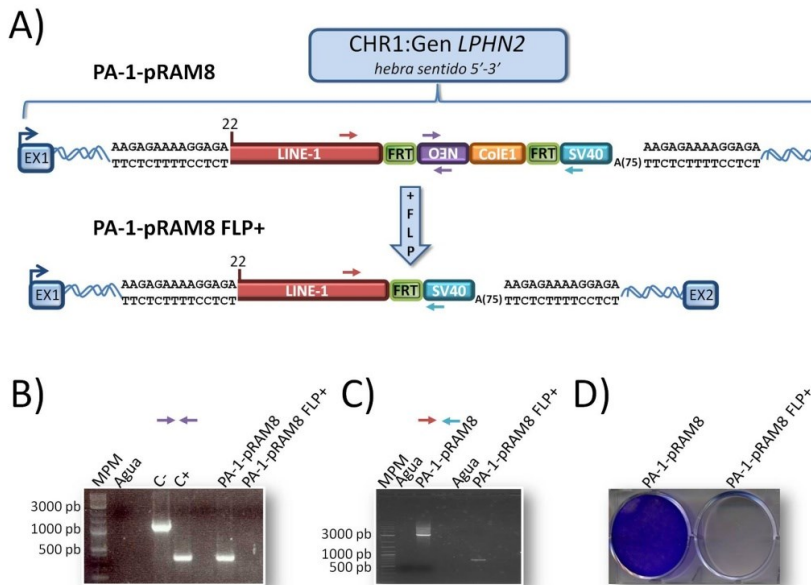
las inserciones *de novo* de LINE-1 analizadas en este estudio. Para abordar esta limitación, el vector pRAM-LRE3 presenta dos secuencias FRT (del inglés *Flippase Recognition Target*)

flanqueando el casete de recuperación o rescate. De esta forma, las secuencias exógenas fusionadas a la inserción de LINE-1 pueden ser eliminadas mediante recombinación mediada por flipasa. El proceso de eliminación o recombinación comienza con la co-

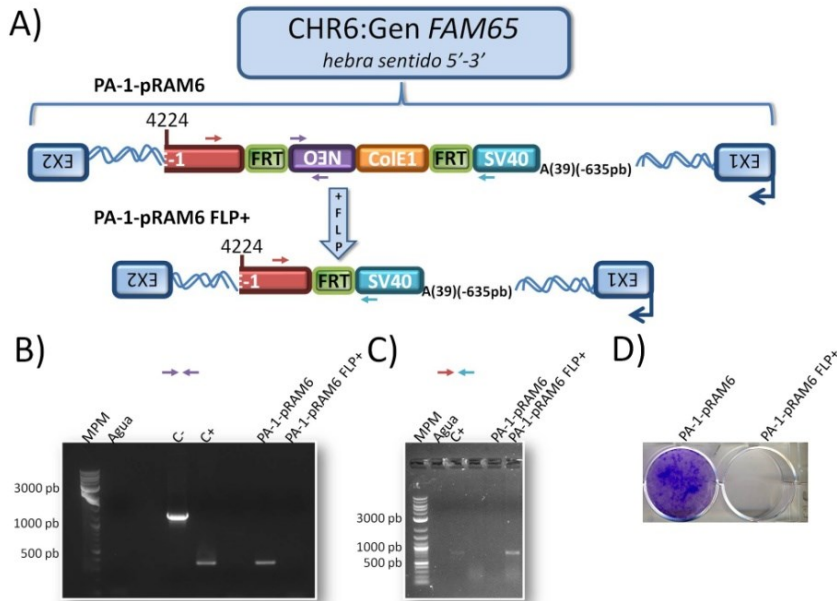


**Figura 43: Esquema del proceso de recombinación por flipasa.** A) La línea clonal con una inserción artificial de LINE-1 (rectángulos rojos) se co-transfecta con un plásmido que expresa la proteína verde fluorescente (pCEP4-EGFP, círculos verdes) y un plásmido que expresa la enzima flipasa (CMV-FLP, círculos azules). En las células transfectadas, la flipasa reconoce los sitios FRT. Las secuencias flanqueadas por las FRTs son escindidas. B) Para encontrar el subclon flipeado, se separan las células por citometría de flujo y se mantienen únicamente las células GFP+. C) Dilución clonal de las células GFP+. D) Búsqueda de clones negativos para la PCR de NEO. E) Posterior validación por PCR sobre el subclon flipeado con otras parejas de oligonucleótidos cuya elección dependerá de la inserción en cuestión.

transfección de cada línea clonal con un plásmido que expresa la enzima flipasa y otro plásmido que expresa EGFP para poder separar las células transfectadas GFP+ por citometría de flujo. Seguidamente, se realiza una dilución clonal de las células GFP+ y por PCR se buscan aquellos nuevos subclones que son negativos para la PCR de *NEO* (descrita en Figura 24C). Además en el caso de las líneas resistentes a neomicina, los nuevos subclones positivos después del tratamiento con flipasa se crecen en presencia de neomicina para comprobar su sensibilidad debido a la pérdida del gen *NEO*. Además se llevan a cabo PCRs alternativas para validar el proceso sobre los distintos subclones positivos (Figura 43A-E). Después de este proceso, la inserción de LINE-1 es prácticamente igual a una inserción de LINE-1 endógena excepto por la presencia de un pequeño fragmento de 180 nucleótidos correspondientes a la señal de poliadenilación SV40 y a una de las dos FRTs que quedan después de la recombinación. De este modo, se genera un subclon de la línea clonal original cuya inserción de LINE-1 presenta una localización



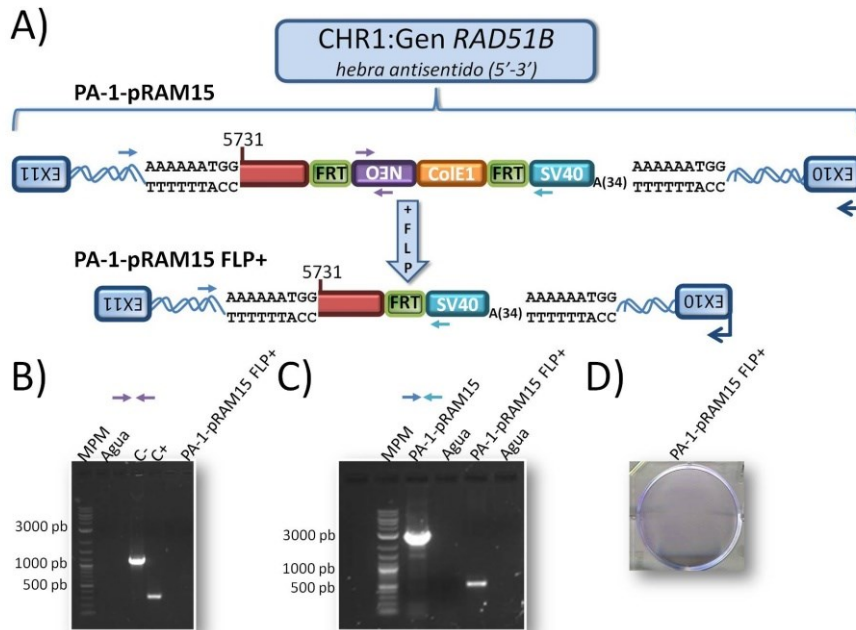
**Figura 44: Generación del subclon flipado en la línea pRAM8. A)** Esquema de la inserción antes y después de la recombinación mediada por flipasa. Las flechas indican sobre la inserción la posición de las parejas de oligonucleótidos utilizadas. **B)** PCR de *NEO* sobre la línea clonal PA-1-pRAM8 y el subclon PA-1-pRAM8 FLP+. **C)** PCR usando los oligonucleótidos 2H sentido (flecha roja) y SV40 REV FLP2 (flecha azul). Los tamaños de los productos de PCR fueron 3120 y 784 pares de bases. **D)** Ensayo de resistencia a neomicina en la línea PA-1-pRAM8 junto con el subclon PA-1-pRAM8 FLP+. En **B)** y **C)**, MPM significa marcador de peso molecular. En **B)** y **C)**, el color de las flechas indica la pareja de oligonucleótidos utilizada. En **B)**, C- significa control negativo y C+ significa control positivo.



**Figura 45: Generación del subclon flipeado en la línea PA-1-pRAM6. A)** Esquema que muestra el aspecto de la inserción antes y después de la recombinación por flipasa. Las flechas indican sobre la inserción la posición de las parejas de oligonucleótidos utilizadas. **B)** PCR de NEO sobre la línea clonal PA-1-pRAM6 y el subclon PA-1-pRAM6 FLP+. **C)** PCR usando los oligonucleótidos 2H sentido (flecha roja) y SV40 REV FLP2 (flecha azul). El tamaño del producto de PCR es 784 pares de bases. **D)** Ensayo de resistencia a neomicina en la línea PA-1-pRAM6 junto con el subclon flipeado (PA-1-pRAM6 FLP+). En **B)** y **C)**, MPM significa marcador de peso molecular. En **B)** y **C)**, el color de las flechas indica la pareja de oligonucleótidos utilizada. En **B)**, C- significa control negativo. En **B)** y **C)**, C+ significa control positivo.

genómica ya conocida pero sin las secuencias artificiales. El nuevo subclon será referido, partir de ahora, bien como “subclon flipeado” o bien con el nombre de la línea clonal seguido de FLP+. De ésta forma, se consiguió generar el subclon flipeado en 3 líneas clonales resistentes (PA-1-pRAM8, PA-1-pRAM6 y PA-1-1pRAM15 y en 1 sola línea clonal sensible (PA-1-pRAM272).

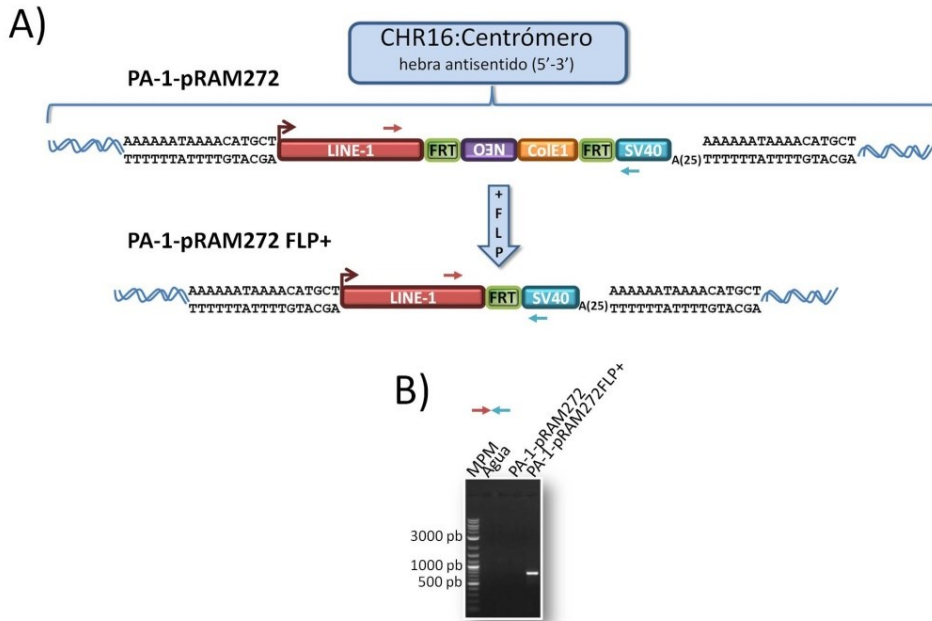
Como se puede comprobar en la figura 44A, se generó el subclon flipeado en la línea resistente PA-1-pRAM8, el cual fue negativo para la PCR de NEO (Figura 44B). Además, la PCR realizada con los oligonucleótidos indicados en azul y rojo (Figura 44A) dio lugar al fragmento pequeño en el subclon flipeado indicativo de la escisión del casete de recuperación (Figura 44C). De hecho, el subclon era sensible a neomicina (Figura 44D). Del mismo modo en la línea PA-1-pRAM6 se consiguió generar el subclon flipeado (Figura 45A) también negativo para la PCR de NEO (Figura 45B). Además, la PCR realizada con los



**Figura 46: Generación del subclon flipeado en la línea PA-1-pRAM15.** **A)** Esquema que muestra el aspecto de la inserción antes y después de la recombinación por flipasa. Las flechas indican sobre la inserción la posición de las parejas de oligonucleótidos utilizadas. **B)** PCR de NEO sobre el subclon flipeado. **C)** PCR usando los oligonucleótidos PRAM15 INSERT 2 (flecha azul oscuro) y SV40REV FLP2 (flecha azul claro). Los tamaños de los fragmentos de PCR son 2706 y 531 pares de bases. **D)** Ensayo de resistencia a neomicina en el subclon flipeado (PA-1-pRAM15 FLP+). En **B)** y **C)**, MPM significa marcador de peso molecular. En **B)** y **C)**, el color de las flechas indica la pareja de oligonucleótidos utilizada. En **B)**, C- significa control negativo y C+ significa control positivo.

oligonucleótidos indicados en azul y rojo, anteriormente mostrada en figura 44C, sobre el subclon flipeado da lugar al fragmento pequeño que no se amplifica en la línea clonal parental (Figura 45C). Por tanto, el subclon es sensible a neomicina (Figura 45D). En la línea PA-1-pRAM15 también se generó el subclon flipeado correspondiente que era negativo para la PCR de NEO (Figura 46A, B). La realización de la PCR con los oligonucleótidos indicados en azul (Figura 46A) dio lugar a la amplificación del fragmento pequeño en el subclon flipeado lo cual es indicativo de la escisión del casete de recuperación o rescate (Figura 46C). Efectivamente, el subclon es sensible a neomicina (Figura 46D). Por último, se generó el subclon flipeado de la línea sensible PA-1-pRAM272 (Figura 47A). Este también fue negativo para la PCR de NEO, sin embargo, por problemas técnicos no se puede mostrar el gel. No obstante, la realización de la PCR con los oligonucleótidos indicados en rojo y azul dio lugar a la amplificación del fragmento

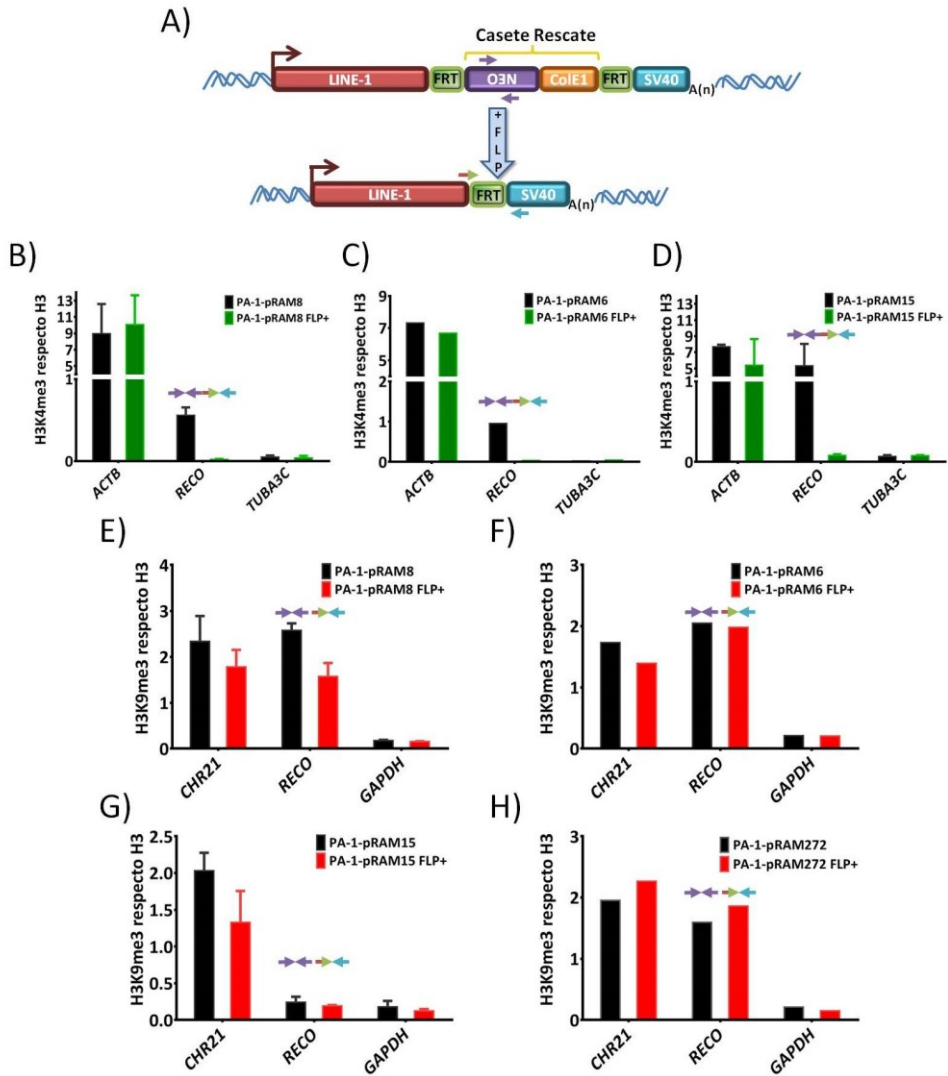
pequeño sólo en el subclon flipeado (Figura 47B). En este caso, no se puede comprobar la sensibilidad a neomicina porque se trata de una línea sensible. Además, en todos los casos, se secuenciaron los productos de PCR para confirmar que se produce la recombinación mediada por flipasa de una forma correcta.



**Figura 47: Generación del subclon flipeado en las líneas PA-1-pRAM272. A)** Esquema de la inserción antes y después de la escisión del casete. Las flechas indican sobre la inserción la posición de la pareja de oligonucleótidos utilizada. **B)** PCR usando los oligonucleótidos 2H sentido (Flecha roja) y SV40 REV FLP2 (Flecha azul). El tamaño del producto de PCR es 784 pares de bases. MPM significa marcador de peso molecular.

#### 4.2.6.1. Análisis de modificaciones de histonas sobre las inserciones de LINE-1 flipeadas

Una vez se generaron los subclones flipeados, mediante CHIP-qPCR se analizó si las modificaciones de histonas enriquecidas sobre las inserciones de LINE-1 se mantenían después del tratamiento con flipasa. La ausencia del casete de rescate o recuperación (en inglés llamado "recovery") en los subclones flipeados implica que para los experimentos de CHIP-qPCR se debe utilizar una pareja de oligonucleótidos alternativa. Esta pareja de oligos alternativa para los subclones flipeados está compuesta por un oligonucleótido Fw que hibrida en la posición 5950, que se corresponde con el final de la secuencia del retroelemento LINE- 1 y 4 cuatro nucleótidos de la secuencia FRT, denominado FLPRECO5



**Figura 48. Modificaciones de histonas enriquecidas sobre las inserciones artificiales de LINE-1 antes y después de la recombinación mediada por flipasa.** A) Esquema del aspecto de la inserción antes y después de la escisión del casete de rescate o “recovery” mediada por Flippasa. Las flechas indican las posiciones de las parejas de oligonucleótidos utilizadas para ChIP-qPCR en cada caso. B), C) y D) Gráfico de barras que muestra el experimento de ChIP-qPCR de H3K4me3 en las líneas clonales PA-1-pRAM8, PA-1-pRAM6 y PA-1-pRAM15 con sus correspondientes subclones flipeados respectivamente. ACTB es un control positivo mientras que TUBA3C es un control negativo. E), F), G), y H) Gráfico de barras que muestra el experimento de ChIP-qPCR de H3K9me3 en las líneas PA-1-pRAM8, PA-1-pRAM6, PA-1-pRAM15 y PA-1-pRAM272. Chr21 es un control positivo de mientras que GAPDH es un control negativo. En B), D), E), G) se muestran los valores medios de dos réplicas



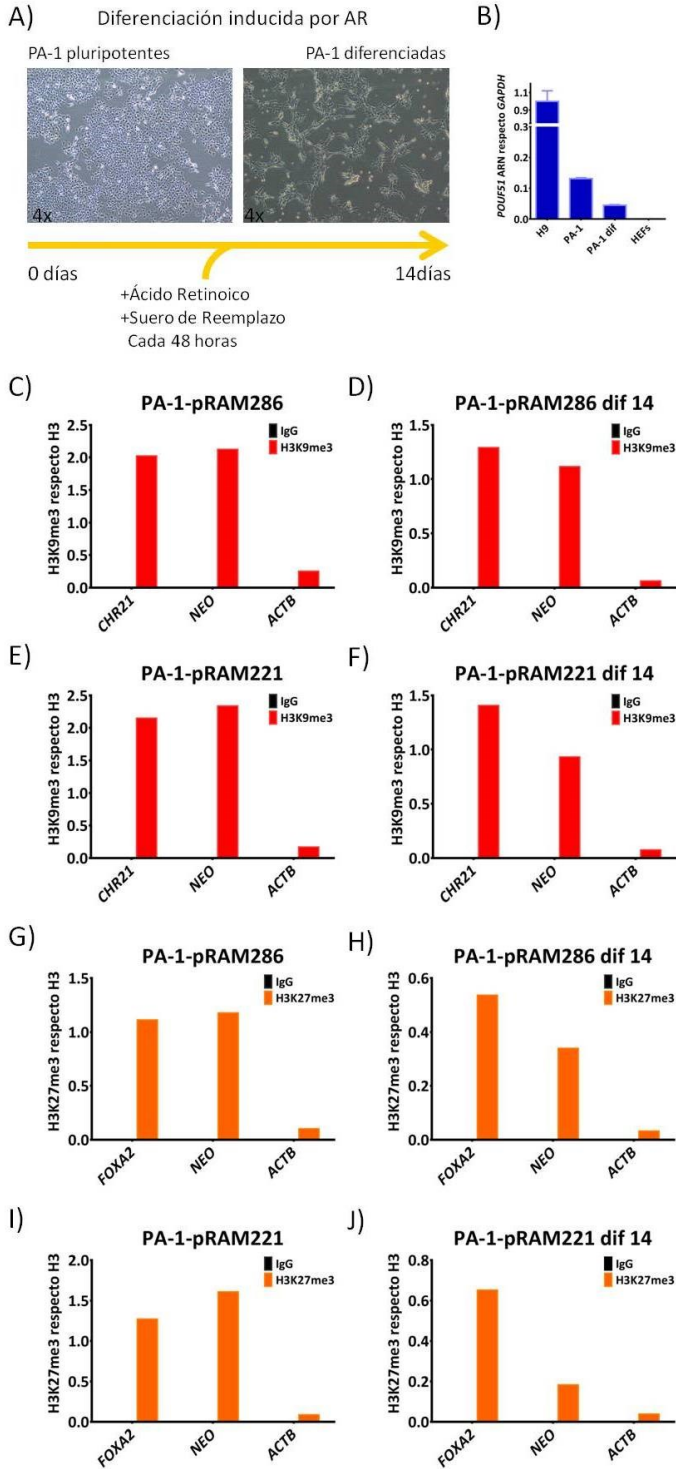
biológicas. Las barras de error muestran el rango de valores. En **C), F) y H)** se muestran los valores de un solo experimento de CHIP-qPCR.

(Figura 48A, flecha roja y verde) y un oligonucleótido RV que anilla en la señal de poliadenilación SV40 denominado SV40 REV FLP (Figura 48A, flecha azul). Esa pareja de oligonucleótidos será referida de aquí en adelante como SV40. Por tanto, utilizando la pareja de oligonucleótidos que anilla sobre el gen *NEO* (Figura 48A, flechas moradas) y la pareja de oligonucleótidos SV40 (Figura 48A, flechas rojo/verde y azul) se puede analizar simultáneamente las modificaciones de histonas enriquecidas sobre las inserciones de LINE-1 antes y después de la recombinación mediada por flipasa utilizando secuencias únicas. Debido a que el gen *NEO* ha sido escindido en los suclones FLP+, a partir de ahora, esa región se denominará indistintamente RECO (del inglés *recovery*) para poder representar conjuntamente los resultados de la línea clonal junto con el subclon flipado.

Por tanto, se analizó por CHIP-qPCR la región RECO antes y después de escindir el casete de recovery o rescate. Para ello se llevaron a cabo experimentos de CHIP-qPCR para H3K4me3 en las líneas PA-1-pRAM8, PA-1-pRAM6 y PA-1-pRAM15 y para H3K9me3 en las líneas PA-1-pRAM8, PA-1-pRAM6, PA-1-pRAM15 y PA-1-pRAM272 junto con los subclones flipados correspondientes. Se observa en la figura 48B, C, D que la eliminación del casete de rescate favorece la pérdida del enriquecimiento de H3K4me3 después de la recombinación mediada por flipasa en los tres subclones flipados estudiados. Por otro lado, cuando se analiza el enriquecimiento de H3K9me3 antes y después de la eliminación del casete, se observa, como el enriquecimiento de H3K9me3 se mantiene sobre la inserción al mismo nivel en todos los casos (Figura 48E, F, G, H). Por tanto, la eliminación del casete de rescate no altera el estatus de H3K9me3 mientras que favorece que desaparezca el enriquecimiento de H3K4me3 sobre la inserción de LINE-1 en los subclones flipados. Estos datos sugieren que el reclutamiento de H3K9me3 es dependiente de la secuencia de LINE-1 y que, por tanto, los cambios de H3K9me3 generados en el genoma adyacente son inducidos por la inserción *de novo* de LINE-1.

#### 4.2.7. Efecto de la diferenciación en la regulación e impacto epigenético de las inserciones artificiales de novo de LINE-1.

A día de hoy se conoce que la mayoría de las inserciones *de novo* de LINE-1 se acumulan durante el desarrollo embrionario temprano. Según los datos mostrados, una nueva inserción de LINE-1 durante el desarrollo embrionario temprano podría generar cambios epigenéticos en el genoma adyacente. En función de la localización de la inserción de LINE-1, estos cambios podrían tener un impacto funcional en fases tempranas de desarrollo embrionario o bien en células diferenciadas en caso de que dichos cambios se

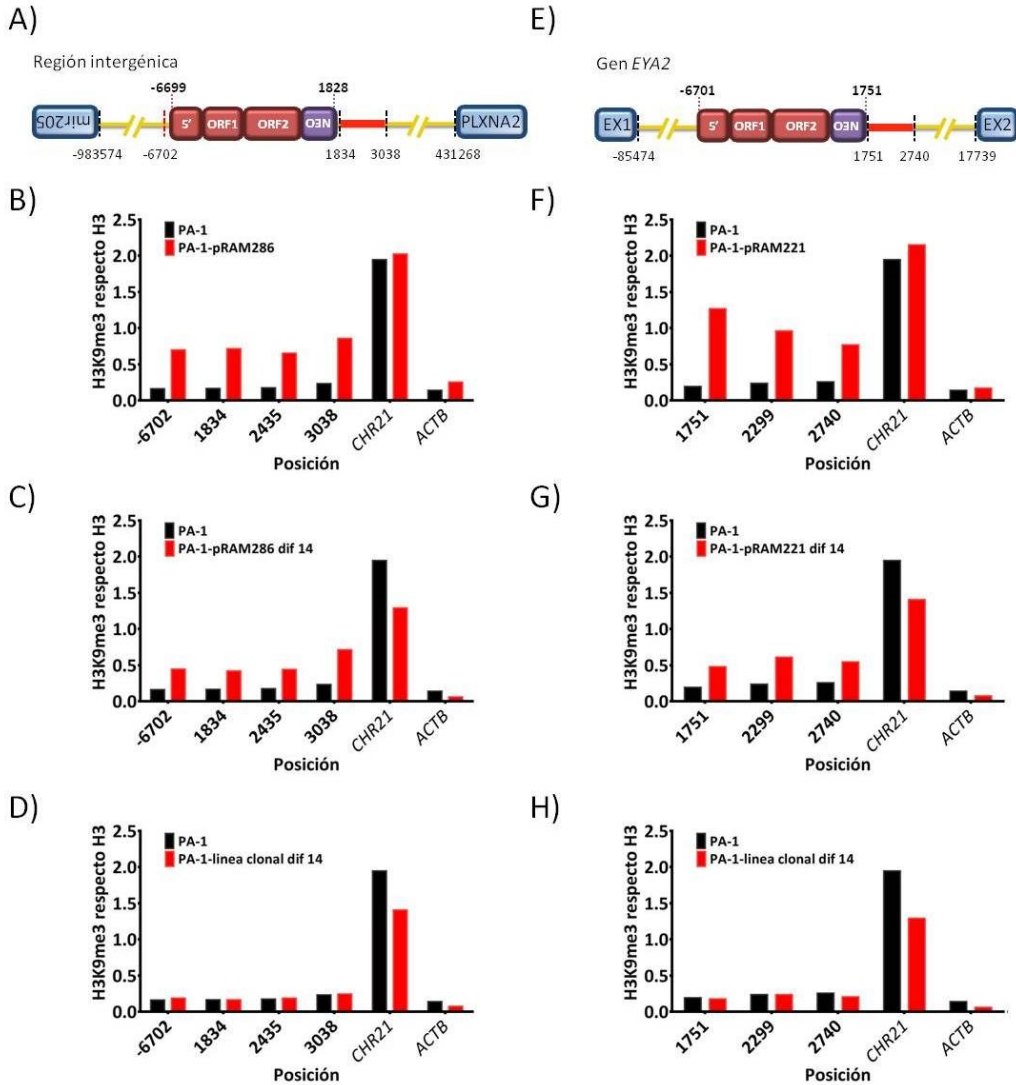


**Figura 49: Regulación epigenética de las inserciones de LINE-1 durante la diferenciación celular.** A) Esquema del proceso de diferenciación. Las fotos muestran la morfología adquirida por las células PA-1 al final del proceso de diferenciación. AR significa ácido retinoico. B) Disminución de la expresión de POU5F1 después del proceso de diferenciación. H9 es un control positivo y MEFs (del inglés Mouse Embryonic Fibroblasts) es un control negativo. C), E) Enriquecimiento de H3K9me3 sobre las inserciones de las líneas PA-1-pRAM286 y PA-1-pRA221 respectivamente. Se muestra un experimento de CHIP-qPCR. D), F) Enriquecimiento de H3K9me3 sobre las inserciones de las líneas PA-1-pRAM286 y PA-1-pRA221 después de 14 días de diferenciación respectivamente. Se muestra un experimento de CHIP-qPCR. G), I) Enriquecimiento de H3K27me3 sobre las inserciones de las líneas PA-1-pRAM286 y PA-1-pRA221 respectivamente. Se muestra un experimento de CHIP-qPCR. H), J) Enriquecimiento de H3K27me3 sobre las inserciones de las líneas PA-1-pRAM286 y PA-1-pRA221 después de 14 días de diferenciación respectivamente. Se muestra un experimento de CHIP-qPCR. En C), D), E), F) CHR21 es un control positivo y ACTB es un control negativo. En G), H), I), J) FOXA2 es control positivo y ACTB es un control negativo.

mantuvieran a lo largo del proceso de diferenciación. Por tanto, para validar esta hipótesis se analizó el efecto del proceso de diferenciación sobre la regulación epigenética de las inserciones *de novo* de LINE-1 y sobre los cambios epigenéticos generados por éstas. Para ello, se aprovechó que las células PA-1 se pueden diferenciar rápida y fácilmente hacia linaje ectodérmico [60] mediante la adición de ácido retinoico y suero de reemplazo al medio de cultivo durante 14 días provocando la bajada de expresión de marcadores de pluripotencia como *POUF51* (Figura 49A, B). Aprovechando este sistema, se diferenciaron las líneas PA-1-pRAM221 y PA-1-pRAM286 cuyas inserciones produjeron cambios de enriquecimiento de H3K9me3 en el genoma adyacente (Figura 38B, E). De este modo, pasados los 14 días se analizó por ChIP-qPCR, el nivel de enriquecimiento de H3K9me3 y H3K27me3 sobre dichas inserciones así como el nivel de enriquecimiento de H3K9me3 en el genoma adyacente.

Como se observa en la figura 49 D, F las inserciones de LINE-1 se mantienen enriquecidas en H3K9me3 después del proceso de diferenciación. De hecho, la inserción de la línea PA-1-pRAM286 presenta un valor de enriquecimiento de H3K9me3 similar al control positivo (Figura 49D) al igual que ocurría en estado pluripotente (Figura 49C). En el caso de la línea PA-1-pRAM221, la inserción presenta un valor de enriquecimiento de H3K9me3 algo menor comparado con el control positivo (Figura 49F) a diferencia de lo que ocurría en estado pluripotente (Figura 49E). Además, las inserciones de LINE-1 parecen mantenerse enriquecidas en H3K27me3 durante la diferenciación celular (Figura 49H, J). Sin embargo, este caso es menos claro ya que la inserción de la línea PA-1-pRAM286 presenta un enriquecimiento intermedio de H3K27me3 (Figura 49H) comparado con el estado pluripotente (Figura 49G) mientras que la inserción de la línea PA-1-pRAM221 presenta un enriquecimiento bajo en estado diferenciado (Figura 49J) en comparación con el estado pluripotente (Figura 49I). Estos datos indican, de forma preliminar, que las inserciones *de novo* de LINE-1 se mantienen enriquecidas en H3K9me3 y en menor medida en H3K27me3 durante la diferenciación celular.

Posteriormente, se analizó si los cambios de H3K9me3 generados en el genoma adyacente se mantenían a lo largo del proceso de diferenciación. En primer lugar, se muestra que los cambios inducidos por la inserción de LINE-1 de la línea PA-1-pRAM286 en estado pluripotente (Figura 50B) se mantienen después de 14 días de diferenciación (Figura 50C). Seguidamente, se muestra que los cambios generados por la inserción de la línea PA-1-pRAM221 en estado pluripotente (Figura 50F) también se mantienen después del proceso de diferenciación (Figura 50G). Cabe destacar que los enriquecimientos detectados en estado diferenciado son menores (Figura 50C, G) que los enriquecimientos detectados en estado pluripotente (Figura 50B, F). Sin embargo, estos enriquecimientos



**Figura 50: Mantenimiento de los cambios de enriquecimiento en el genoma adyacente de las líneas PA-1-pRAM221 y PA-1-pRAM286 en estado diferenciado. A), B)** Esquema de la localización de las inserciones de las líneas PA-1-pRAM286 y PA-1-pRAM221 respectivamente. Se muestran las posiciones analizadas en rojo. **B), F)** Gráfico de barras que muestra un ChIP-qPCR de H3K9me3 en el genoma adyacente en las líneas PA-1-pRAM286 y PA-1-pRAM221. **C), G)** Gráfico de barras que muestra un ChIP-qPCR de H3K9me3 en el genoma adyacente en las líneas PA-1-pRAM286 y PA-1-pRAM221 después de 14 días de diferenciación. **D), H)** Gráfico de barras que muestra un ChIP-qPCR de H3K9me3 en el genoma adyacente en la línea parental PA-1 y en una línea clonal de PA-1 después de 14 días de diferenciación.

son superiores al control negativo así como a los valores de enriquecimiento detectados en las mismas posiciones en la línea parental PA-1 wt (Figura 50C, G). Además, en ambos casos, los enriquecimientos detectados en el genoma adyacente no son producto del proceso de diferenciación. De hecho, una línea clonal de PA-1 diferenciada también durante 14 días presenta los mismos valores de enriquecimiento de H3K9me3 en las posiciones analizadas que la línea parental PA-1 wt (Figura 50D, H). Estos datos, aunque preliminares, sugieren que los cambios epigenéticos en el genoma adyacente generados por una inserción *de novo* de LINE-1 que se integre en etapas tempranas del desarrollo embrionario podrían tener un impacto funcional en etapas más tardías del proceso de desarrollo cuando se produce la especificación de linaje celular.

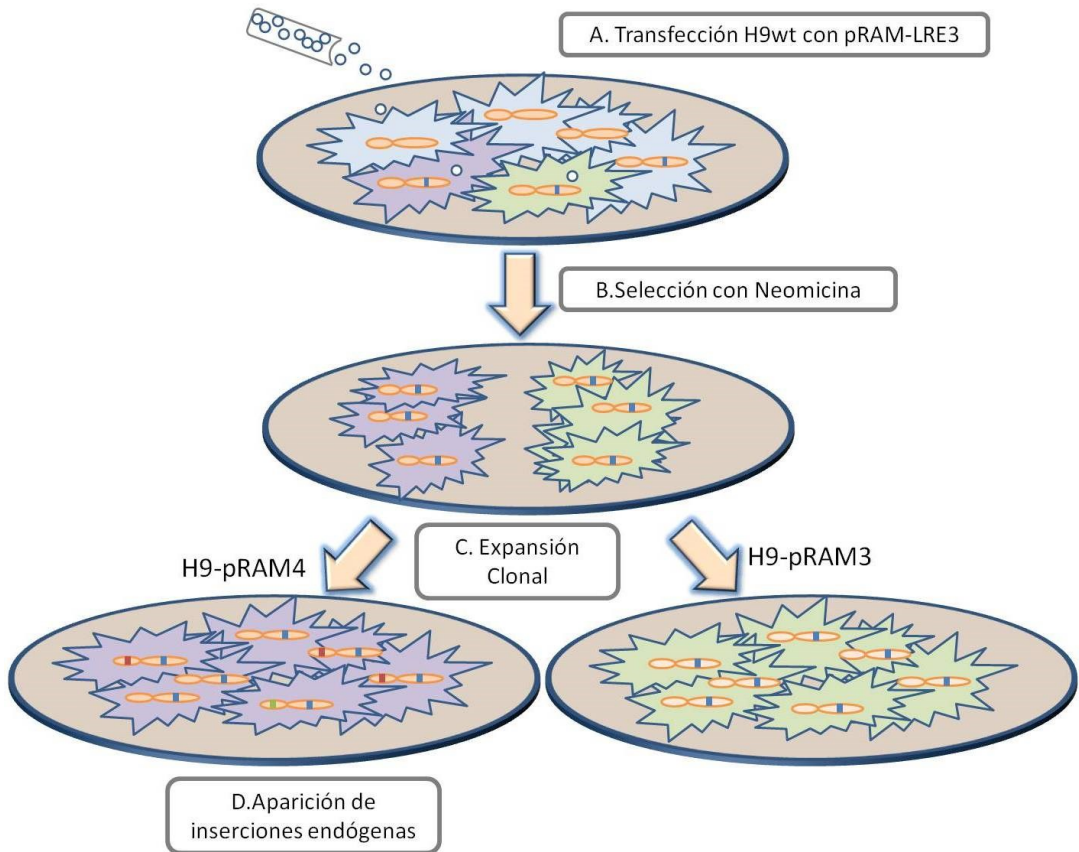
#### 4.2.8. Regulación epigenética de inserciones endógenas de novo de LINE-1 en células H9

Durante la realización de esta tesis doctoral, en nuestro laboratorio, un compañero postdoctoral descubrió mediante RC-seq (del inglés Retrotransposon Capture Sequencing) [82] que la retrotransposición de elementos LINE-1 endógenos ocurre en células madre embrionarias humanas en cultivo. De hecho, nuevas inserciones endógenas de LINE-1 se acumulan en el genoma de estas células cuando se mantienen en cultivo durante varios pases (Muñoz-Lopez et *al.*, en preparación). Afortunadamente, dichas inserciones fueron detectadas en las líneas H9-pRAM3 y H9-pRAM4, las cuales fueron generadas para el objetivo de esta tesis y que ya han sido introducidas a lo largo de la misma. Por tanto, se aprovechó este hecho para extender el tipo de análisis realizado sobre las inserciones artificiales de LINE-1, a las inserciones endógenas encontradas en el estudio mencionado (Muñoz-Lopez M. et *al.*, en preparación). Las inserciones encontradas en H9-pRAM4 y H9-pRAM3 junto con sus características se resumen en la tabla 10.

##### *4.2.8.1. Determinación del porcentaje celular con las inserciones endógenas identificadas*

Como se observa en la figura 51, a lo largo del proceso de expansión clonal para la generación de líneas clonales resistentes en células H9 aparecieron también estas inserciones endógenas propias de H9-pRAM4 y H9-pRAM3 (Tabla 10). En primer lugar, se comprobó el número de células que contienen esta inserción mediante PCR cuantitativa. Es bastante probable que la inserción 1-927 se encuentre en el 100% de las células ya que está presente en la línea H9-pRAM4 y en la línea H9-pRAM3 lo que indica que se produjo en el cultivo de original de células H9 (Figura 51A, barras azules). No obstante, se determinó el nº de copias relativo de dicha inserción para confirmar esta asunción. Para ello, el nº de copias relativo fue obtenido mediante qPCR normalizando respecto al nº de

copias del gen de la subunidad 25 de la ribonucleasa P que se encuentra en 1 copia por genoma diploide [73]. Por tanto, el nº de copias relativo será equivalente al nº de células donde se encuentre la inserción. Como se observa en la figura 52A, el nº de copias de la inserción 1-927 es igual al nº de copias del gen de referencia lo cual indica que esta inserción se encuentra en el 100% de las células.



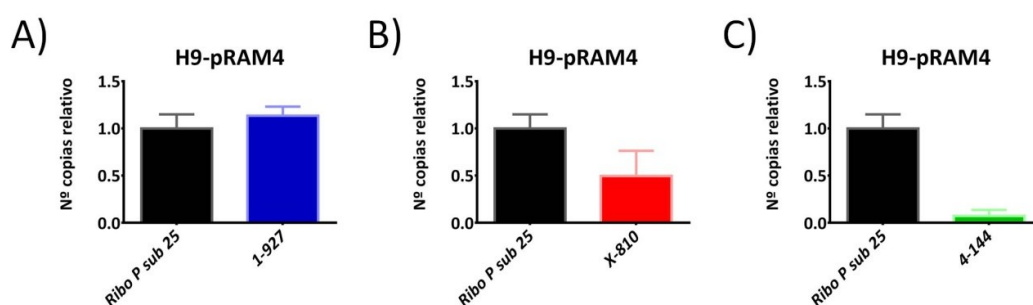
**Figura 51: Aparición inserciones endógenas de LINE-1 en las líneas clonales H9-pRAM4 y H9-pRAM3.** **A)** Células H9 wt fueron transfectadas con el vector pRAM-LRE3 para generar líneas clonales resistentes. Algunas de las células del cultivo ya contenían una nueva inserción de LINE-1 endógena: 1-927 (líneas gruesas azules). **B)** Las células transfectadas y aquellas donde ocurrió la retrotransposición desde el vector pRAM-LRE3 fueron las mismas donde había aparecido la inserción 1-927 (líneas gruesas azules). **C)** Estas colonias se expandieron separadamente generándose dos líneas clonales independientes (H9-pRAM4 y H9-pRAM3). **D)** Durante el proceso descrito en C) para H9-pRAM4 se produce la aparición de las inserciones X-810 (líneas gruesas rojas) y 4-144 (línea gruesa verde) propias de esta línea clonal.

Por otro lado, las inserciones X-810 y 4-144, sólo se encuentran en la línea clonal H9-pRAM4 (Figura 51D, barras rojas y barra verde respectivamente). Esto indica que se han producido durante el proceso de expansión clonal o durante la selección con neomicina (Figura 51B, C). Por tanto, no se encontrarían en el 100% de las células a menos que se hubieran producido en la primera célula que dio origen a la línea clonal. De la misma forma que para la inserción 1-927, se calculó en nº relativo de copias que equivale al nº de células donde se encuentran ambas inserciones. Como se esperaba, ninguna de las inserciones se encuentra en el 100% de las células (Figura 52B,C).

Inserción	Punto Truncación	DSI <sup>1</sup>	Longitud Poli A	Deleción	<sup>4</sup> Localización	Intrón/ Exón	Sentido/ Antisentido <sup>2</sup>
1-927	Completa	16	35	No	Crom 1/Gen RPAP2	Intrón 7/13	Antisentido
X-810	Completa	11	77	No	Crom X/ Intergénica		
4-144	Completa	14	175	No	Crom 4/ Intergénica		

**Tabla 10:** Inserciones endógenas nuevas presentes en la línea H9-pRAM4.<sup>1</sup>Duplicaciones del sitio de inserción. <sup>2</sup> Crom significa cromosoma. <sup>4</sup>Orientación de la inserción respecto al gen.

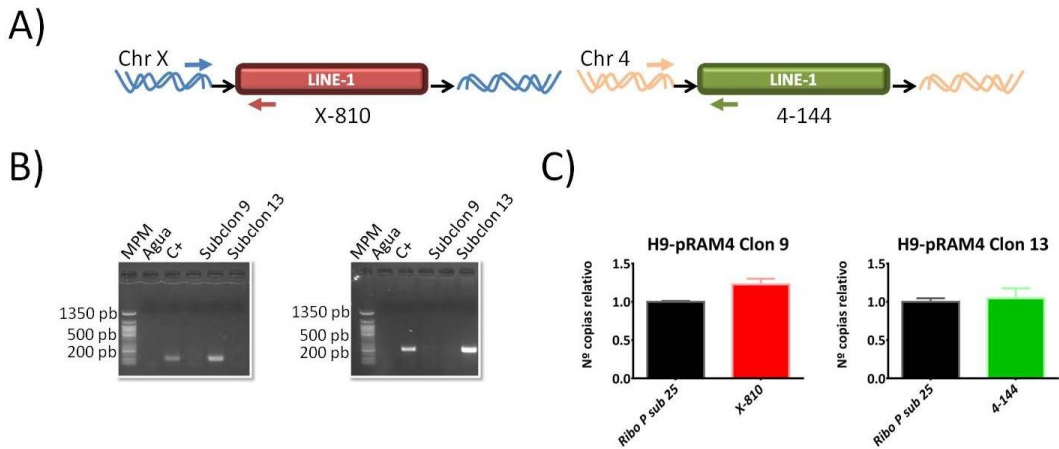
Según estos datos, el estudio epigenético se puede llevar a cabo sobre la inserción 1-927 mientras que para las otras dos inserciones existe la posibilidad de que la dilución de la inserción enmascare los enriquecimientos existentes sobre las inserciones. Por esta razón es necesario enriquecerlas en la población celular. Para ello se realizó una dilución clonal de la línea H9-pRAM4. Posteriormente, se realizó la búsqueda de subclones positivos por PCR convencional utilizando oligonucleótidos de genotipado específicos para cada



**Figura 52:** Determinación del nº de copias relativo de las inserciones endógenas. **A)** PCR cuantitativa que muestra el nº de copias relativo para la inserción 1-927. **B)** PCR cuantitativa que muestra la determinación del nº de copias relativo para la inserción X-810. **C)** PCR cuantitativa que muestra la determinación del nº de copias para la inserción 4-144. En **A), B), C)** se muestra la media de dos réplicas biológicas. Las barras de error representan el rango de valores. En **A), B), C)** se normaliza respecto al gen de la subunidad 25 de la ribonucleasa P.

inserción (Figura 53A). De esta forma se consiguió encontrar subclones que sólo contenían una inserción o la otra (Figura 53B). Finalmente, mediante qPCR se confirmó que el proceso de dilución clonal, efectivamente, había favorecido el enriquecimiento de las inserciones X-810 y 4-144 ya que se encuentran ahora en el 100% de las células del subclon identificado (Figura 53C).

Por tanto, este proceso ha permitido disponer de 2 líneas clonales donde el 100% de las células de cada una presenta una de las 2 inserciones endógenas de novo de LINE-1.

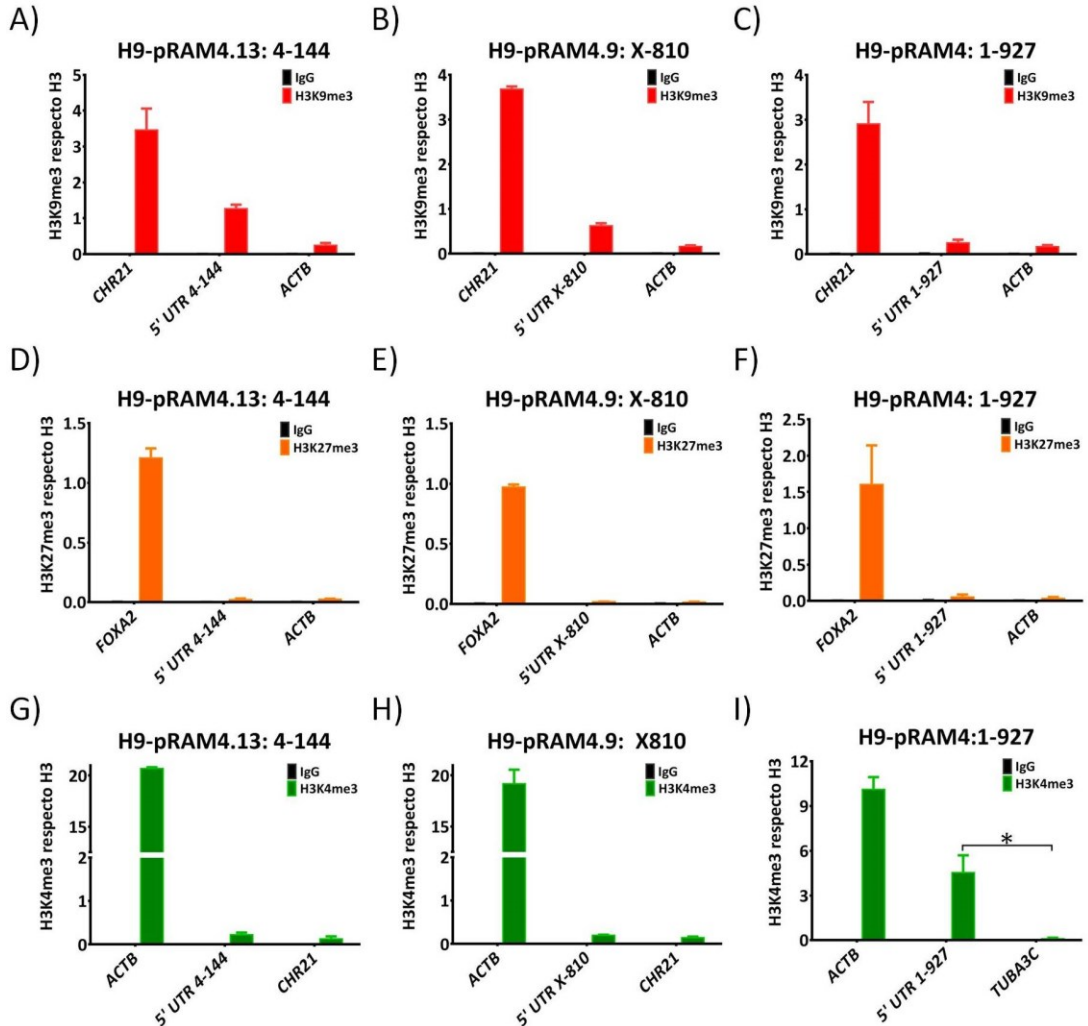


**Figura 53: Proceso de enriquecimiento de las inserciones X-810 y 4-144 en la población celular.** **A)** Oligonucleótidos de genotipado específicos para cada inserción endógena. **B)** Identificación de subclones positivos mediante PCR convencional. **C)** Cuantificación mediante qPCR usando los oligonucleótidos de genotipado. Se muestra la media de dos réplicas biológicas. Las barras de error muestran el rango de valores. Se normaliza respecto al gen de la subunidad 25 de la ribonucleasa P.

#### 4.2.8.2. Estudio epigenético de las inserciones endógenas de novo de LINE-1

En primer lugar se determinaron las modificaciones de histonas que se encuentran enriquecidas sobre las inserciones endógenas de LINE-1. Afortunadamente, las tres inserciones endógenas de las que se dispone para este estudio son de copia completa y, por tanto, tienen un promotor intacto. De esta forma, el análisis de regulación epigenética podía ser realizado directamente sobre el promotor de LINE-1. Para este análisis no se puede obviar la existencia de otras copias completas de LINE-1 en el genoma humano, de tal forma que se pueden producir contaminaciones durante la amplificación por qPCR. Por esta razón, se usa la intersección entre la inserción de LINE-1 y el sitio flanqueante del genoma como “sitio único” para diseñar oligonucleótidos de PCR cuantitativa. De esta forma, se diseña una pareja de oligonucleótidos posicionando uno fuera del promotor y otro dentro de éste. Teniendo en cuenta los resultados obtenidos en la primera parte del





**Figura 54: Determinación de las modificaciones de histonas enriquecidas sobre las inserciones endógenas de novo de LINE-1.** A-C) Gráfico de barras que muestra el enriquecimiento de H3K9me3 sobre los promotores de las tres inserciones endógenas (4-144; X-810; 1-927). CHR21 es un control positivo mientras que ACTB es un control negativo. D-F) Gráfico de barras que muestra el enriquecimiento de H3K27me3 sobre los promotores de las tres inserciones endógenas (4-144; X-810; 1-927). FOXA2 es un control positivo mientras que ACTB es un control negativo. G-I) Gráfico de barras que muestra el enriquecimiento de H3K4me3 sobre los promotores de las tres inserciones endógenas (4-144; X-810; 1-927). ACTB es un control positivo mientras que CHR21 y TUBA3C son controles negativos. En A), B), D), E), G) y H) se muestra la media de dos experimentos de CHIP-qPCR. Las barras de error muestran el rango de valores. En C), F) e I) se muestra la media de tres experimentos de CHIP-qPCR con su desviación estándar. El p-valor obtenido en I) es 0,002686.

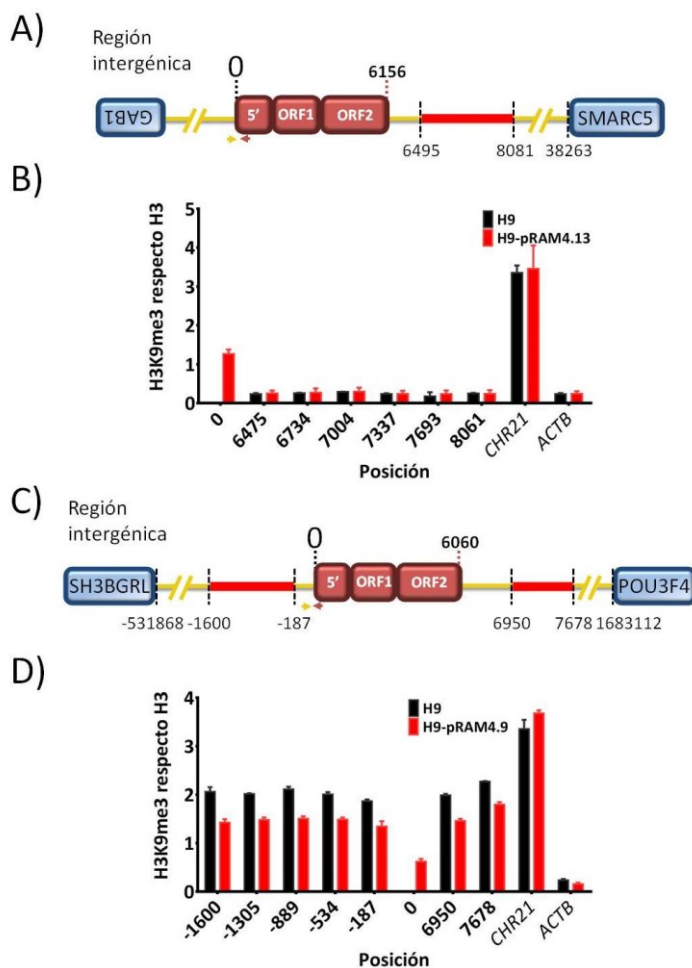
estudio donde se analizó la regulación epigenética de las nuevas inserciones artificiales del retroelemento LINE-1, se optó por dar prioridad a las modificaciones de histonas asociadas

con cromatina inactiva como son H3K9me3 y H3K27me3. Sin embargo, también se incluyó H3K4me3. A pesar de su desaparición sobre las inserciones de los subclones flipeados, no se puede descartar que tenga un papel en la regulación de las inserciones endógenas de LINE-1. De hecho, H3K4me3 aparece enriquecida en las cercanías del promotor de elementos LINE-1 endógenos transcripcionalmente activos en el genoma humano [71].

Los ChIPs-qPCR para H3K9me3, H3K27me3 y H3K4me3 mostraron que el promotor de las inserciones 4-144 y X-810 se encuentran enriquecidas en H3K9me3 (Figura 54A, B), siendo el enriquecimiento mayor para la inserción 4-144 (Figura 54A). Por otro lado, la inserción 1-927 mostró un enriquecimiento marginal de H3K9me3 (Figura 54C). Inesperadamente, ninguna de las tres inserciones estudiadas presenta enriquecimiento alguno de H3K27me3. De hecho, los valores de enriquecimiento de H3K27me3 son similares a los valores medios de los controles negativos en las tres inserciones estudiadas (Figura 54D, E, F). Sorprendentemente, la inserción 1-927 presenta un elevado enriquecimiento de H3K4me3 sobre su promotor (Figura 54I). De hecho, este enriquecimiento es significativo respecto a los controles negativos. Sin embargo, las otras dos inserciones X-810 y 4-144 presentan un enriquecimiento similar al valor medio del control negativo (Figura 54G, H).

Seguidamente, se llevó a cabo el análisis de los posibles cambios de enriquecimiento de las modificaciones de histonas que pueden ser inducidos por las inserciones *de novo* de LINE-1. Para ello se llevaron a cabo, experimentos de ChIP-qPCR de H3K9me3 para las inserciones 4-144 y X-810 y de H3K4me3 para la inserción 1-927. Teniendo en cuenta las distancias máximas que alcanzaron los cambios de enriquecimiento en las líneas clonales con inserciones artificiales, se decidió diseñar parejas de oligonucleótidos cada 200-300 pares de bases cubriendo una distancia total de 1-1,5kb. Para estos análisis, hay que tener en cuenta que el punto 0 ha cambiado de localización. Ahora se toma como punto 0, el amplicón de qPCR utilizado para medir los enriquecimientos sobre el promotor de LINE-1. En este caso, se considera el nucleótido -1 y +1, al primer nucleótido situado aguas arriba y aguas abajo de dicho amplicón respectivamente. En todos los casos, se utilizó la línea H9 wt como control.

En primer lugar, se analizaron los posibles cambios de H3K9me3 en el genoma adyacente generados por la inserción 4-144. Esta inserción tiene el hándicap de que no se pueden diseñar oligonucleótidos aguas arriba de la misma ya que se encuentra integrada dentro de una inserción endógena de LINE-1. Por tanto, tan solo se puede analizar el genoma adyacente aguas abajo de la inserción (Figura 55A). En este caso no se detectaron cambios en la composición de H3K9me3 en las posiciones analizadas aguas abajo de la inserción (Figura 55B).

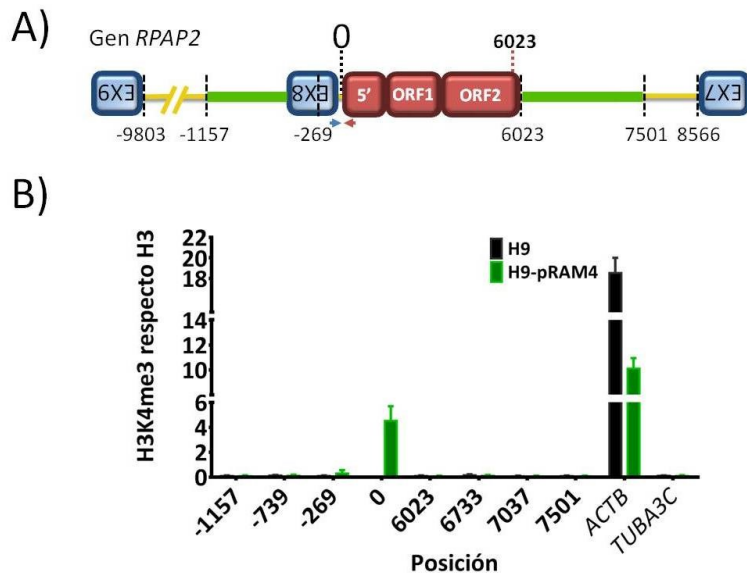


**Figura 55. Ausencia de cambios de enriquecimiento de H3K9me3 en el genoma adyacente a las inserciones endógenas 4-144 y X-810. A), C)** Esquema que muestra la localización de las inserciones endógenas 4-144 y X-810 respectivamente. Las flechas de color amarillo y rojo indican la posición del punto 0. **B)** Gráfico de barras que muestra los resultados de ChIP-qPCR de H3K9me3 en el genoma adyacente a la inserción 4-144 en la línea clonal H9-pRAM 4.13 y en la línea parental H9. Se muestran los valores medios de dos réplicas biológicas. Las barras de error muestran el rango de valores. **D)** Gráfico de barras que muestra los resultados de ChIP-qPCR de H3K9me3 en el genoma adyacente a la inserción X-810 en la línea clonal H9-pRAM 4.9 y en la línea parental H9. Se muestran los valores medios de dos réplicas biológicas. Las barras de error muestran el rango de valores. En **B)** y **C)** CHR21 es un control positivo mientras que ACTB es un control negativo.

En segundo lugar, se analizaron los posibles cambios de H3K9me3 en el genoma adyacente generados por la inserción X-810. Esta inserción se encuentra en una zona altamente heterocromatinizada debido a los valores de enriquecimiento de H3K9me3 que se observan en la línea control (Figura 55D, barras negras). Como se observa en la gráfica,

esta inserción no genera cambios en el enriquecimiento de H3K9me3 en el genoma adyacente (Figura 55D). Curiosamente, se observa en todas las posiciones una disminución del enriquecimiento de H3K9me3 en la línea clonal que contiene la inserción X-810 (Figura 55D, barras rojas).

Por último, se analizan los posibles cambios de enriquecimiento de H3K4me3 en el genoma adyacente a la inserción 1-927. La inserción 1-927 se encuentra localizada dentro de un gen denominado *RPAP2* específicamente dentro del intrón número 7 muy cerca del exón nº 8 (Figura 56A). Este gen, se expresa a niveles, que se pueden considerar relevantes biológicamente, ya que se expresa unas 5 veces menos que *NANOG* que es un gen bastante expresado en células madre embrionarias humanas (Ver Anexo IV) [223]. Según



**Figura 56: Ausencia de cambios de enriquecimiento de H3K4me3 en el genoma adyacente a la inserción endógena 1-927. A)** Esquema de localización de la inserción. Las flechas de color azul y rojo indican la posición del punto 0. **B)** Gráfico de barras que muestra el ChIP-qPCR de H3K4me3 en el genoma adyacente a la inserción 1-927 en la línea clonal H9-pRAM4 y en la línea parental H9. Se muestran los valores medios de tres experimentos independientes con su desviación estándar. ACTB es un control positivo mientras que TUBA3C es un control negativo.

se muestra en la figura 56B, no se detectan cambios en el enriquecimiento de H3K4me3 en las posiciones analizadas. Cabe destacar, que es una región desprovista de H3K4me3 (Figura 56B, barras negras).

En resumen, todos estos datos conjuntamente sugieren que las inserciones endógenas *de novo* de LINE-1 se regulan después de su integración mediante H3K9me3 y H3K4me3. Según las inserciones endógenas de novo de LINE-1 analizadas, éstas no producen cambios en el enriquecimiento de H3K9me3 en el genoma adyacente. Sin embargo, este estudio muestra por primera vez que una inserción endógena de LINE-1 de copia completa puede presentarse enriquecida en H3K4me3 después de su integración en el genoma humano. Esto significa que podría estar transcripcionalmente activa y efectuar un nuevo ciclo de retrotransposición.



4.2.9. ANEXO IV: INSERCIONES ARTIFICIALES Y ENDÓGENAS DE LINE-1





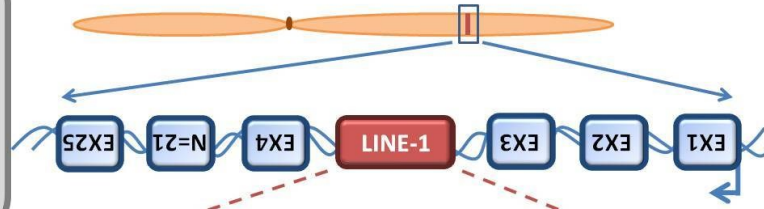
En este anexo incluye información acerca de:

1. Localización de la inserción: se muestra su posición en el cromosoma así como su posición respecto a los exones y los intrones del gen en caso de que la inserción sea génica.
2. Esquema de la inserción donde se puede apreciar las principales características de la inserción.
3. Para inserciones génicas, se muestra el nivel de expresión del gen mediante una RT-qPCR en las células wt (PA-1 o H9).
4. Esquema de la inserción donde se representan las distancias respecto a los exones u otros genes en función de su localización siempre que su cálculo fuera posible. Las flechas discontinuas se corresponden con la posición del último par de oligonucleótidos diseñados en 5' y en 3' para los experimentos de ChIP-qPCR. Se muestra siempre la hebra sentido.
5. Entrada del UCSC Genome Browser con distintas pistas donde se pueden ver las modificaciones de histonas asociadas al promotor del gen (para las inserciones génicas), a la zona analizada por ChIP-qPCR en cada línea así como los niveles de expresión del gen en cuestión. Las líneas discontinuas negras marcan aproximadamente la región analizada por ChIP-qPCR que se corresponde con la región delimitada por las flechas discontinuas en el esquema de localización de la inserción. Las pistas mostradas contienen la siguiente información:
  - a. ARNm-seq de células H1 (células madre embrionarias humanas).
  - b. Sitios de hipersensibilidad a DNAasa I: señala enhancers y promotores.
  - c. Modificaciones de histonas
    - i. H3K27me3: modificación asociada a represión génica [157].
    - ii. H3k36me3: modificación asociada a transcripción activa [137], [157].
    - iii. H3K4me1: modificación asociada a enhancers [152], [153].
    - iv. H3K4me3: modificación asociada fundamentalmente a promotores activos [136], [137].
    - v. H3k9ac: modificación asociada fundamentalmente a promotores activos [151].
    - vi. H3K9me3: modificación asociada a represión génica [183], [184].
    - vii. H3K27ac: modificación asociada a enhancers activos [152], [153].
    - viii. H4K5ac y H4K8ac: modificaciones asociadas con promotores activos [140]

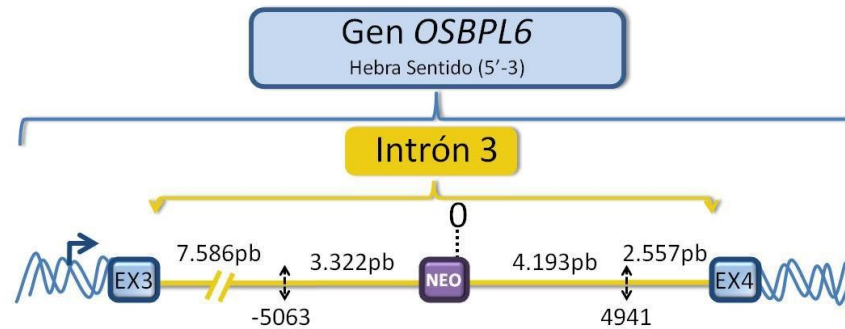
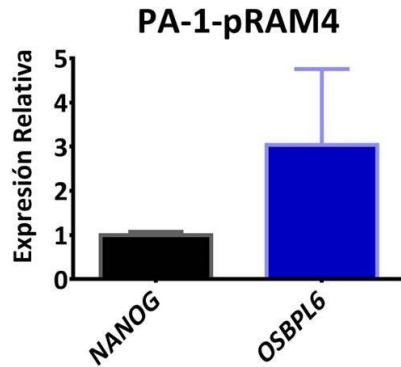
# PA-1-pRAM4

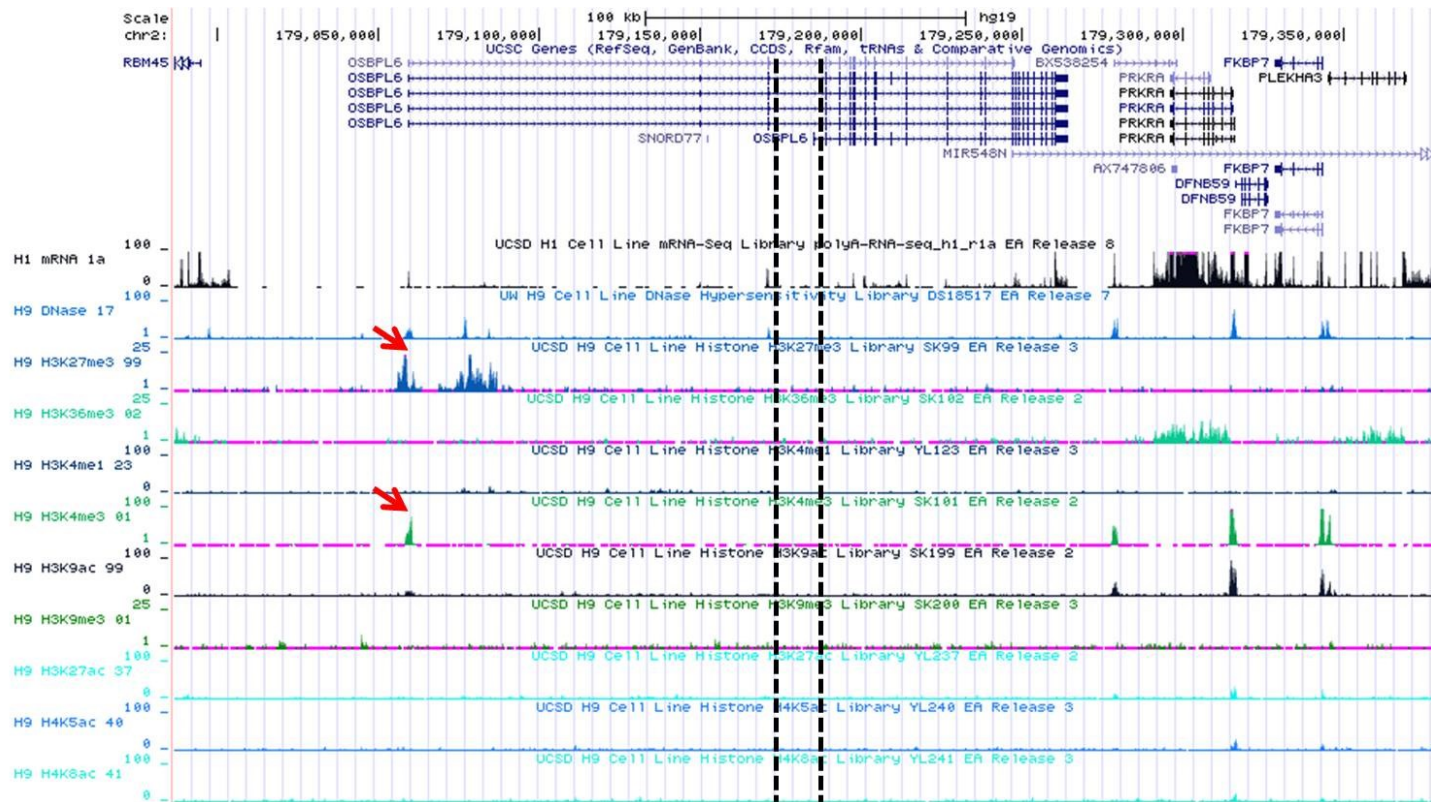
- Gen *OSBPL6*
- Hebra antisentido
- Truncado en 5963
- Cola de poli A 38
- DSI 16 pb
- ADN "relleno" 8 pb

Chrm2 (q31.2)



***OSBPL6***: Este gen pertenece a la familia de receptores intracelulares de lípidos que unen oxysterol (*OSBP*). Expresión mayoritaria en cerebro, piel y músculo.

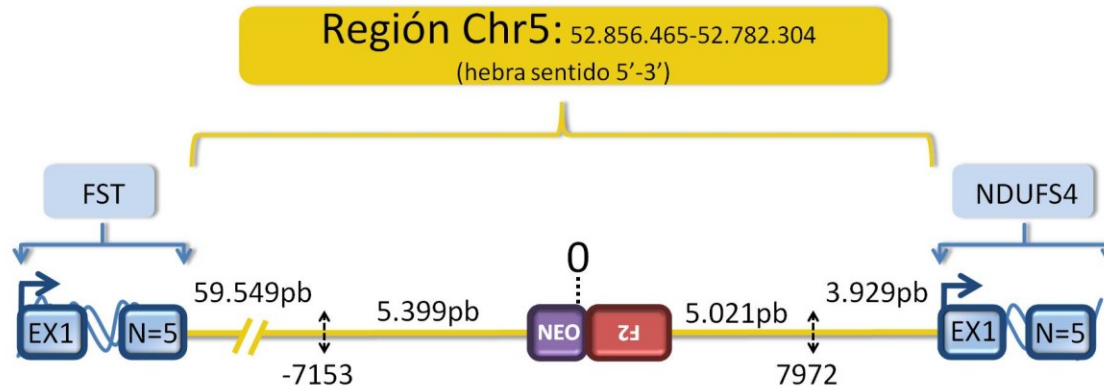
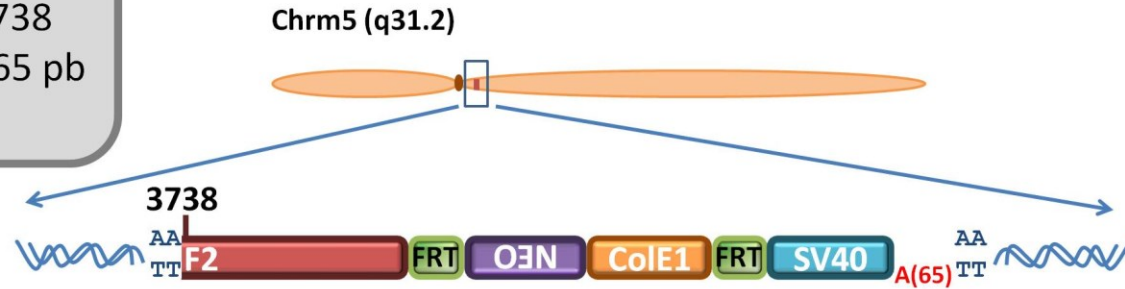


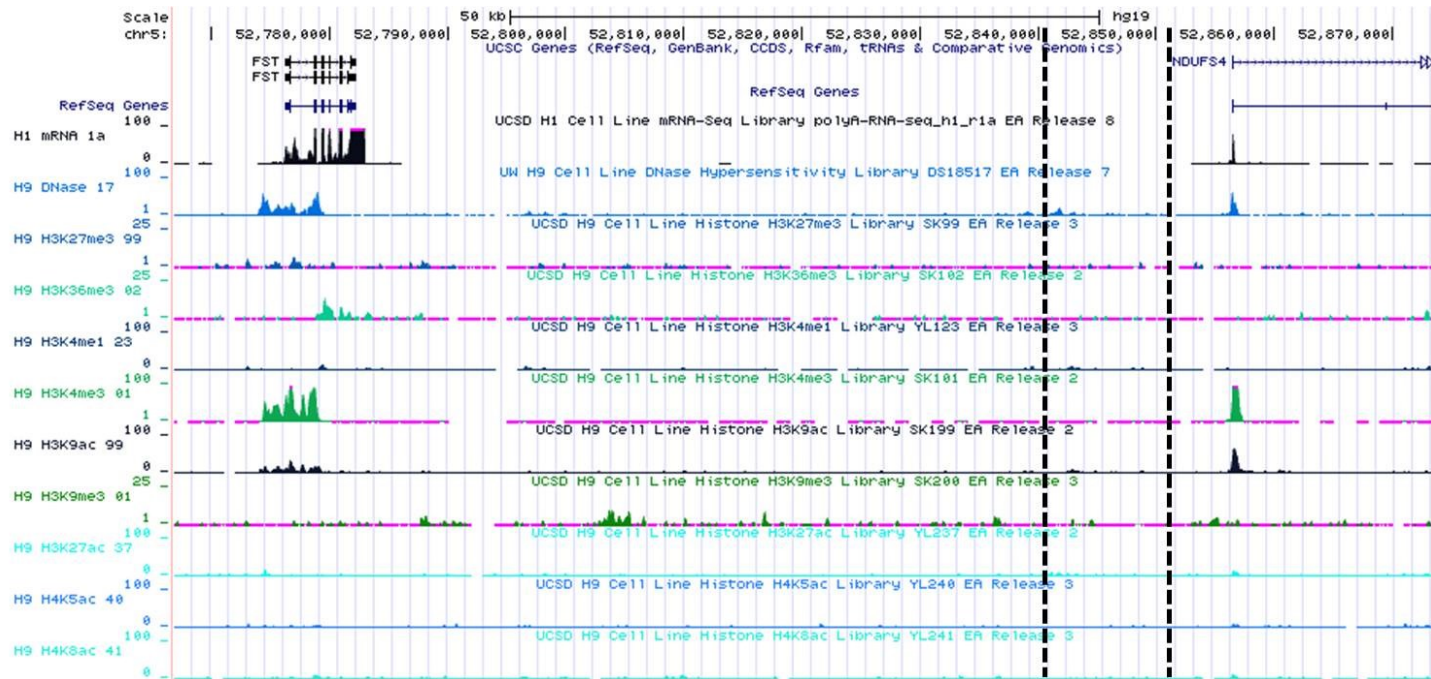


El gen donde se encuentra integrada esta inserción de LINE-1 es el *OSBPL6*. Este gen presenta H3K4me3 y H3K27me3 sobre su promotor de tal forma que es un gen bivalente (flechas rojas). Por ello, este gen se debe expresar poco o muy poco. De hecho, los datos de ARNm-seq en células H1 demuestran que efectivamente es un gen poco expresado ya que se detectan pequeños picos sobre alguno de los exones. De hecho, existe una ausencia casi total de H3K36me3 a lo largo del cuerpo del gen. Por tanto, estos datos concuerdan con el hecho de que este gen presente unos niveles de expresión aproximadamente entre 2 y 5 veces superiores a *NANOG* en células PA-1, cuya proteína no es detectable por western-blot. [224]. En general, el área analizada delimitada por las líneas discontinuas es una zona bastante desprovista de las modificaciones estudiadas.

- Intergénica
- Hebra antisentido
- Truncado en 3738
- Cola de poli A 65 pb
- DSI 2pb

# PA-1-pRAM26



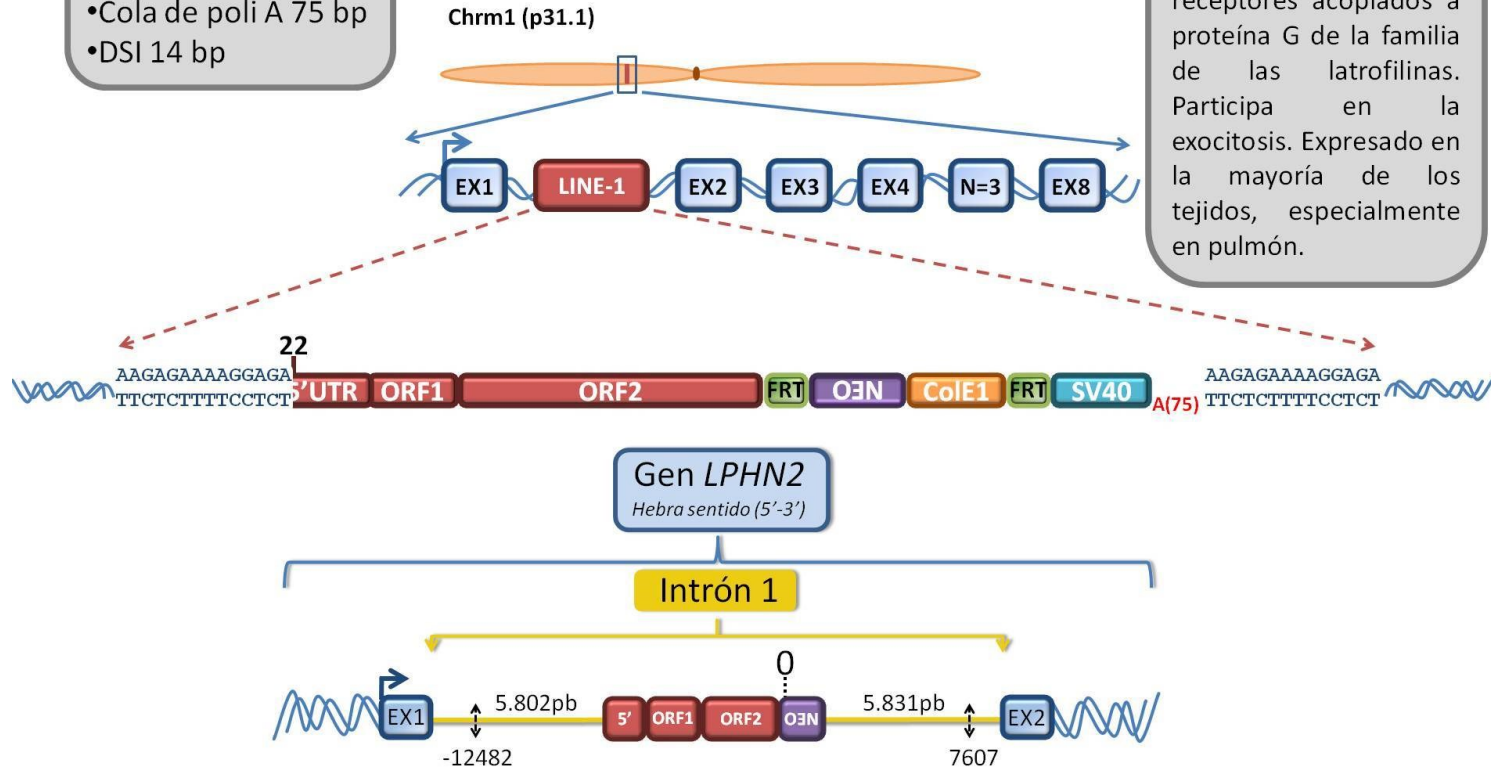


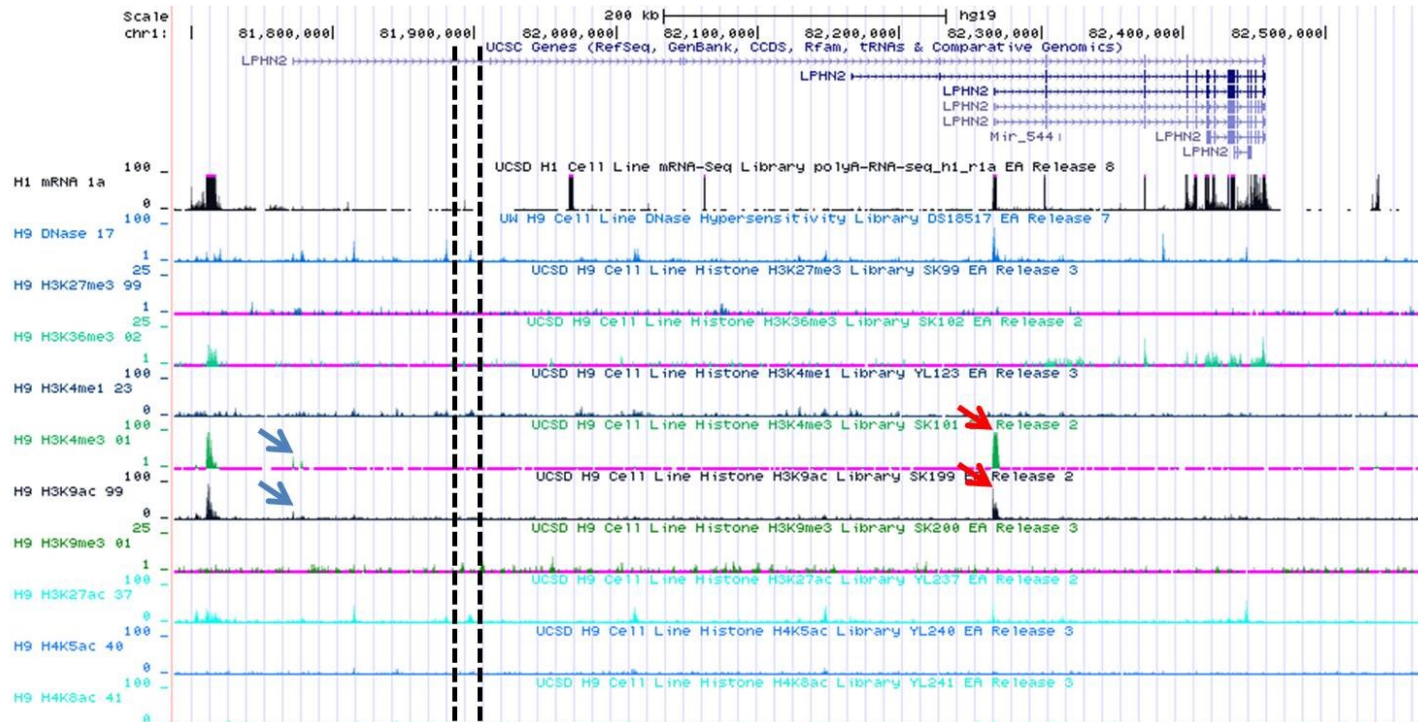
Como se puede observar, el área analizada por CHIP-qPCR en la línea PA-1-pRAM26, parece un lugar bastante despoblado de cualquiera de las modificaciones de histonas utilizadas en el estudio (zona delimitada por líneas discontinuas negras).

- Gen *LPHN2*
- Hebra sentido
- Truncado en 22
- Cola de poli A 75 bp
- DSI 14 bp

# PA-1-pRAM8

**LPHN2:** Este gen pertenece a la subfamilia de los receptores acoplados a proteína G de la familia de las latrofilinas. Participa en la excitación. Expresado en la mayoría de los tejidos, especialmente en pulmón.



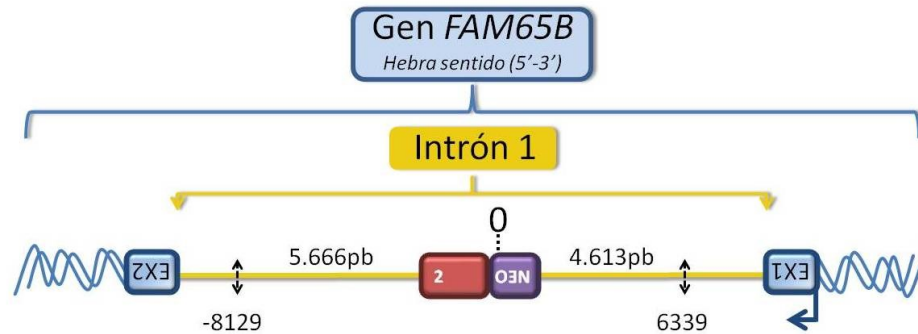
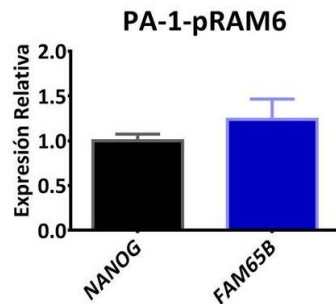
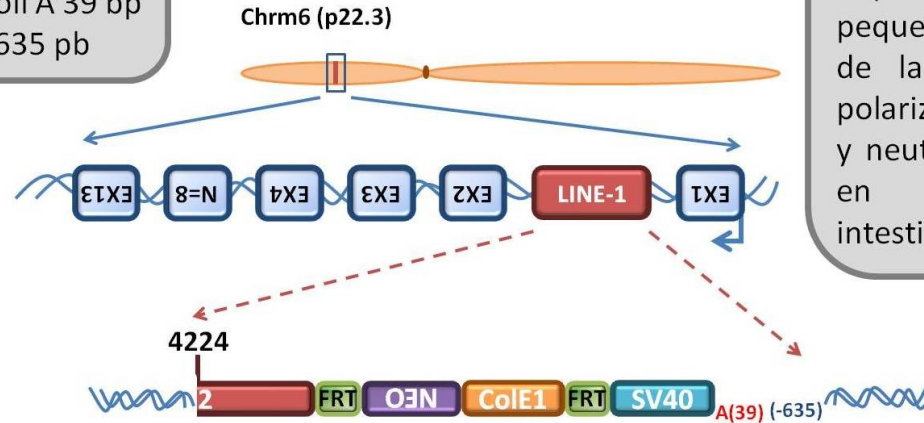


El gen donde se encuentra integrada esta inserción es el *LPHN2*. Según la información disponible en células H9, se observa que este gen debe estar activo, ya que en su promotor aparecen enriquecimientos de H3K4me3 y H3K9ac sobre todo para las isoformas más cortas (flechas rojas). Además, los picos del ARNm-seq de células H1 son de gran magnitud sobre todo sobre para los exones de las isoformas cortas (derecha). Sin embargo, la isoforma más larga donde se encuentra integrada la inserción de LINE-1 (izquierda), parece que se expresa menos ya que existen enriquecimientos mínimos de H3K4me3 y H3K9ac sobre su promotor (flechas azules). Existe, de hecho, algún pico procedente del ARNm-seq a lo largo de la misma. En la zona analizada por ChIP-qPCR (líneas discontinuas negras) es una zona poco enriquecida en cualquiera de las modificaciones estudiadas.

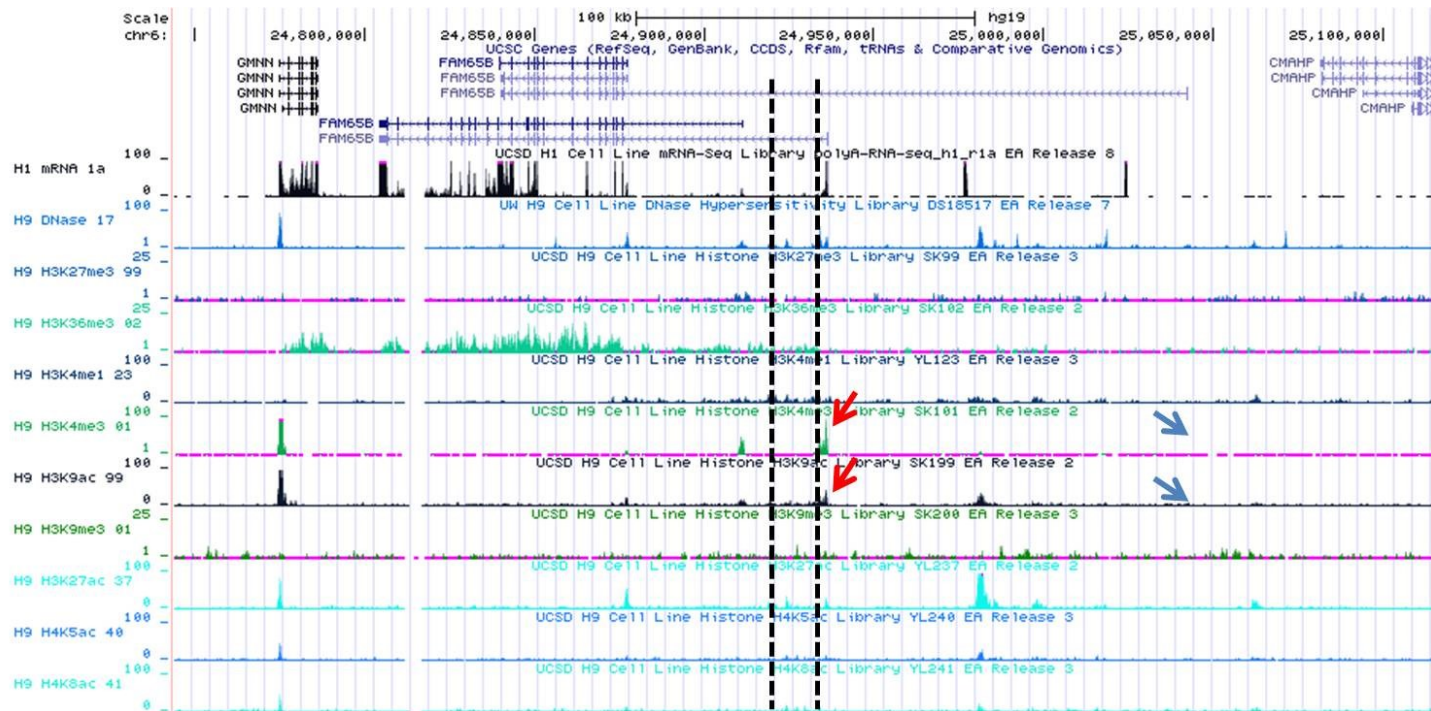
# PA-1-pRAM6

- Gen *FAM65B*
- Hebra sentido
- Truncado en 4224
- Cola de poli A 39 bp
- Delección 635 pb

**FAM65B:** Este gen codifica un inhibidor atípico de la protein G pequeña RhoA. El efecto de la inhibición es la polarización de células T y neutrófilos. Se expresa en sangre, útero, intestino delgado y bazo.





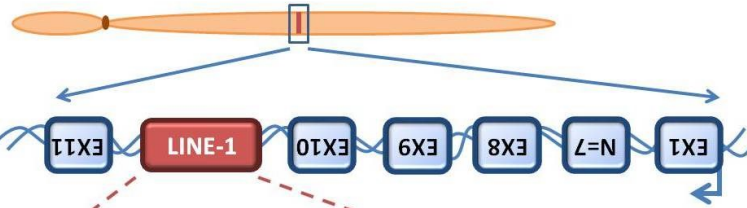


El gen donde se encuentra integrada la inserción de LINE-1 es el *FAM65B*. En este caso, la región donde se encuentra la inserción se corresponde con el intrón 1 de 2 isoformas de este gen. De estas 2 isoformas, la más corta parece ser activa debido a los enriquecimientos de H3K4me3 y H3K9ac existentes sobre su promotor (flechas rojas). Además, los datos de ARNm-seq indican que esta isoforma se expresa ya que existen picos sobre los exones de la misma. Además estos picos procedentes del ARNm-seq se solapan con enriquecimientos de H3K36me3. Por otro lado, la isoforma más larga no presenta enriquecimiento de ninguna modificación sobre su promotor. Sin embargo, en células PA-1, según indica la RT-qPCR, este gen parece expresarse poco ya que presenta un nivel de expresión parecido a *NANOG* cuya proteína no es detectable por western-blot en células PA-1 [224]. En la región analizada se observan enriquecimientos mínimos de algunas modificaciones como H3K27ac, H3K36me3, H3K4me1 o H3K9me3.

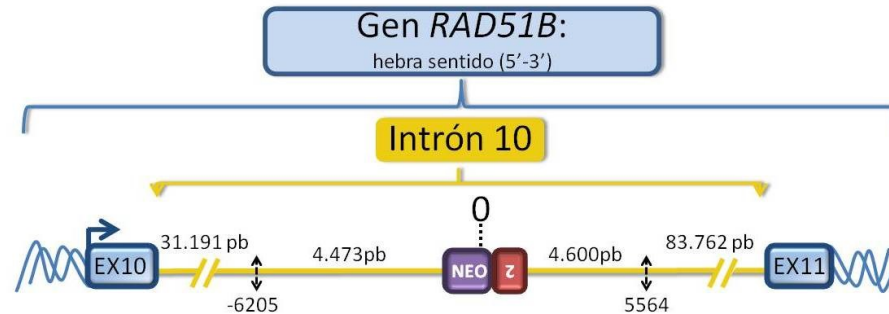
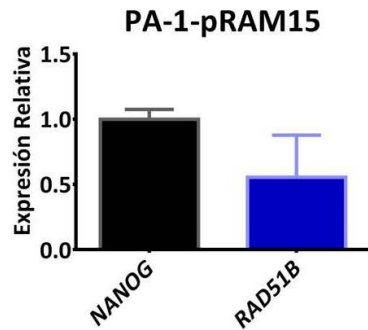
# PA-1-pRAM15

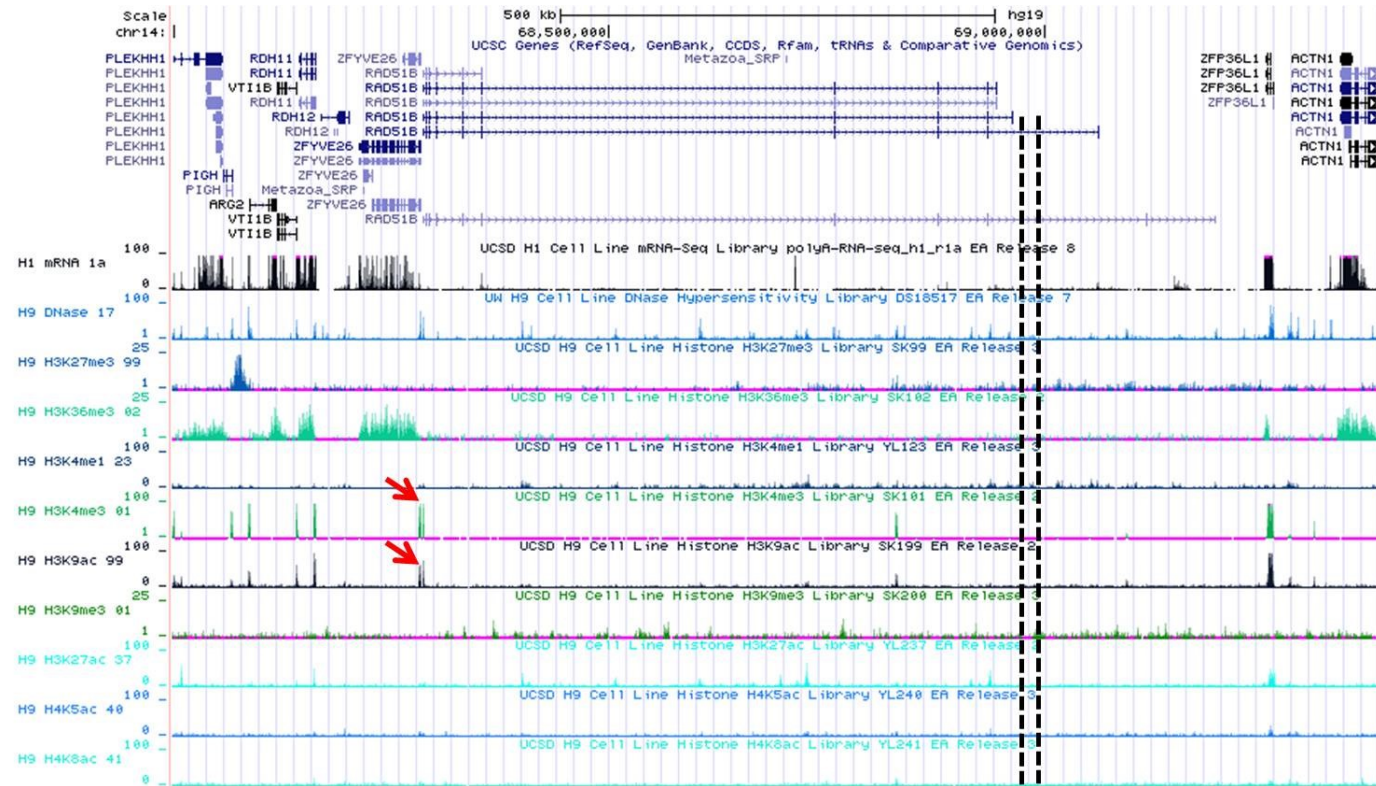
- Gen *RAD51B*
- Hebra antisentido
- Truncado en 5731
- Cola de poli A 34 pb
- DSI 8 pb

Chrm14 (q14.1.3)



**RAD51B:** Codifica una proteína implicada en la reparación del ADN mediante recombinación homóloga. Expresado en la mayoría de los tejidos, y en mayor medida en ovario, útero, glándula salivar y mama.



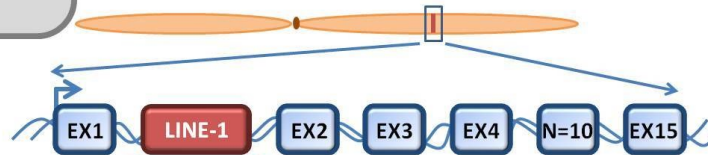


Como se puede observar, este gen (*RAD51B*) presenta un enriquecimiento elevado de H3K4me3 y H3K9ac en el promotor (flechas rojas). Por otro lado, no se detectan picos de ARNm-seq sobre los exones de las distintas isoformas. De la misma forma, no existen enriquecimientos remarcables de H3K36me3 a lo largo del cuerpo del gen. Se deduce, por tanto, que el gen, puede presentar una actividad transcripcional baja o muy baja a pesar del enriquecimiento de H3K9ac y H3K4me3 sobre su promotor. Por otra parte en células PA-1, este gen presenta un nivel de expresión bajo ya que se expresa entre 1,35 y 3 veces menos que *NANOG* cuya proteína no es detectable por western blot [224]. Cómo se puede observar, la zona analizada por CHIP-qPCR en la línea PA-1-pRAM15 está bastante desprovista de casi todas las modificaciones de histonas estudiadas (líneas negras discontinuas).

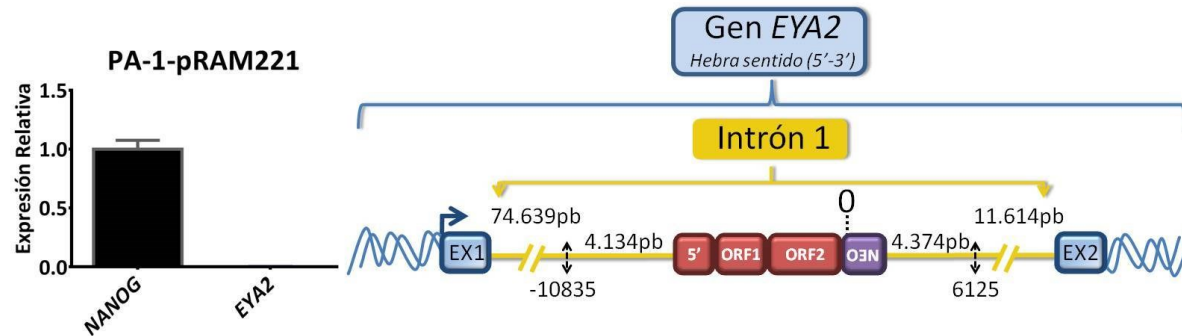
- Gen *EYA2*
- Hebra sentido
- Copia completa
- Cola de poli A 56 pb
- DSI 14 pb

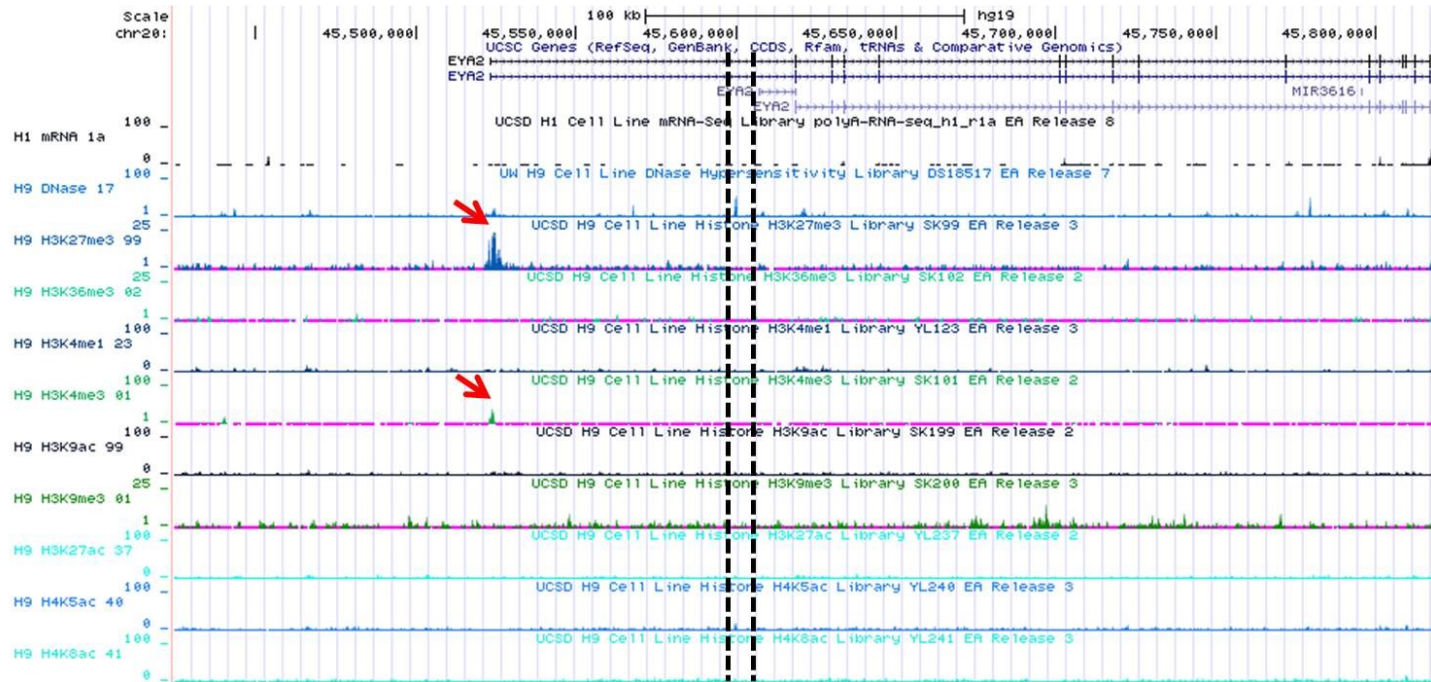
## PA-1-pRAM221

Chrm20 (q13.12)



**EYA2:** Pertenece a la familia de las proteínas de ojos ausentes. Dicha proteína puede intervenir en el desarrollo del ojo. Expresado en varios tejidos, en gran medida en nervio tibial, próstata, vagina, cuello uterino, mucosa del esófago, tiroides, glándula salivar y mama.



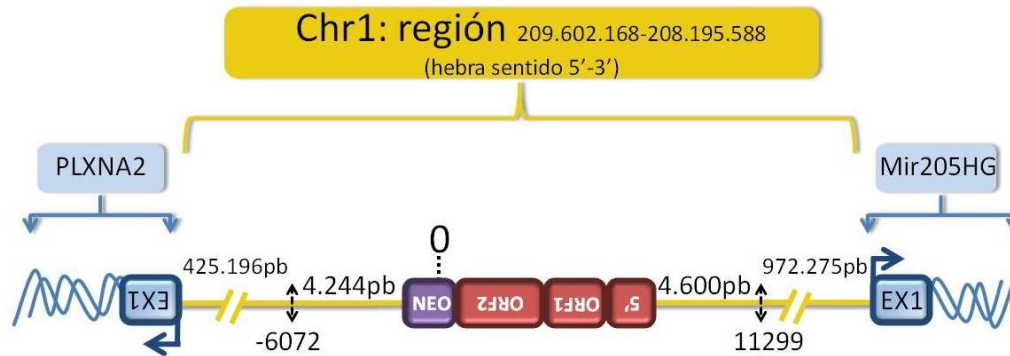


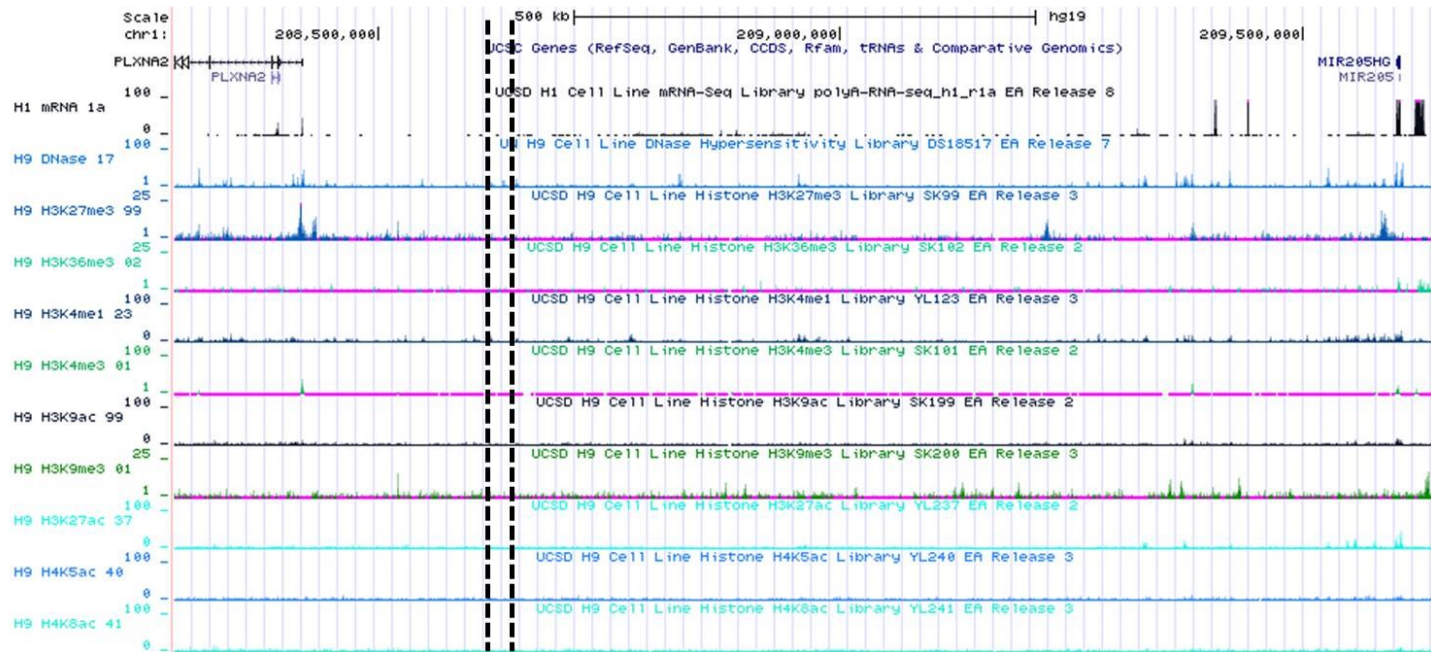
En este caso, el promotor de este gen (*EYA2*) aparece muy enriquecido en H3K27me3 y además también presenta un mínimo enriquecimiento de H3K4me3 por lo que se trata de un gen, seguramente, inactivo (flechas rojas). De hecho, no se detectan picos provenientes del ARNm-seq de células H1 así como tampoco se detectan enriquecimientos de H3K36me3 a lo largo del mismo. De hecho, esto coincide con los datos de expresión en células PA-1 obtenidos mediante RT-qPCR en células PA-1 ya que el ARN de este gen es casi indetectable por RT-qPCR. Tanto es así, que se expresa muchísimo menos (225-830 veces menos) que *NANOG* cuya proteína no es detectable por western-blot [224]. Las líneas discontinuas marcan la región analizada.

# PA-1-pRAM286

- Intergénica
- Hebra antisentido
- Copia completa
- Cola de poli A 131 pb
- DSI 12 pb

Chrm1 (q32.2)



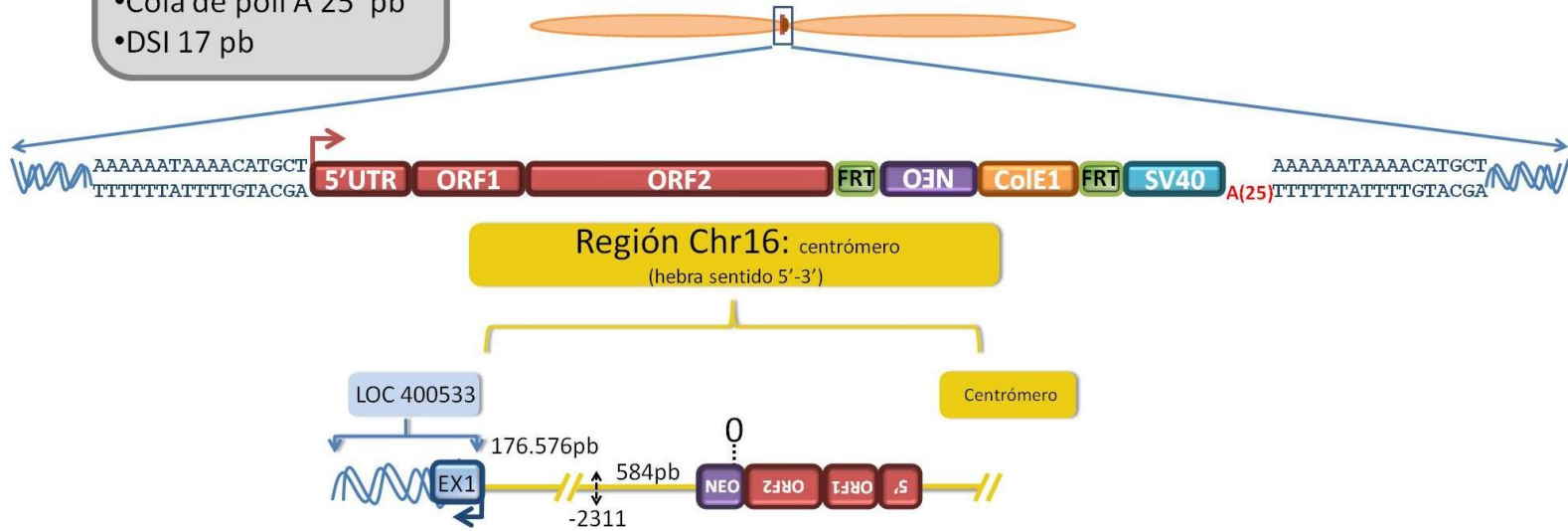


Cómo se puede observar, la zona analizada por CHIP-qPCR en la línea PA-1-pRAM286 es una zona bastante desprovista de las modificaciones de histonas que se han utilizado en este estudio (líneas discontinuas negras).

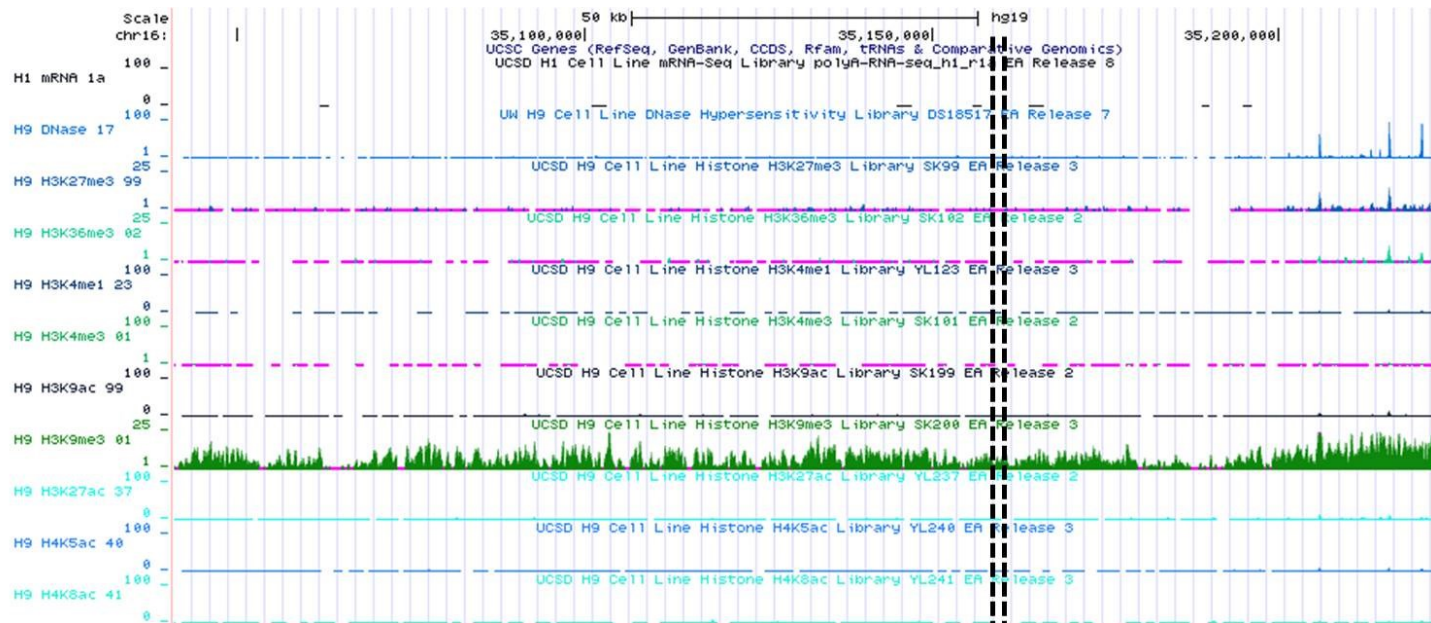
- Intergénica
- Hebra antisentido
- Copia completa
- Cola de poli A 25 pb
- DSI 17 pb

# PA-1-pRAM272

Chrm16 (p11.1)





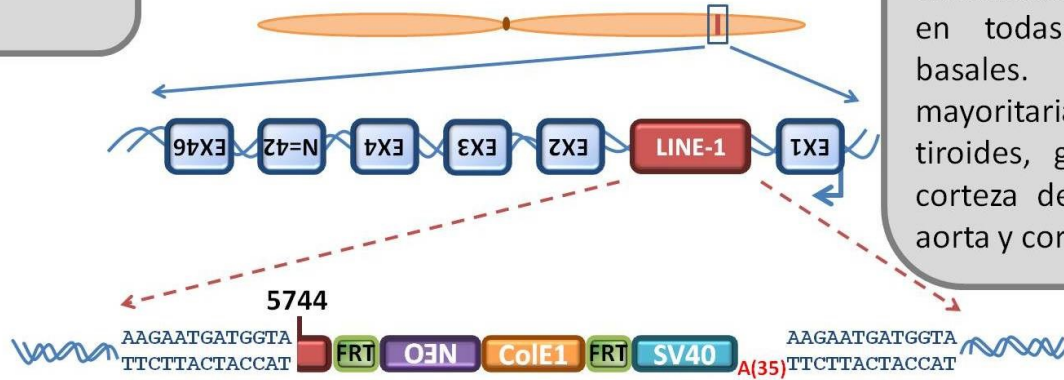


Cómo se puede observar, la zona analizada por CHIP-qPCR en la línea PA-1-pRAM272 está desprovista de todas las modificaciones de histonas estudiadas excepto de H3K9me3, la cual se encuentra bastante enriquecida en dicha zona (líneas discontinuas negras).

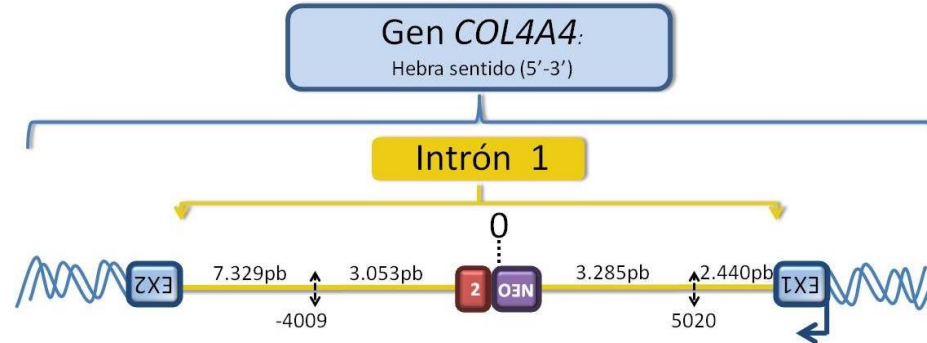
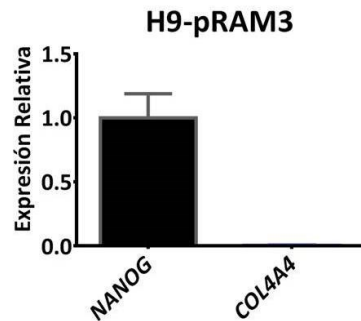
- Gen COL4A4
- Hebra sentido
- Truncado en 5744
- Cola de poli A 35 bp
- DSI 13

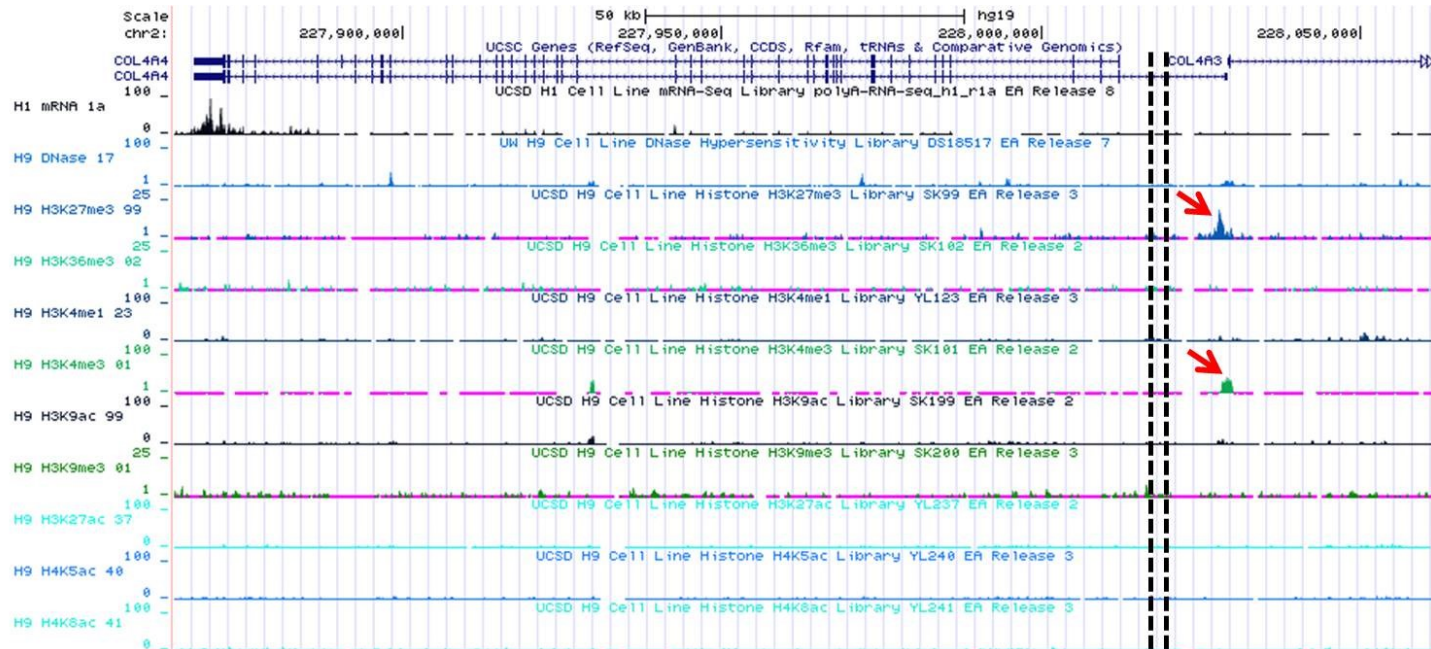
# H9-pRAM3

Chrm2 (q36.3)



**COL4A4:** codifica una de las seis subunidades del colágeno tipo IV. Esta subunidad no está presente en todas las membranas basales. Expresado mayoritariamente en tiroides, glándula pituitaria, corteza del riñón y arterias aorta y coronaria.



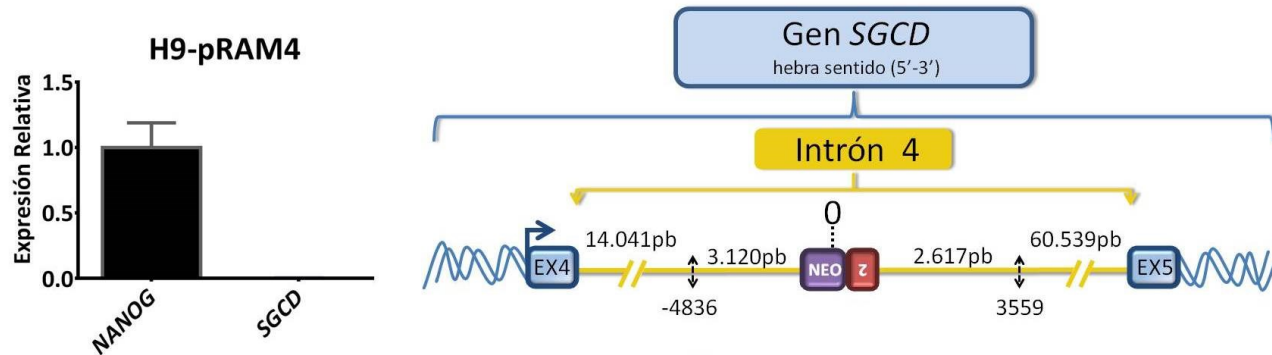
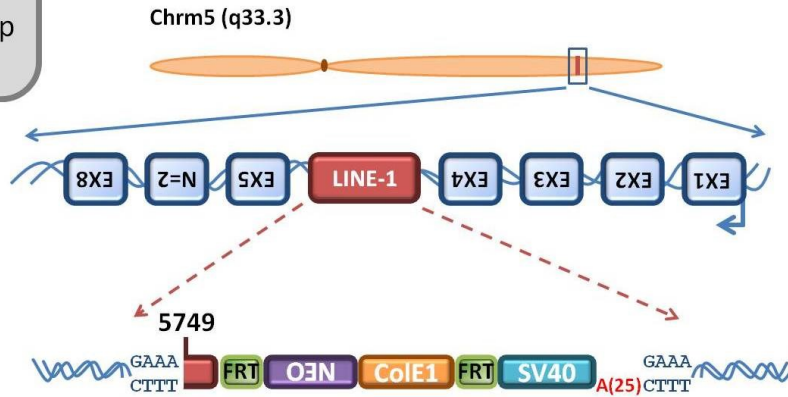


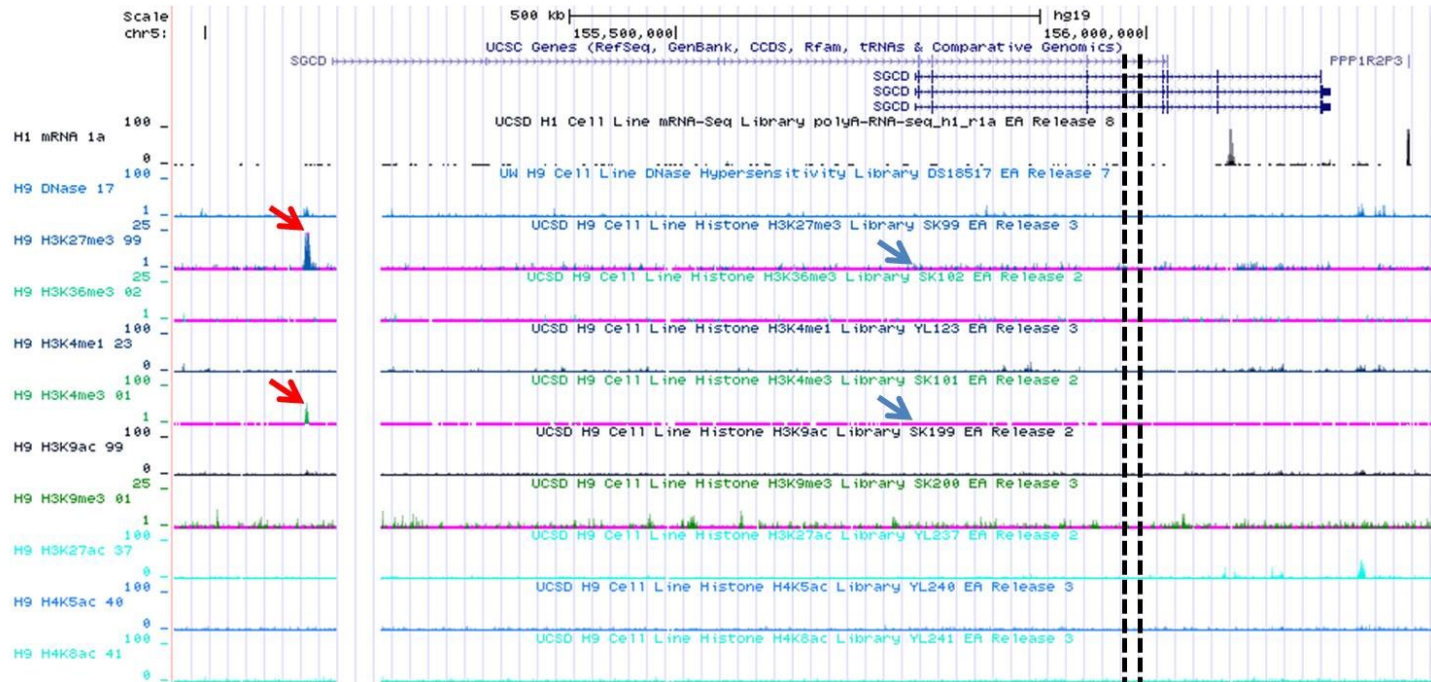
Este gen (*COL4A4*) parece encontrarse inactivo, ya que su promotor presentan un cierto enriquecimiento de H3K27me3 y un ligero enriquecimiento de H3K4me3 (flechas rojas). Esto hace que sea un gen probablemente bivalente por lo que presentará un nivel de expresión bajo. De hecho, no se observan picos procedentes del ARNm-seq sobre la secuencia del mismo así como tampoco se observan enriquecimientos de H3K36me3 a lo largo de esta zona. Esto concuerda con los datos de expresión obtenidos mediante RT-qPCR en células H9 wt ya que este gen se expresa a un nivel muy bajo (252-215 veces menos) comparado con *NANOG* que en células H9 es detectable en la mayoría de las células en cultivo [223]. No se detectan enriquecimientos considerables de ninguna modificación en la zona estudiada (líneas negras discontinuas).

# H9-pRAM4

- Gen *SGCD*
- Hebra antisentido
- Truncado en 5749
- Cola de poli A 25 bp
- DSI 4

**SGCD:** Codifica uno de los 4 componentes del complejo sarcoglicano. Este complejo forma parte del complejo distrofina-glicoproteína que une la actina F del citoesqueleto y la matriz extracelular. Muy expresado en músculo esquelético, corazón, arteria aorta, coronaria y tibial.



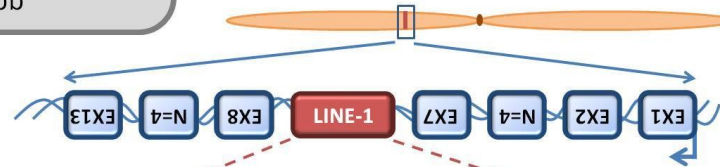


La inserción se encuentra integrada en el intrón 3 o 4 dependiendo de la isoforma considerada del gen *SGCD*. Independientemente de la isoforma, este gen se encuentra inactivo ya que bien presenta un enriquecimiento de H3K27me3 en el promotor (flechas rojas) de la isoforma larga (izquierda) o bien el promotor se encuentra totalmente desprovisto de toda modificación (flechas azules) como es el caso del resto de isoformas más cortas (derecha). De hecho, no se detecta ningún pico de ARNm-seq a lo largo de la secuencia del gen. Esta información coincide con los datos de expresión obtenidos mediante RT-qPCR en células H9 wt, donde este gen se expresa a un nivel muy bajo (487-920 veces menos) comparado con *NANOG* el cual se expresa en la mayoría de las células H9 del cultivo celular [223]. No se detectan enriquecimientos de las modificaciones de histonas estudiadas en la zona analizada (líneas negras discontinuas).

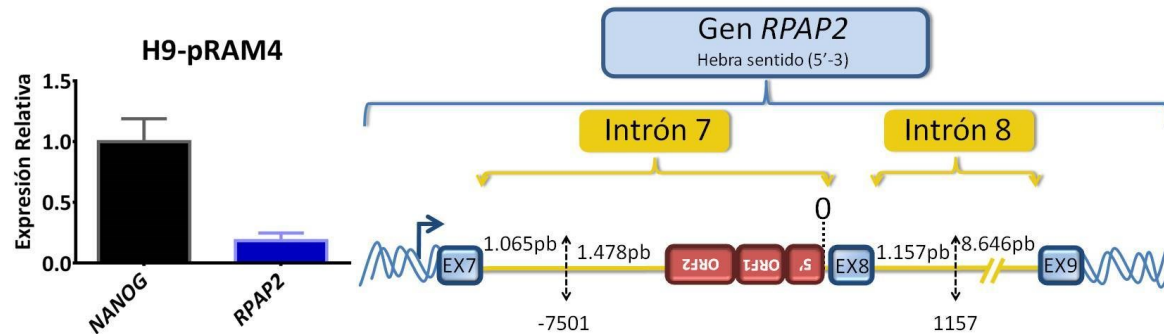
- Gen *RPAP2*
- Hebra antisentido
- Copia completa
- Cola de poli A 35 pb
- DSI 16 pb

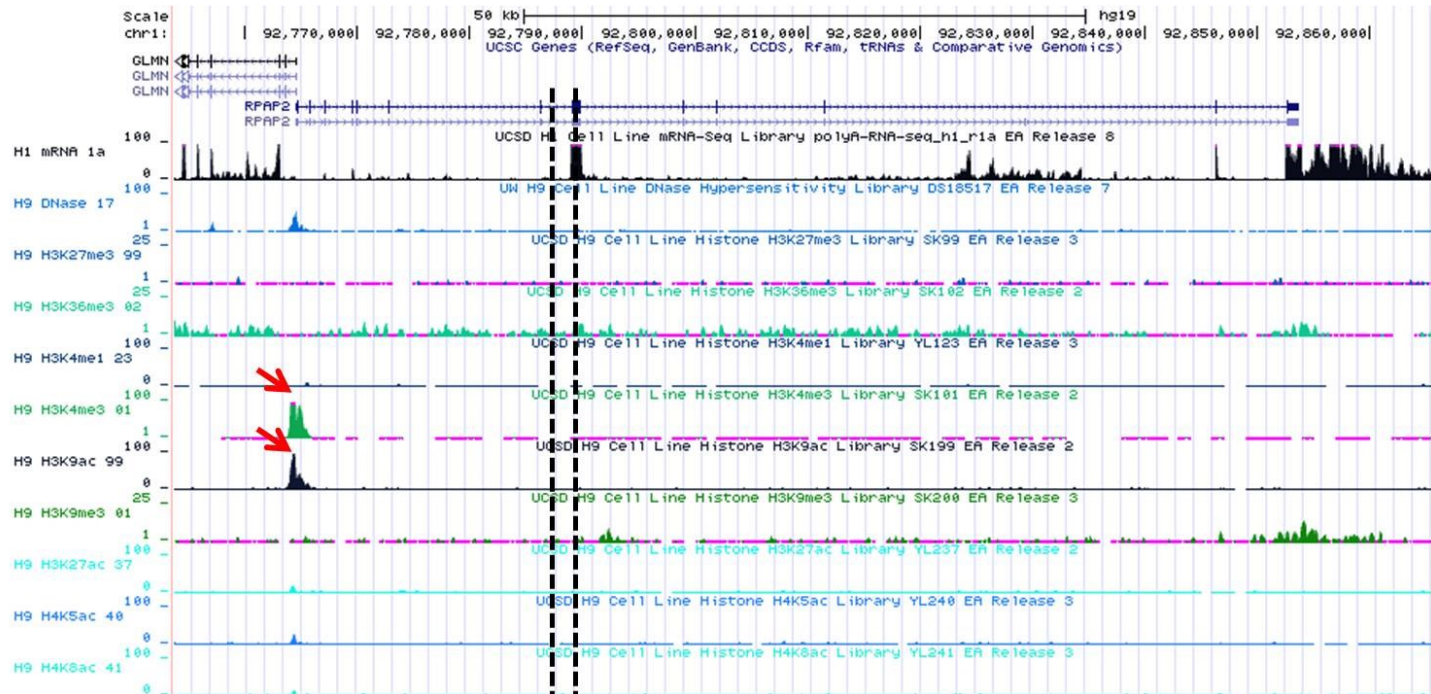
1-927

Chrm1 (p22.1)

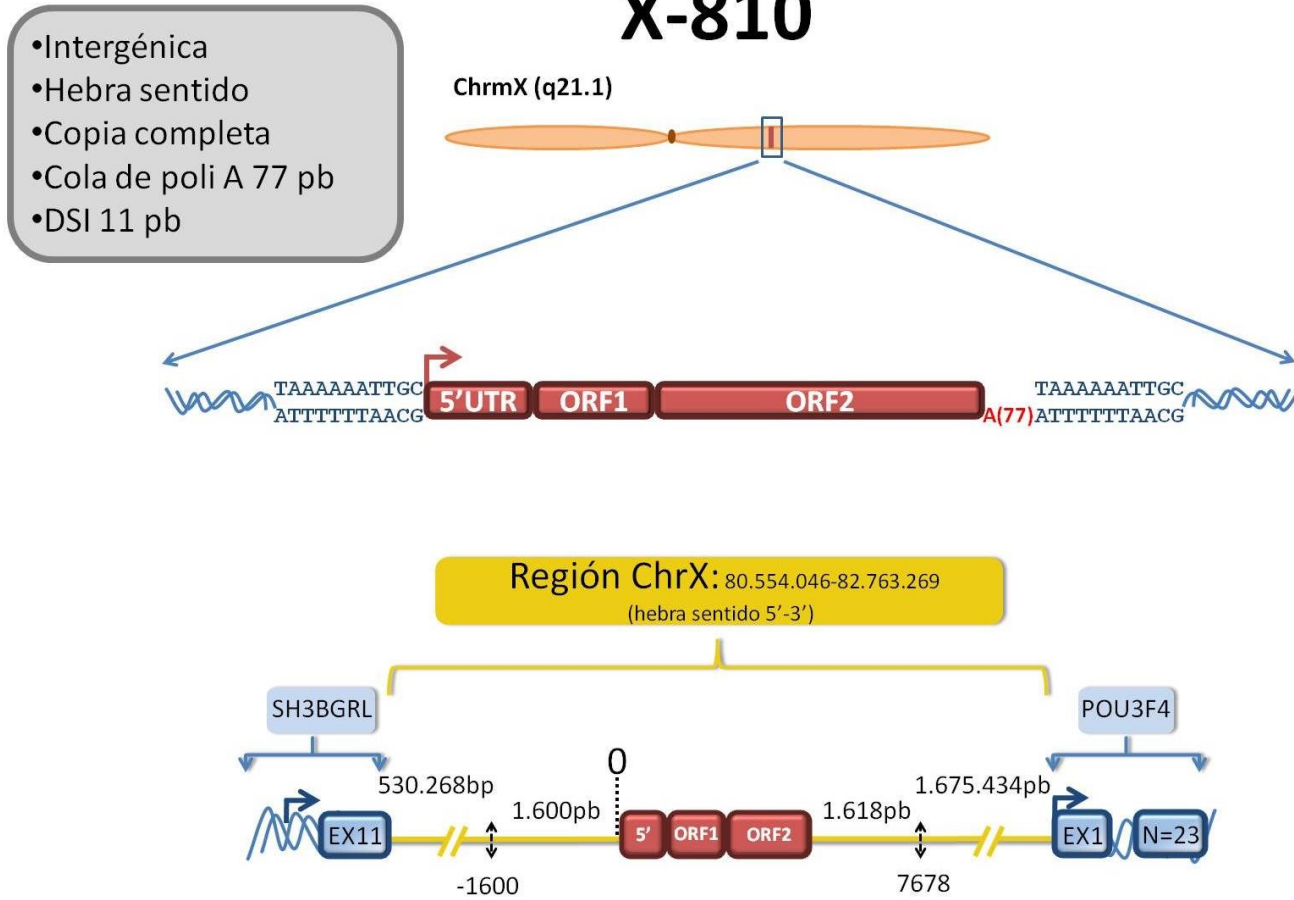


**RPAP2:** Proteína 2 asociada a la ARN polimerasa II. Gen expresado en la mayoría de los tejidos, aunque lo hace en mayor medida en cerebro.

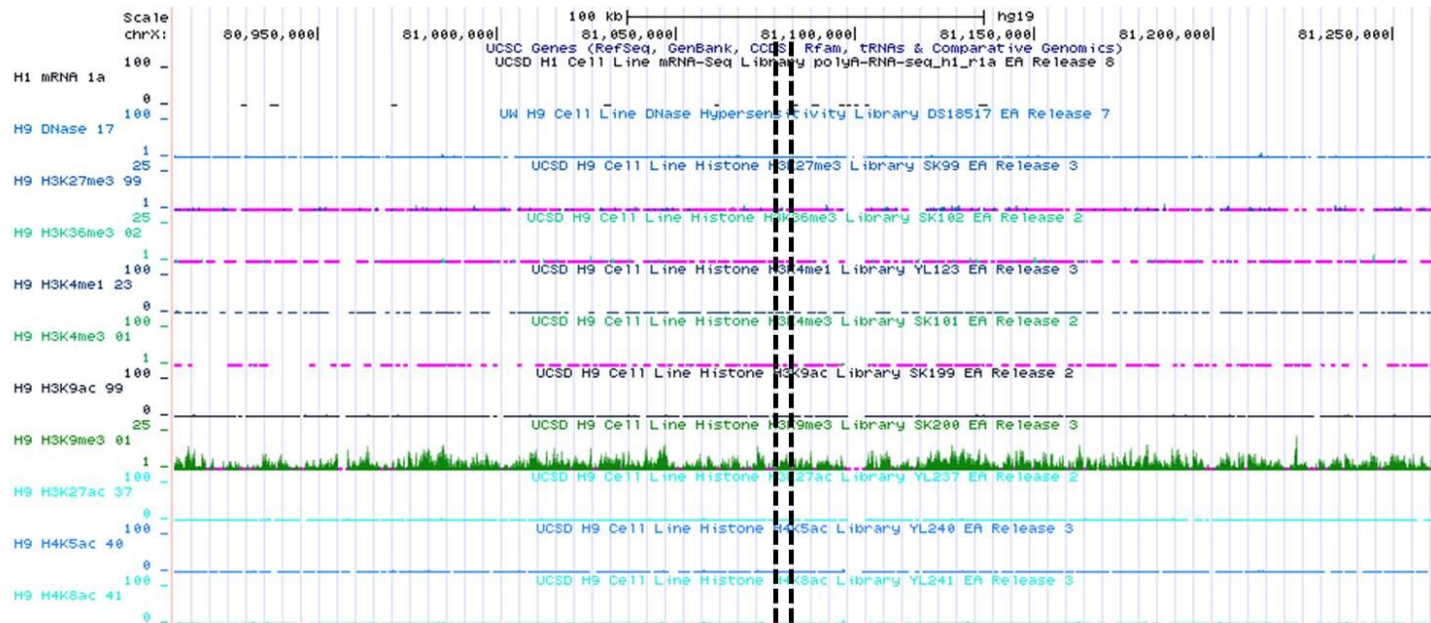




Esta inserción se encuentra dentro del gen *RPAP2* que debe estar activo, ya que su promotor, presenta un elevado enriquecimiento de H3K4me3, de H3K9ac y también aparecen ligeros enriquecimientos de otras acetilaciones como H3K27ac, H4K5ac y H4K8ac (flechas rojas). Además, se detectan picos de ARNm-seq a lo largo de la secuencia del gen. Estos coinciden con un ligero enriquecimiento de H3K36me3. Por otro lado, según los datos de expresión obtenidos por RT-qPCR en células H9 wt, este gen se expresa a un nivel que representa aproximadamente un cuarto de las expresión de NANOG cuya proteína es detectable por citometría de flujo en todas las células del cultivo [223]. Por otra parte, la región estudiada por CHIP-qPCR (líneas negras discontinuas), se encuentra desprovista de cualquier modificación de histona excepto un mínimo enriquecimiento de H3K36me3.







En este caso, la región analizada por CHIP-qPCR así como las zonas adyacentes se encuentran enriquecidas en H3K9me3.

# 4-144

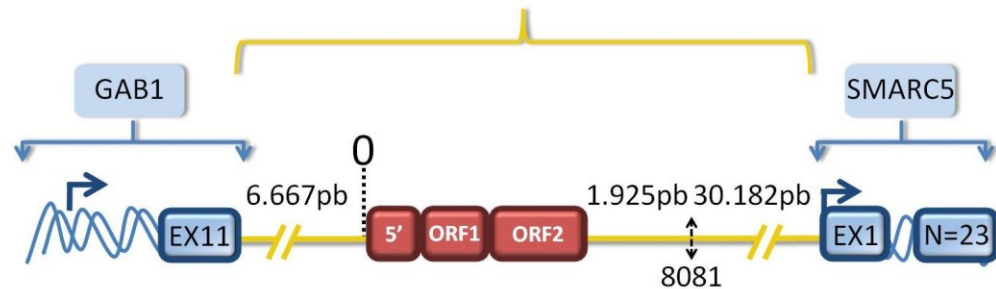
- Intergénica
- Hebra sentido
- Copia completa
- Cola de poli A 175 pb
- DSI 14 pb

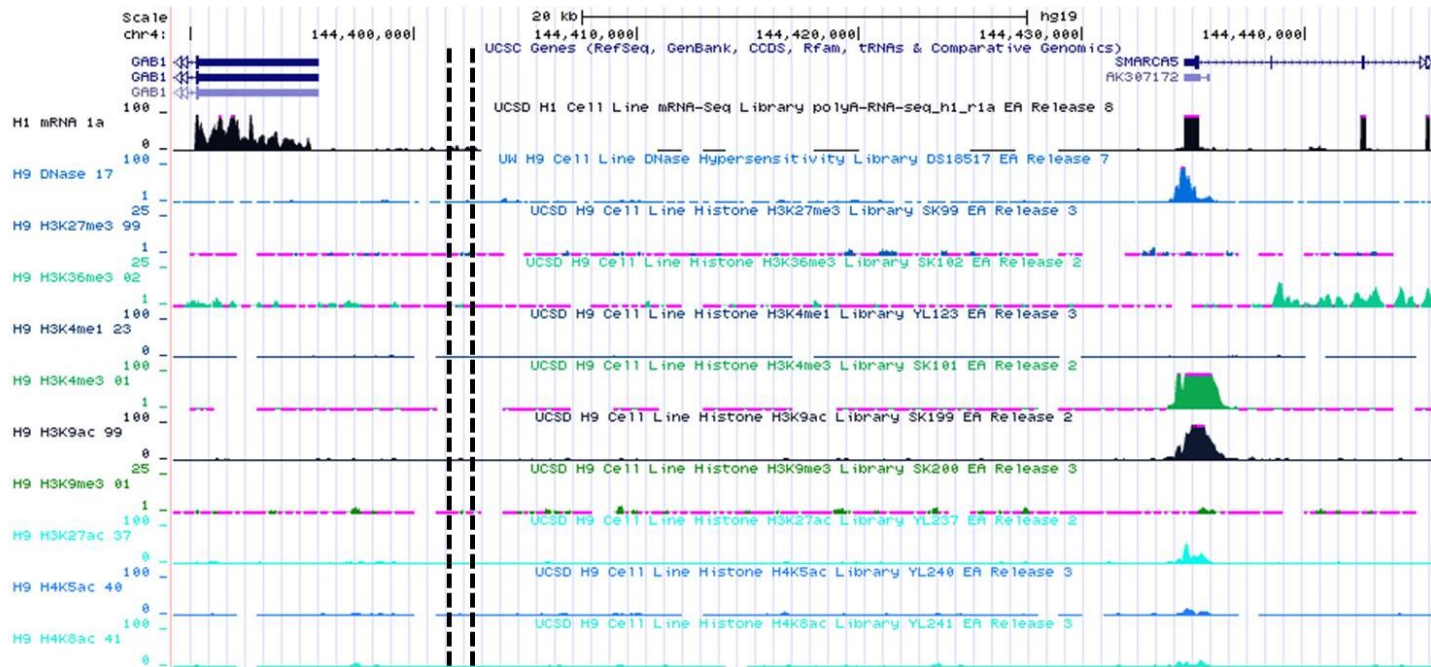
Chrm4 (q31.21)



Región Chr4: 144.395.718-144.434.616.782.304

(hebra sentido 5'-3')





Al igual en el caso anterior, la inserción se encuentra localizada en una región intergénica desprovista de las modificaciones de histonas estudiadas (líneas discontinuas negras).



# **DISCUSIÓN**



## 5. Discusión

### 5.1. Regulación de la movilidad de LINE-1 por el microprocesador (DROSHA/DGCR8).

La regulación de los elementos LINE-1 a distintos niveles es importante para mantener un control severo sobre estos elementos y minimizar sus oportunidades para retrotransponerse. Por ello, existen distintos mecanismos que actúan durante el ciclo de retrotransposición a nivel transcripcional, post-transcripcional e integracional. Se puede incluir como parte de este circuito de regulación al microprocesador (DROSHA/DGCR8) implicado en la biogénesis de microRNAs [225], [226]. Según los resultados obtenidos, el complejo microprocesador regula negativamente la retrotransposición de LINE-1. Esto se debe a que DGCR8, componente del microprocesador, es capaz de unir otros ARNs que no sean los pri-microRNAs [124]. De hecho, es capaz de regular ARNs como ARN de cadena larga intergénicos así como también ARNs mensajeros codificantes de proteína [124]. Esto se debe a que DGCR8 reconoce una estructura secundaria del ARN similar, a aquella presente en el pri-microRNA, que consiste en una zona central de 33 nucleótidos con 3 horquillas y una parte final de ARN de cadena simple en ambos extremos [123], [227]. De este modo, si el ARN diana en cuestión puede formar esa estructura secundaria, seguramente pueda ser reconocido y procesado por el microprocesador. De acuerdo con esta teoría, dicha estructura secundaria fue predicha para un gran porcentaje de los ARNs que se pueden unir a DGCR8, los cuales eran procesados *in vitro* [124]. Por otro lado, DGCR8 también puede unir los ARNs de retrotransposones como LINE-1 y retrovirus endógenos [124]. De hecho, en el artículo que acompaña a esta tesis, se demuestra que existen regiones a lo largo del ARN de LINE-1 que son sitios de unión para DGCR8 [228]. Estos sitios de unión se encuentran localizados a lo largo de los tres elementos de la secuencia del ARN de LINE-1 (5'UTR, ORF1 y ORF2) [228]. Además, se demostró que una parte de la región 5'UTR podía formar una estructura secundaria reconocida por el microprocesador [228]. Estos datos concuerdan con el hecho de que el microprocesador regule negativamente la retrotransposición.

La regulación de la retrotransposición está mediada por una interacción entre el microprocesador y la región 5'UTR del retroelemento LINE-1 que conlleva su procesamiento. De hecho, la retrotransposición en ausencia de un microprocesador funcional es mayor en presencia de la región 5'UTR lo que indica una actividad del microprocesador sobre esa zona (Figura 20C). De acuerdo con esto, el microprocesador puede procesar *in vitro* la región 5'UTR de LINE-1 en las regiones 1-300 y 200-500 [228]. Estas regiones por sí solas, sobre todo la región 1-300, inducen un aumento de la

retrotransposición en ausencia de un microprocesador funcional similar al inducido en presencia de la región 5'UTR completa (Figura 22A, B). Cabe destacar, que el resto de regiones en que fue dividida la región 5'UTR, en base a los resultados del HITS-CLIP, no son procesadas *in vitro* e igualmente no tienen efecto sobre la retrotransposición en ausencia de un microprocesador funcional. Por otro lado, la retrotransposición en presencia del DN-Drosha también se eleva ligeramente en ausencia de la 5'UTR lo cual puede ser debido al procesamiento del ARNm de LINE-1 en otras zonas ya que existen sitios de unión, como se ha comentado más arriba, también en las regiones codificantes ORF1 y ORF2. Consistentemente con el procesamiento *in vitro* del ARNm de LINE-1, células embrionarias de ratón deficientes para DGCR8 también presenta unos niveles elevados de luciferasa cuando son transfectadas con un constructo que contiene la 5'UTR fusionada al gen de la luciferasa [228]. Estos resultados junto con el procesamiento *in vitro* demuestran que el aumento de la retrotransposición en ausencia de un microprocesador funcional, tiene que ver con un procesamiento deficiente del ARNm de LINE-1. De ahí que se detecte un aumento de los niveles de ARN de LINE-1 y por consiguiente de ORF1p en células que carecen de DGCR8. La acumulación de ARNm y proteína de LINE-1 llevaría al incremento de la retrotransposición observada.

La ausencia de un microprocesador funcional como ocurre en células *DGCR8<sup>-/-</sup>* da lugar a una falta casi total de microRNAs maduros [229]. Por tanto, el incremento de la retrotransposición debido a la falta de función del microprocesador también se puede deber a la falta de microRNAs maduros que regulan la actividad de LINE-1 como mir-128 [120], o microRNAs que regulen factores celulares que sean requeridos para la movilidad de LINE-1. Sin embargo, este no parece ser el caso ya que la ausencia de la región 3'UTR del LINE-1, donde se unirían la mayoría de miRNAs reguladores, no tuvo efecto sobre la tasa de retrotransposición en ausencia de un microprocesador funcional (Figura 23B, C). Cabe destacar que se elimina la región 3'UTR del constructo porque la región 3'UTR de los genes, en general, es la región canónica de unión para los microRNAs [220]. Por otro lado, no se puede descartar totalmente que una cierta parte del aumento de la retrotransposición en ausencia de un microprocesador funcional se pueda deber a la ausencia de algún microRNA maduro. De hecho, mir-128 presenta su sitio de unión en la región ORF2 [120], con lo que su efecto sobre la retrotransposición en ausencia de un microprocesador funcional no se puede testar eliminando la región 3'UTR de LINE-1.

Es importante destacar que la regulación negativa ejercida por el microprocesador podría tener lugar durante el desarrollo embrionario temprano justo cuando se acumulan la mayoría de los nuevos eventos de retrotransposición [59], [77], [78] ya que DGCR8 es



fundamental para el desarrollo embrionario temprano [229]. Además DGCR8 es también importante para abandonar el estado de pluripotencia y comprometerse hacia la diferenciación de los distintos tejidos [229], [230]. Estas observaciones cuadran con el hecho de que se observe un aumento de los niveles de ARN y proteína de LINE-1 en células embrionarias de ratón *DGCR8*<sup>-/-</sup> lo que permite situar temporalmente esta regulación. Por último, destacar que DGCR8 es un gen ausente en un gran número de pacientes afectados por el Síndrome de DiGeorge o CATCH22. De hecho, una hipótesis de investigación derivada de este estudio propone que el genoma de estos pacientes contenga más inserciones de LINE-1 endógenos, lo cual está siendo testado en el laboratorio por otros compañeros.

## **5.2. Regulación epigenética de las inserciones *de novo* de LINE-1.**

La mayoría de las inserciones de *de novo* de LINE-1 se acumulan durante el desarrollo embrionario temprano [59], [77], [78]. Como se describe en la introducción existen distintos mecanismos de control a lo largo del ciclo de vida del retrotransposón LINE-1. Sin embargo, se sabe poco acerca de los mecanismos que intervienen a nivel post-integracional para evitar que se produzcan rondas sucesivas de retrotransposición y, por tanto, que haya una actividad excesiva de retrotransposición. Estudios pioneros llevados a cabo por el Dr. José Luí García Pérez durante su etapa postdoctoral demostraron que la mayoría de las inserciones artificiales *de novo* de LINE-1 son silenciadas mediante desacetilación de histonas en células pluripotentes, al menos de la secuencia reportera [60]. Estos datos sugerían la implicación de modificaciones de histonas mediando este silenciamiento. Esta tesis doctoral, derivada de dicho estudio, pretende completar los hallazgos realizados y trata de encontrar cuales son exactamente las modificaciones de histonas involucradas en la regulación epigenética de las nuevas inserciones de LINE-1. Para ello, se analizó la presencia de 5 modificaciones de histonas en un panel de 10 nuevas inserciones artificiales de LINE-1 con características diferentes además de 3 inserciones endógenas *de novo* de LINE-1. Los resultados obtenidos tras este análisis muestran que son tres las modificaciones que aparecen enriquecidas en las inserciones de *de novo* de LINE-1: H3K9me3, H3K27me3 y H3K4me3.

### *Silenciamiento epigenético mediante H3K9me3*

Sorprendentemente, H3K9me3 se encuentra enriquecida tanto en inserciones de líneas clonales resistentes como en inserciones de líneas clonales sensibles (8/10 inserciones totales) (Figura 31C-J). Es de destacar, que tanto sobre las inserciones de líneas resistentes como sobre las inserciones de líneas sensibles, esta modificación se encuentra enriquecida a niveles que son biológicamente relevantes. De hecho, 6/8 inserciones

presentan enriquecimientos altos similares al control positivo (Figura 31E-J) mientras que 2/8 presentan enriquecimientos intermedios/bajos (Figura 31C, D). De las 6 inserciones altamente enriquecidas, 3 pertenecen a líneas clonales resistentes y 3 a líneas clonales sensibles. Por otro lado, el nº de inserciones altamente enriquecidas en H3K9me3 es mayor que el nº de inserciones altamente enriquecidas en H3K4me3. Solo 3/10 de las inserciones presentan enriquecimientos intermedios/altos de H3K4me3 (Figura 28A-C) Además, las tres inserciones pertenecen a líneas clonales resistentes. Del mismo modo, sólo 2/10 inserciones están altamente enriquecidas en H3K27me3, perteneciendo ambas a líneas clonales sensibles (Figura 32 H, I). Por tanto, estos datos sugieren que H3K9me3 es la modificación de histona que puede jugar un papel central en la regulación post-integracional de las nuevas inserciones artificiales de LINE-1, al menos dentro de las modificaciones estudiadas en esta Tesis. Probablemente, esta regulación mediante H3K9me3 esté relacionada con la represión transcripcional de la inserción. De hecho, los datos de expresión del gen *NEO* en inserciones artificiales *de novo* de LINE-1 indican que aquellas inserciones que acumulan relativamente más H3K9me3 que H3K4me3 son las que presentan una menor expresión del gen *NEO* (Figura 33D). Esto sugiere que, el enriquecimiento de H3K9me3 sobre las inserciones de LINE-1 está asociado con su represión transcripcional y, por tanto, podría impedir que se vuelvan a retrotransponer en un genoma pluripotente después de su integración. Estos resultados están de acuerdo con un estudio publicado donde las inserciones endógenas de LINE-1 en células embrionarias de ratón son reactivadas transcripcionalmente después de la depleción de la enzima responsable del depósito de H3K9me3 [117].

Además, estos resultados concuerdan con los datos existentes acerca del silenciamiento de las nuevas inserciones de otros retroelementos como son los retrovirus endógenos. Las nuevas inserciones artificiales de éstos también se ha demostrado que son silenciadas epigenéticamente a través de la metilación de H3K9. De hecho, vectores derivados del retrovirus endógeno MLV (del inglés *Murine Leukemia Virus*) son rápidamente silenciados después de la integración mediante H3K9me2 en células embrionarias de ratón [231] o mediante H3K9me3 en células de carcinoma embrionario de ratón [232], [233]. Además, en una línea celular derivada de carcinoma embrionario de ratón, el silenciamiento de una inserción *de novo* de un vector derivado del M-MLV (del inglés *Moloney Murine Leukemia Virus*) se produce a través de H3K9me2 [201]. El hecho de compartir con otros retrotransposones y retrovirus el mecanismo por el cual las inserciones artificiales *de novo* son silenciadas, refuerza que la modificación central en este silenciamiento post-integracional de las inserciones *de novo* de LINE-1 sea H3K9me3. En el caso de retrovirus endógenos este silenciamiento mediante H3K9me3 está mediado por una serie factores represores específicos de secuencia como son KAP-1 y proteínas de la

familia KRAB-ZFN, que interaccionan con SETDB1 produciéndose el depósito de la marca. Sin embargo, los factores represores que reconocen la nueva inserción de LINE-1 y que median el depósito de H3K9me3 a través, probablemente de Suv39h1/2 son desconocidos hasta la fecha.

Además, los experimentos de CHIP-qPCR sobre los subclones flipeados revelaron que la modificación H3K9me3 se mantenía enriquecida sobre la inserción de LINE-1 una vez las secuencias exógenas eran eliminadas (Figura 48E, F, H). Cabe destacar que esta es una situación contraria a lo que sucedía con H3K4me3, cuyo enriquecimiento disminuía hasta niveles residuales después de la eliminación de las secuencias anexas (Figura 48B-D). Por tanto, estos datos confirman que existe un reclutamiento de H3K9me3 independiente de las secuencias exógenas presentes en el sistema utilizado y, por ello, este reclutamiento probablemente está asociado con la propia secuencia de LINE-1. Sin embargo, no se puede De hecho, 2/3 inserciones endógenas *de novo* estudiadas en células madre humanas embrionarias reclutan H3K9me3 a niveles biológicamente relevantes sobre su promotor confirmando los resultados obtenidos sobre los subclones flipeados (Figura 54A, B).

Por tanto, todos estos datos conjuntamente sugieren que uno de los mecanismos de control post-integracional de las inserciones *de novo* de LINE-1 en un genoma pluripotente sea el silenciamiento epigenético a través del reclutamiento de H3K9me3.

#### *Silenciamiento mediante H3K27me3*

Por otra parte, según los datos obtenidos, las inserciones *de novo* de LINE-1 podrían estar reguladas por H3K27me3 ya que 4/10 se encuentran enriquecidas en H3K27me3 (Figura 32G-J)). Además, estas 4 inserciones se encuentran altamente enriquecidas en H3K9me3 (Figura 31G-J). De hecho, no se encontraron inserciones enriquecidas en H3K27me3 poco o nada enriquecidas en H3K9me3 (Figura 32; Figura 31). Además, las únicas dos inserciones altamente enriquecidas en H3K27me3 también presentan un enriquecimiento elevado de H3K9me3 (Figura 32H, I; Figura 31H, I). Ambas inserciones pertenecen a las líneas sensibles PA-1-pRAM286 y PA-1-pRAM221. Estos datos sugieren que el enriquecimiento de H3K27me3 podría estar ligado a un enriquecimiento previo de H3K9me3. Por esta razón, la función de H3K27me3 podría estar relacionada con la consolidación del silenciamiento establecido por H3K9me3. Por tanto, es posible que el doble marcaje H3K9me3/H3K27me3 sobre las inserciones de las líneas sensibles (PA-1-pRAM286 y PA-1-pRAM221) de lugar a un silenciamiento más intenso o estable de la inserción *de novo* de LINE-1. De hecho, este fenómeno también está descrito para elementos LINE-1 endógenos de copia completa de ratón donde después de una liberación de la represión mediante metilación, se produce un rápido silenciamiento mediante un

doble marcaje H3K9me3 y H3K27me3, siendo H3K27me3 la modificación que consolida el silenciamiento [198]. Sin embargo, esta teoría no concuerda con la relación funcional que H3K9me3/H3K27me3 mantienen habitualmente sobre promotores celulares. Por ejemplo, existen genes reguladores del desarrollo en células embrionarias controlados por promotores bivalentes (H3K4me3/H3K27me3) que además suman H3K9me3 para reforzar el silenciamiento y mantener el estado de pluripotencia [161]. Además, en células madre embrionarias humanas, los promotores bivalentes (H3K4me3/H3K27me3), que controlan genes propios del estado de pluripotencia, se resuelven hacia un doble marcaje H3K9me3/H3K27me3 en células diferenciadas asegurando el mantenimiento del estado diferenciado [163]. Siguiendo estas observaciones, estos resultados pueden interpretarse de manera alternativa. H3K9me3 y H3K27me3 pueden cooperar sinérgicamente reprimiendo las inserciones *de novo* de LINE-1. Sin embargo, esta cooperación puede depender del contexto genómico/transcripcional donde se integre la inserción de LINE-1. En un contexto represivo, como el existente en la localización de las inserciones de las líneas sensibles se depositan ambas modificaciones. En un contexto menos reprimido transcripcionalmente, las inserciones de novo de LINE-1 puede reclutar H3K9me3 pero no H3K27me3. Esto explicaría que esta marca no esté presente en las inserciones de líneas resistentes.

Alternativamente, el enriquecimiento de H3K27me3 también se puede explicar a través de la longitud de la inserción ya que las 4 inserciones enriquecidas en H3K27me3 son inserciones de copia completa excepto la inserción de la línea resistente PA-1-pRAM8 que está truncada en el nucleótido 22, lo cual indica que es casi de copia completa. Esta posible relación con la longitud de la inserción podría explicar parte de su ausencia sobre el resto de inserciones de líneas resistentes ya que todas están truncadas en 5' (Ver Anexo IV). De una forma u otra, las inserciones de las líneas sensibles (PA-1-pRAM286 y PA-1-pRAM221) son las únicas altamente enriquecidas en ambas modificaciones (H3K9me3/H3K27me3).

En todo caso, el análisis realizado sugiere que la presencia de dos marcas de represión está relacionada con una mayor represión independientemente del orden de reclutamiento y de las circunstancias. Efectivamente, los datos sobre la expresión de *NEO* apuntan a que el doble marcaje H3K9me3/H3K27me3 de estas dos inserciones puede tener una función de refuerzo de la represión ya que estas líneas muestran una expresión menor comparada con la expresión de otras líneas cuyas inserciones presentan únicamente un enriquecimiento elevado de H3K9me3 (Figura 33C, D). Por otro lado, la línea PA-1-pRAM272, cuya inserción está altamente enriquecida en H3K9me3 y poco enriquecida en H3K27me3, presenta una expresión menor de *NEO* que estas dos líneas

sensibles. No obstante, dicha inserción se encuentra en una región pericentromérica donde ya seguramente exista una gran compactación de la cromatina y represión de la transcripción debido a un alto nivel de H3K9me3 (Figura 40B) lo cual explica la baja expresión del gen *NEO* en esta línea. Probablemente, la localización de esta inserción sea el motivo del bajo enriquecimiento de H3K27me3 sobre dicha inserción a pesar de pertenecer a una línea sensible [170]. Por otro lado, también sería importante confirmar que la modificación H3K27me3 podría permanecer sobre la inserción una vez que las secuencias exógenas son eliminadas. Desafortunadamente, por problemas técnicos no se pudieron generar subclones flipados para las líneas sensibles cuyas inserciones presentan el enriquecimiento más alto de H3K27me3. Sin embargo, esta hipótesis se podría comprobar en la línea PA-1-pRAM8, la cual presenta un enriquecimiento bajo de H3K27me3, ya que se dispone de su correspondiente subclon flipado.

Para confirmar que el enriquecimiento de H3K27me3 tiene la función de consolidar el silenciamiento de estas inserciones, se llevarán a cabo experimentos de inhibición química de las enzimas que depositan estas marcas para determinar cuál es el efecto de cada una de las modificaciones y si existe un efecto sinérgico. Muy probablemente, la enzima que cataliza el depósito de H3K9me3 sobre las nuevas inserciones de LINE-1 puede ser Suv39h1/2 ya que es la principal encargada del silenciamiento de las inserciones endógenas de LINE-1 de ratón [117]. Por otro lado, el depósito de H3K27me3 será seguramente catalizado por el complejo PRC2 (Polycomb Repressive Complex 2) [193], [194] ya que además éste deposita tal modificación sobre retrovirus endógenos [206]. Para ambas enzimas existen drogas que son inhibidores específicos como la caetocina que inhibe las Suv39h1/2 [234] mientras que para PRC2 existe multitud de drogas que se están usando incluso en ensayos clínicos para el tratamiento de ciertos tipos de cáncer [235]. De esta forma, mediante el tratamiento de las líneas sensibles en cultivo con estas drogas, y la posterior medida de expresión del gen *NEO* se podrá determinar la contribución de cada una las modificaciones sobre el silenciamiento de las inserciones artificiales *de novo* de LINE-1.

Inesperadamente, 2/2 inserciones endógenas *de novo* analizadas que se encuentran enriquecidas en H3K9me3 (Figura 54A, B) no presentan este doble marcaje H3K9me3/H3K27me3 sobre el promotor (Figura 54D, E). Sin embargo, este hecho podría deberse a la existencia de una distribución asimétrica de ambas modificaciones sobre la secuencia de la inserción localizándose un mayor enriquecimiento de H3K9me3 hacia el extremo 5' y un mayor enriquecimiento de H3K27me3 hacia el extremo 3' de la inserción. De hecho, este es un fenómeno descrito en células embrionarias de ratón donde elementos LINE-1 de copia completa endógenos son silenciados mediante H3K9me3 y

H3K27me3 después de sufrir una reactivación transcripcional a causa de una desmetilación global [198]. Por tanto, esta posible distribución asimétrica de ambas modificaciones podría explicar que analizando únicamente el extremo 5' de la inserción no se detectara el enriquecimiento de H3K27me3. Por esta razón, se propone analizar mediante qPCR el extremo 3' de las inserciones endógenas en un futuro cercano. Por otro lado, otra razón que podría explicar la ausencia de H3K27me3 sobre las inserciones endógenas es el nivel de enriquecimiento de H3K9me3. De hecho, las inserciones endógenas 4-144 y X-810 presentan un enriquecimiento medio y bajo respectivamente. Según los datos procedentes del análisis de las inserciones artificiales, H3K27me3 esté enriquecido sobre inserciones altamente enriquecidas en H3K9me3. Por ello, H3K27m3 podría estar únicamente enriquecida en inserciones endógenas con un enriquecimiento alto de H3K9me3. Además, es posible que el contexto transcripcional no favorezca el reclutamiento de H3K27me3 como parece ocurrir en inserciones artificiales *de novo* de LINE-1 de líneas sensibles.

#### Regulación mediante H3K4me3

H3K4me3, se encuentra enriquecida a niveles intermedios altos en 3/10 inserciones artificiales *de novo* las cuales pertenecen a líneas resistentes (Figura 28A-C). Asimismo, se trata de tres inserciones muy truncadas en 5', con lo que la mayor parte de la inserción la conforman secuencias artificiales. Además, se demostró que cuando las secuencias exógenas son eliminadas mediante recombinación por flipasa, las inserciones analizadas pierden esta modificación lo que sugiere que H3K4me3 probablemente sea reclutada por dichas secuencias (Figura 48B-D). Posiblemente, el sistema utilizado para este estudio no sea ideal para el análisis de modificaciones de histonas asociadas a actividad transcripcional debido a la inclusión de un promotor exógeno. De cualquier forma, estos resultados no pueden descartar el posible papel regulador de H3K4me3 sobre la actividad transcripcional de elementos LINE-1, ya que inserciones endógenas activas transcripcionalmente presentan H3K4me3 en la región genómica flanqueante [71] o sobre su promotor [198]. El hecho de que 1/3 de las inserciones endógenas *de novo* analizadas se encuentre enriquecida en H3K4me3 confirma que las inserciones *de novo* de LINE-1 pueden ser reguladas mediante el reclutamiento de H3K4me3 (Figura 54I). Este resultado constituye la primera demostración de que una inserción de copia completa de *de novo* de LINE-1 en el genoma humano muestra marcas epigenéticas asociadas a actividad transcripcional tras su integración. Esto podría posibilitar su transcripción lo cual podría acabar en un nuevo ciclo de retrotransposición. Desde un perspectiva evolutiva, este resultado concuerda con el hecho de que es necesario el mantenimiento de algunas copias retrocompetentes epigenéticamente activas para que se sigan generando nuevas inserciones, ya que un silenciamiento excesivo podría reducir notablemente el número de

elementos LINE-1 activos lo cual podría resultar en la extinción de estos elementos en el genoma humano.

### Regulación mediante acetilación

Consistentemente con los datos discutidos más arriba, tan sólo una inserción (PA-1-pRAM4) de las 4 analizadas se encuentra enriquecida en H4ac y H3K9ac (Figura 29A; Figura30A). Esta inserción es muy activa según los datos de expresión de *NEO* (Figura 33C) y con un enriquecimiento intermedio/alto de H3K4me3 (Figura 28A). Esto es esperable ya que H3K9ac y la acetilación, en general, está asociada con la presencia de H3K4me3 [136], [140]. Sin embargo, las tres inserciones restantes que no están enriquecidas en H4ac o H3K9ac presentan un elevado enriquecimiento de H3K9me3 o incluso un elevado enriquecimiento de H3K9me3/H3K27me3. Estos datos sugieren la existencia de una relación de exclusión entre la acetilación y las modificaciones represivas H3K9me3 y H3K27me3. Esto concuerda con estudios publicados en genomas de mamíferos donde existe una correlación inversa entre estas modificaciones represivas y la acetilación. De hecho, estos datos están de acuerdo con lo publicado en García-Pérez et al., [60] ya que las inserciones artificiales *de novo* de LINE-1 son silenciadas mediante desacetilación de histonas acompañado de un aumento del enriquecimiento de H3K9me2. Del mismo modo, nuevas inserciones de un vector retroviral derivado del retrovirus endógeno MLV (del inglés *Murine Leukemia Virus*), son silenciadas mediante desacetilación de la histona H3 en células de carcinoma embrionario de ratón y un aumento del enriquecimiento de H3K9me3 y H3K27me3 [232]. Asimismo, retrovirus endógenos como IAP son reactivados mediante la depleción condicional del co-represor KAP-1 lo que conlleva un aumento de la acetilación y una disminución del enriquecimiento de H3K9me3 [204]. Por tanto, todos estos datos indican que la ausencia de acetilación detectada sobre 3/4 inserciones artificiales *de novo* de LINE-1 es una forma de favorecer y contribuir a su silenciamiento epigenético.

### Expresión génica, longitud de inserción y reclutamiento de modificaciones de histonas

Analizando las inserciones que están localizadas dentro de genes se observa que, en 5/6 inserciones analizadas, existe una relación entre el enriquecimiento relativo de las modificaciones H3K4me3 y H3K9me3 y la expresión del gen donde se encuentra integrada dicha inserción (Figura 33E). Es decir, que inserciones que presentan un balance de reclutamiento relativo de  $H3K4me3/H3K9me3 > 1$  se encuentran dentro de genes que presenta unos niveles de expresión superiores a aquellos genes que albergan inserciones cuyo balance de enriquecimiento relativo de  $H3K4me3/H3K9me3 < 1$ . Además, las inserciones con un balance de reclutamiento relativo  $H3K4me3/H3K9me3 > 1$  localizadas

dentro de genes más activos coinciden con aquellas que presentan una mayor expresión del gen *NEO* y viceversa (Figura 33E, C). Estos resultados se encuentran en una línea similar a un estudio realizado sobre un panel de líneas celulares humanas que ha demostrado que los LINEs endógenos de copia completa de la subfamilia L1Hs transcripcionalmente más expresados, se encuentran dentro de genes que presentan un nivel de expresión superior a aquellos genes que albergan inserciones poco o nada expresadas [71]. Además, estos autores demostraron que LINEs transcripcionalmente activos de copia completa están enriquecidos en H3K4me3 en la región flanqueante de su promotor en una de las líneas celulares. Por tanto, sería interesante, determinar el enriquecimiento de H3K4me3 y H3K9me3 en un mayor número de inserciones génicas pertenecientes tanto a líneas resistentes como a líneas sensibles, además de determinar los niveles de expresión de dichos genes hospedadores para aumentar la potencia estadística de dicha relación. Además, habría que incluir un mayor número de inserciones de copia completa en distintos escenarios de actividad transcripcional para determinar cuál es el comportamiento en cada caso del genoma hospedador. Sin embargo, estos datos preliminares junto con los datos de Philippe et al., [71] apuntan a que el estado transcripcional del gen hospedador podría influir en las modificaciones de histonas que finalmente son reclutadas para regular las nuevas inserciones artificiales *de novo* de LINE-1 las cuales tienen un efecto sobre su transcripción.

La longitud de la inserción es potencialmente muy determinante en términos de reclutamiento de H3K9me3. De hecho, las inserciones con una mayor longitud son aquellas que presentan un mayor enriquecimiento de H3K9me3 (Figura 33F). Estos resultados concuerdan con el hecho de que las inserciones endógenas de LINE-1 en células embrionarias de ratón, también presentan un mayor reclutamiento de esta modificación a medida que la longitud de las mismas aumenta. De hecho, las inserciones de copia completa son las más enriquecidas en H3K9me3 [117]. Por otro lado, las inserciones endógenas de la subfamilia LIPA4 y LIPA5 con una longitud superior a 5000 pares de bases son aquellas, que presentan un enriquecimiento mayor de H3K9me3 sobre la 5'UTR en células madre embrionarias humanas [210]. Conjuntamente, estos datos sugieren que secuencias presentes en los elementos de copia completa serán responsables del reclutamiento de H3K9me3 posiblemente atrayendo la maquinaria enzimática encargada de su depósito.

Se deduce de lo expuesto anteriormente que la regulación epigenética de una nueva inserción de LINE-1 será el resultado probablemente de la combinación de ambos factores y, por tanto, del balance de fuerzas reguladoras dependientes del contexto genómico y de la propia secuencia de LINE-1. Por ello, es posible especular con que este balance haya sido



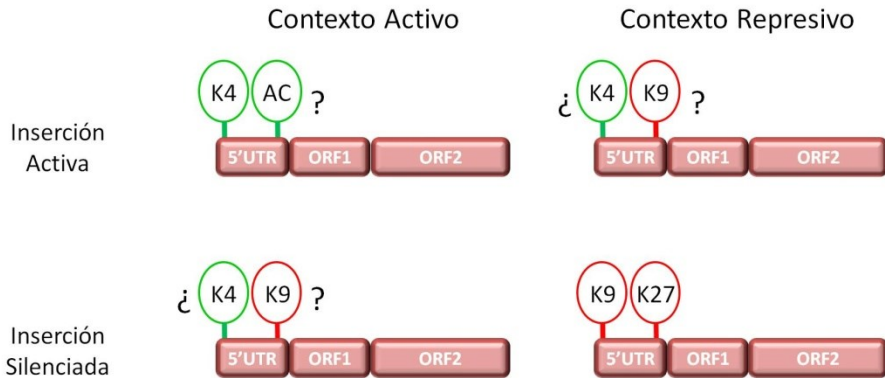
seleccionado durante la evolución para minimizar el potencial mutagénico de LINE-1 así como para reducir el impacto de la regulación de una nueva inserción de LINE-1 sobre el genoma hospedador. De esta forma, la regulación de una nueva inserción de LINE-1 será diferente en función del impacto que pueda causar sobre el genoma silenciar o mantener activa una nueva inserción. Por ejemplo, la inserción 1-927 se encuentra enriquecida en H3K4me3 (Figura 54I) siendo una inserción de copia completa que, por definición, es retrocompetente. Sin embargo, esta inserción no se encuentra enriquecida en H3K9me3 como se podría esperar. Esto se podría deber a que esta inserción se encuentra en un gen denominado *RPAP2* que se expresa solo 4 veces menos que *NANOG* el cual es un gen bastante expresado en células embrionarias [223]. De hecho, este gen debe ser esencial ya que codifica para una proteína asociada a ARN polimerasa II que se expresa en la mayoría de los tejidos. Este dato puede explicar que una inserción de copia completa no sea silenciada mediante H3K9me3 y por el contrario reclute una marca asociada a actividad transcripcional. Por tanto, es posible que el silenciamiento del retrotransposón en este determinado contexto pudiera tener consecuencias negativas para el genoma hospedador. En la esquina opuesta se encuentra la inserción artificial *de novo* de la línea sensible PA-1-pRAM221 la cual es de copia completa y además se encuentra en un gen *EYA2* cuyo ARNm es casi indetectable por RT-qPCR (Ver Anexo IV). Esto explica que en este caso exista un alto reclutamiento de H3K9me3 y H3K27me3 (Figura 31I; Figura 32I) para silenciar la inserción, ya que el coste de silenciar la inserción dentro de un contexto ya reprimido deber ser menor que en el caso de la inserción endógena 1-927. Sin embargo, es posible que el silenciamiento de la inserción de la línea PA-1-pRAM221 tenga un impacto sobre el genoma en etapas posteriores aumentando el coste de su silenciamiento. Por tanto, el nivel de expresión de un gen en pluripotencia podría ser un factor que defina el posible efecto de una inserción en etapas posteriores de desarrollo o tejidos diferenciados.

En conclusión, la regulación epigenética de una nueva inserción de LINE-1 podría depender tanto del contexto transcripcional donde se inserte así como de la presencia de secuencias concretas presentes en los elementos LINE-1 de copia completa.

#### *Modelo de regulación de las inserciones de novo de LINE-1*

Finalmente teniendo en cuenta todos los datos generados en esta tesis sobre la regulación epigenética así como los estudios publicados anteriormente, se podría establecer el siguiente modelo para las inserciones *de novo* de LINE-1 de copia completa que potencialmente podrían tener un impacto sobre el genoma. Inserciones, que se integran en contextos altamente activos, se mantienen activas mediante el reclutamiento de H3K4me3 (Inserción 1-927, Figura 54I) y probablemente mediante el reclutamiento de acetilaciones como H3K9ac y H4ac (Figura 57, arriba izquierda). Por otro lado, las

inserciones artificiales de líneas sensibles (PA-1-pRAM286/PA-1-pRAM221) integradas en un contexto, por definición, muy represivo se silencian mediante el reclutamiento de H3K9me3 y de H3K27me3 (Figura 31H, I; Figura 32H, I) (Figura 57, abajo derecha). Ambos tipos de inserciones representan las dos clases extremas fácilmente identificables.



**Figura 57: Modelo de regulación de la inserciones de novo de LINE-1.** Detalles en el texto. Los signos de interrogación al lado de cada marca significan que se desconoce si la modificación en cuestión aparece sobre inserciones de copia completa. **Arriba izquierda)** Se desconoce si las acetilaciones estudiadas están enriquecidas sobre el promotor de inserciones activas endógenas de copia completa. **Arriba derecha y abajo izquierda)** Datos provenientes de inserciones artificiales de líneas resistentes truncadas en 5'. Se desconoce la presencia de estas dos marcas de forma combinada sobre el promotor de una inserción de copia completa en los contextos representados salvo la inserción de la línea sensible PA-1-pRAM286. **Abajo derecha)** Modificaciones enriquecidas sobre inserciones artificiales de copia completa. Se desconoce su presencia combinada sobre el promotor de inserciones endógenas de copia completa. Las abreviaciones usadas son las siguientes: K4, H3K4me3; AC, H3K9ac y H4ac; K9, H3K9me3 y K27, H3K27me3.

Por otro lado, se podrían establecer dos categorías intermedias de inserciones. En primer lugar, se podría hablar de inserciones activas en un contexto represivo (Figura 57, arriba derecha) y, en segundo lugar, se podría hablar de inserciones silenciadas en un contexto activo (Figura 57, abajo izquierda). Los datos provenientes del análisis de las inserciones artificiales de líneas resistentes sugieren que en ambas clases o categorías las inserciones podrían estar enriquecidas en H3K4me3 y H3K9me3. De hecho, existen inserciones (H9-pRAM4 y PA-1-pRAM26) que presentan un mayor enriquecimiento relativo de H3K9me3 respecto a H3K4me3 (Figura 33D) y su nivel de expresión del gen *NEO* es inferior al de las inserciones de las líneas resistentes con un mayor enriquecimiento relativo de H3K4me3 respecto a H3K9me3 (Figura 33C, D). Por tanto, estas inserciones se pueden considerar inserciones silenciadas en un contexto activo ya que se trata de líneas resistentes (Figura 57, abajo izquierda). En el lado contrario, también existe una inserción

(H9-pRAM3) que presenta un mayor enriquecimiento relativo de H3K4me3 respecto a H3K9me3 (Figura 33D) y su expresión del gen *NEO* es superior a la de las inserciones de líneas resistentes con un enriquecimiento relativo de H3K9me3 superior respecto a H3K4me3 (Figura 33C, D). Además, esta inserción se encuentra en un contexto más represivo que las inserciones más activas de líneas resistentes (Figura 33E). Por tanto, esta inserción podría ser una inserción activa en un contexto represivo (Figura 57, arriba derecha).

Cabe destacar que estos datos de las clases intermedias provienen fundamentalmente de inserciones artificiales truncadas en 5', mientras que los datos de las dos clases extremas provienen de inserciones endógenas y artificiales de copia completa. Además, cabe recordar que el enriquecimiento de H3K4me3 en las inserciones artificiales estaba mediado por las secuencias exógenas. Por tanto, se desconoce si las inserciones de las clases intermedias podrían ser encontradas de forma endógena en el genoma. De hecho, de las 3 inserciones endógenas analizadas ninguna presenta el doble enriquecimiento H3K4me3 y H3K9me3. Sin embargo, la inserción de la línea sensible PA-1-pRAM286 presentaba un enriquecimiento significativo pero mínimo de H3K4me3 (Figura 28H) junto con un enriquecimiento alto de H3K9me3 (Figura 31H), lo cual indica que inserciones de copia completa pertenecientes a las categorías intermedias quizás podrían existir. Por ello, como se mencionaba anteriormente, sería de gran ayuda poder analizar un gran número de inserciones endógenas de copia completa en distintos contextos transcripcionales para completar este modelo de una forma más fehaciente.

### **5.3. Cambios en el enriquecimiento de modificaciones de histonas en la cromatina adyacente.**

Hasta la fecha, se conocen distintos mecanismos a través de los cuales los elementos LINE-1 pueden tener un efecto sobre el genoma hospedador. La mayoría de las enfermedades asociadas a la actividad de LINE-1 se han producido debido a la integración de una nueva copia en un exón, a la alteración del proceso de splicing o a la generación de deleciones, lo cual conlleva en todos los casos a una alteración de la función del gen. Como se puede deducir, todos estos mecanismos son efectos directos sobre la secuencia del ADN en el lugar de inserción. Sin embargo, se sabe poco acerca del efecto que puede tener una nueva inserción de LINE-1 mediante la alteración de la epigenética local del sitio de inserción. A raíz de las observaciones del Dr. José Luís García Pérez sobre el silenciamiento epigenético de las nuevas inserciones *de novo* de LINE-1 mediante desacetilación de histonas, se planteó si este silenciamiento podría tener un impacto sobre el contexto epigenético que rodea a una nueva inserción. De ser así, sería posible que una nueva

inserción de LINE-1 generara una enfermedad o un desorden mediante un efecto epimutágeno sin alterar directamente la secuencia del ADN

*Ausencia de cambios de H3K4me3 y H3K27me3 en el genoma adyacente*

Una vez realizado el análisis del panel de 10 líneas clonales con 5 modificaciones de histonas, se analizó si las inserciones enriquecidas en estas modificaciones podían generar cambios en el genoma adyacente. Los cambios de enriquecimiento de H3K4me3 en el genoma adyacente producidos por las inserciones artificiales *de novo* de LINE-1 (Figura 34B, E; Figura 35B, E) son producto, muy probablemente, de la presencia de secuencias exógenas asociadas a la inserción. La disminución del enriquecimiento de H3K4me3 hasta niveles residuales en los subclones flipeados sugería que los cambios en el genoma adyacente podrían desaparecer también (Figura 48B-D). De hecho, los cambios de H3K4me3 detectados en las líneas PA-1-pRAM6 y PA-1-pRAM8 (Figura 35B, E) desaparecen en el genoma adyacente de los subclones flipeados procedentes de ambas líneas (datos no mostrados). Por otro lado, no se detectaron cambios en el genoma adyacente a la inserción endógena 1-927 (Figura 56B), la cual presentaba un enriquecimiento medio de H3K4me3 sobre su promotor. Por tanto, conjuntamente, estas evidencias sugieren que las inserciones *de novo* de LINE-1 no generan cambios de enriquecimiento de H3K4me3 sobre el genoma adyacente. La ausencia de cambios en el enriquecimiento de H3K4me3 en la cromatina adyacente podría tener que ver con la propia naturaleza de esta modificación. De hecho, esta modificación suele concentrarse en torno al sitio de inicio de la transcripción sobre el promotor de los genes [137], [151]. Además los dominios de H3K4me3 en general son de una menor extensión que los dominios de otras modificaciones como puede ser H3K27me3 [134], [160]. Asimismo, en general, está aceptado que a medida que se produce el proceso de diferenciación celular se produce una expansión de los dominios de marcas represivas tales como H3K9me3 y H3K27me3 [162], [163], [165], no estando documentado la expansión de H3K4me3 según lo encontrado en la literatura. En conclusión, estos datos sugieren que la ausencia de cambios de H3K4me3 en el genoma adyacente se encuentra dentro de lo esperable.

Sorprendentemente, no se detectaron cambios en el enriquecimiento de H3K27me3 en el genoma adyacente en aquellas líneas clonales con inserciones que presentan un elevado enriquecimiento de esta marca (Figura 41A, C). Las inserciones analizadas que presentaban un elevado enriquecimiento de H3K27me3 (Figura 32H, I) se corresponden con aquellas que estaban altamente enriquecidas en H3K9me3 (Figura 31H, I) y las cuales generaron los cambios de H3K9me3 en el genoma adyacente (PA-1-pRAM286 y PA-1-pRAM221) (Figura 38B, E). Es posible que un contexto de represión donde están presentes ambas modificaciones sea más favorable para la expansión de los cambios de H3K9me3.

Además, la ausencia de cambios de H3K27me3 también se puede deber al patrón de distribución de H3K27me3 sobre la inserción de LINE-1. Como se ha mencionado anteriormente, el cultivo de células embrionarias de ratón en medio 2i+Vit C conlleva que los elementos LINE-1 del genoma de ratón sean reactivados transcripcionalmente debido a una desmetilación global. Estos son silenciados mediante el enriquecimiento asimétrico de H3K9me3 y H3K27me3 localizándose ésta última hacia el extremo 3' de la inserción restringiendo su enriquecimiento a la secuencia del retroelemento. Al contrario que H3K9me3, que está más localizada hacia el extremo 5' y se adentra ligeramente en el genoma adyacente [198]. Por tanto, la ausencia de cambios de H3K27me3 en el genoma adyacente también se podría explicar desde esta perspectiva. Contrariamente a lo que ocurre con H3K4me3, se esperaría que hubieran cambios en el enriquecimiento de H3K27me3 en el genoma adyacente ya que, en general, su distribución en otras regiones del genoma se asocia con grandes dominios de varias kilobases de media [160], [163], [164]. Además, está documentado la expansión de estos dominios de H3K27me3 durante el proceso de diferenciación celular [162], [163], [165]. Asimismo, la enzima responsable de esta modificación PRC2 también presenta este tipo de distribución ocupando dominios de varias kilobases [192]. Por tanto, el comportamiento de esta modificación sobre los elementos LINE-1 puede ser diferente al mostrado en otras regiones del genoma explicándose así la ausencia de cambios en el genoma adyacente.

#### *Cambios de H3K9me3 en el genoma adyacente*

De todas las inserciones analizadas que se encontraban enriquecidas en H3K9me3, 2/8 inserciones artificiales generaron cambios en el estatus de H3K9me3 en la cromatina adyacente (Figura 38B, E). Cabe destacar que estas 2 inserciones pertenecen a dos líneas sensibles (PA-1-pRAM286 y PA-1-pRAM221) las cuales se encuentran altamente enriquecidas en H3K9me3 y H3K27me3 (Figura 31H, I; Figura32H, I). Además, estos cambios se producen únicamente en líneas sensibles probablemente porque son dependientes del contexto transcripcional, de la longitud de las inserciones y del enriquecimiento conjunto de marcas de represión (H3K9me3/H3K27me3). En primer lugar, las inserciones de líneas sensibles son las únicas de copia completa de todo el panel de inserciones analizadas. En segundo lugar, estas inserciones se encuentran en un contexto represivo por definición ya que pertenecen a líneas sensibles. En efecto, la inserción de la línea PA-1-pRAM221 se encuentra integrada en el gen *EYA2* cuyo ARNm era casi indetectable por RT-qPCR (Ver Anexo IV). Por último, el doble marcaje H3K9me3/H3K27me3 puede favorecer la generación de estos cambios en contraposición al doble marcaje H3K4me3/H3K9me3 encontrado en inserciones de líneas resistentes incapaces de generar cambios de H3K9me3 con enriquecimientos igualmente elevados

(Tabla 9). Asimismo, existen evidencias de que, efectivamente, estos cambios son inducidos por la propia inserción de LINE-1. En primer lugar, se descarta que los cambios se deban a la naturaleza clonal de la línea analizada según demuestran los experimentos de estabilidad epigenética de las líneas clonales de PA-1 (Figura 42F, H). En segundo lugar, los experimentos en los subclones flipeados demostraron que la modificación se mantenía sobre la inserción después de la eliminación de las secuencias exógenas (Figura 48E, F, H). Estos datos sugieren que probablemente estos cambios en el estatus de H3K9me3 en el genoma adyacente sean dependientes de la inserción. Por tanto, las inserciones *de novo* de LINE-1 son capaces de expandir la modificación H3K9me3 depositada sobre la misma hacia el genoma adyacente.

En cuanto a la distancia que pueden alcanzar estos cambios, es de destacar que en las líneas sensibles pueden alcanzar, de media, las 1427 pares de bases aguas abajo de la inserción. Por otro lado, los cambios aguas arriba generados en el estatus de H3K9me3 llegan hasta las 1100 pares de bases. Tanto la distancia aguas arriba como aguas abajo hasta donde llegan los cambios de enriquecimiento de H3K9me3, son similares a los cambios generados por una inserción polimórfica del retrovirus endógeno IAP en un línea de células embrionarias de ratón (1-2,5kb) [236]. Además, esta distancia concuerda con los enriquecimientos de H3K9me3 encontrados en las regiones flanqueantes de ciertos retrovirus endógenos de clase I y clase II en células embrionarias de ratón (1kb) a los que se le atribuye dichos enriquecimientos [185]. De esta forma, se observa como los cambios generados por la inserción *de novo* de LINE-1 genera cambios a una distancia similar a los cambios observados para otros elementos móviles presentes en el genoma de ratón.

Por otro lado, las inserciones endógenas enriquecidas en H3K9me3 no generaron cambios en la composición de esta modificación en la cromatina adyacente (Figura 55B, D). Sin embargo, esto se puede deber probablemente a la localización genómica de las inserciones. En el caso de la inserción 4-144 sólo se pudo analizar la región aguas abajo de la inserción donde el primer punto en el genoma se encuentra a 6475 pares de bases desde el punto 0. Esto se debe a que la inserción se encuentra dentro de un elemento LINE-1 endógeno lo cual impidió el diseño de oligonucleótidos aguas arriba de la misma. Por otro lado, en el caso de la inserción X-810 su localización hace imposible la determinación de estos cambios pues se trata de una región muy enriquecida en H3K9me3 tanto aguas abajo como aguas arriba de la inserción. Esto es un caso similar al de la línea sensible PA-1-pRAM272 cuya inserción se encuentra en la región pericentromérica donde elevados enriquecimientos de H3K9me3 impiden detectar los posibles cambios generados en el genoma adyacente por la inserción (Figura 40B). Además, ambas inserciones presentaban un nivel de enriquecimiento medio/bajo (4-144) y bajo (X-810) de H3K9me3

(Figura 54A, B) lo cual, por lo ya mostrado para inserciones artificiales, parece que no es compatible con la generación de cambios. Por tanto, la localización y el nivel enriquecimiento de H3K9me3 podrían ser la causa para que no se hayan podido detectar los cambios de enriquecimiento de H3K9me3 en el genoma adyacente alrededor de estas dos inserciones endógenas.

Estos cambios podrían tener una serie de consecuencias en función de la localización de las nuevas inserciones de LINE-1. De hecho, podrían alterar elementos reguladores como pueden ser promotores y enhancers lo cual podría tener un impacto funcional sobre la expresión del gen controlado por dicho elemento. La teoría de la capacidad de los elementos móviles para generar cambios en el genoma adyacente ha estado presente en el campo de la epigenética y de los elementos móviles desde hace tiempo. De hecho, existen varios estudios en el campo de los retrovirus endógenos donde se asocian los cambios en la expresión de genes adyacentes a los cambios epigenéticos que tienen lugar sobre un retrovirus endógeno cercano. En células embrionarias de ratón, la eliminación de elementos relacionados con la represión de retrovirus endógenos tales como la lisina desmetilasa KDM1a, una lisina metil transferasa SETDB1, el represor KAP1 o co-represores de la familia de las proteínas KRAB-ZFP da lugar a la activación de genes vecinos de retrovirus endógenos. Dicha activación está asociada con la ganancia de marcas asociadas a actividad transcripcional (H3K4me3, H3K27ac, H3K4me3, H3ac) y la pérdida de marcas asociadas a represión transcripcional (H3K9me3 y H3K9me2) sobre el retrovirus endógeno cercano. En estos casos, el retrovirus endógeno funciona bien como promotor ya que los transcritos son quiméricos y llevan una parte de la secuencia del retrovirus endógeno [185], [208], [237] o bien funcionan como enhancers ya que la ganancia de marcas propias de enhancers correlaciona con un aumento de la expresión de dichos genes vecinos [209]. Por tanto, estos estudios muestran que existe una relación de dependencia entre la expresión génica y la regulación epigenética de retrovirus endógenos cercanos. Por otro lado, en células embrionarias de ratón los enriquecimientos de H3K9me3 localizados entre 1 y 2kb alrededor de un retrovirus endógeno clase I y clase II se atribuyen a la presencia de dicho retrovirus endógeno [141], [185]. En ninguno de los casos señalados, se puede hablar de epimutaciones además de que todos estos estudios se basan en copias de retrovirus endógenos ya existentes en el genoma murino. Sin embargo, existe un estudio en células embrionarias de ratón que muestra que una inserción polimórfica del retrovirus endógeno IAP genera cambios de H3K9me3 en la cromatina adyacente. Estos cambios alcanzan al gen *BGALTL3* situado a 300 pares de bases del mismo provocando una disminución de su expresión [236]. Tan solo este último estudio demuestra que un elemento móvil murino después de insertarse puede alterar la cromatina adyacente afectando la expresión de un gen cercano. Por tanto, esta tesis es el primer estudio donde

se observan cambios en la composición de modificaciones de histonas en la cromatina adyacente debido a la integración de una inserción *de novo* de un retrotransposón activo en el genoma humano a tiempo real.

Aunque se ha demostrado que el enriquecimiento de H3K9me3 sobre el elemento LINE-1 puede expandirse hacia el genoma adyacente, no se ha podido mostrar cómo estos cambios pueden afectar a la expresión génica. En primer lugar, estos cambios podrían afectar *en cis* a elementos reguladores tales como promotores y enhancers. Sin embargo, la ausencia de elementos reguladores cercanos a las inserciones que generan los cambios (<1-1,5kb) impide validar esta hipótesis. En segundo lugar, estos cambios podrían afectar *en trans* a alguna región del genoma que contuviera un elemento regulador, que debido a la estructura 3D de la cromatina pudiera estar en contacto con la región donde se han generado los cambios de H3K9me3, generándose en él un cambio que tuviera un efecto sobre la expresión génica. Por ello, actualmente se está llevando a cabo un ARN-seq para determinar si existen cambios de expresión de genes *en trans*. De ser así, se intentaría analizar por ChIP-seq los cambios existentes en las zonas candidatas y se intentaría determinar la relación física entre dichas regiones del genoma utilizando la técnica 4C. Por último, se podría especular con que el silenciamiento de una inserción génica *de novo* de LINE-1 y la expansión de dicho silenciamiento podría afectar a la expresión génica mediante la inclusión de un dominio de H3K9me3 dentro del marco abierto de lectura del gen. Este dominio de H3K9me3 podría ralentizar el paso de la polimerasa II durante la transcripción del gen debido a una mayor compactación de la cromatina. Desafortunadamente, esta hipótesis, no puede ser testada en la única inserción génica que induce cambios de H3K9me3 de la línea sensible PA-1-pRAM221 ya que el gen hospedador (*EYA2*) se expresa a un nivel casi indetectable. No obstante, quizás esta hipótesis se podría validar después de diferenciar esta línea clonal en caso de que este gen se expresara a un nivel considerable en estado diferenciado. Además, como parte de otra tesis doctoral del laboratorio, se midió la expresión del gen receptor de la inserción de la línea PA-1-pRAM8 antes y después de la integración. De esta forma, se demostró que la expresión disminuía después de que se produjera la inserción de LINE-1. Esta inserción presenta un elevado enriquecimiento de H3K9me3. Por tanto, se podría atribuir la bajada de la expresión a la existencia de un dominio de H3K9me3 que podría ralentizar el paso de la Pol II. De igual forma, no se puede descartar que los cambios generados se deban a un efecto sobre el splicing o maduración de este gen.

Según los datos provenientes de los experimentos de diferenciación en ambas líneas clonales, parece que los cambios de H3K9me3 generados en el genoma adyacente se podrían mantener a lo largo del proceso de diferenciación (Figura 50C, G). Este dato,

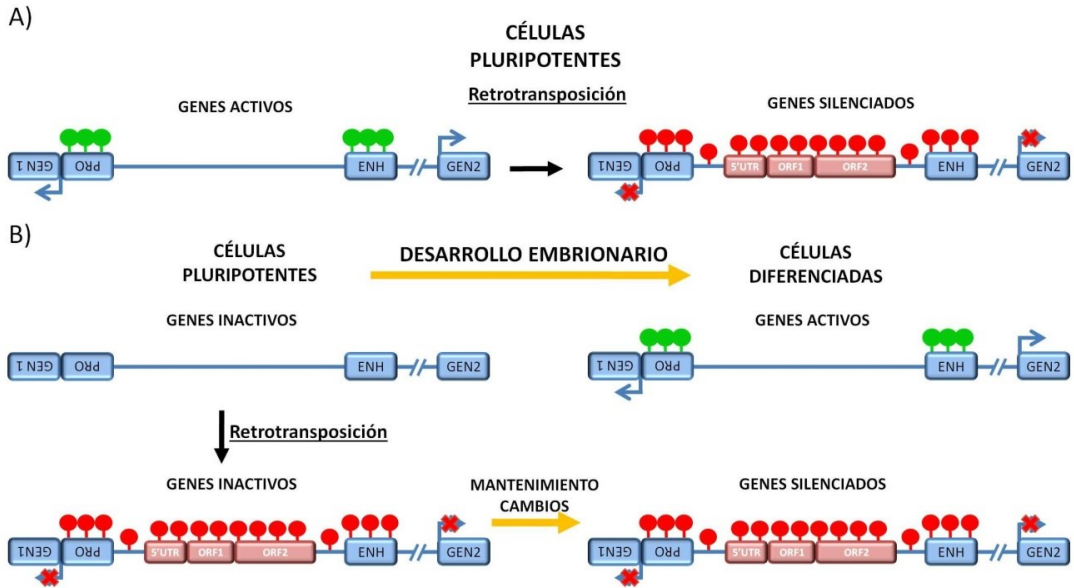


aunque preliminar, dota al proyecto de una importancia adicional ya que inserciones generadas durante el desarrollo embrionario temprano podrían tener un impacto funcional en etapas posteriores del desarrollo. De hecho, inserciones *de novo* de LINE-1 integradas cerca de elementos reguladores requeridos durante el proceso de diferenciación podrían afectar este proceso alterando epigenéticamente dichos elementos. Por tanto, una inserción *de novo* de LINE-1 que ocurriera durante el desarrollo embrionario temprano podría tener un efecto sobre el fenotipo de una célula totalmente diferenciada. Por ello, estos experimentos de diferenciación confieren una nueva perspectiva al proyecto, ya que expanden la ventana temporal durante la cual una inserción *de novo* de LINE-1 podría alterar la epigenética de elementos reguladores vecinos. De hecho, inserciones *de novo* de LINE-1 que se pudieran considerar neutras en las etapas tempranas del desarrollo embrionario temprano podrían ser deletéreas en etapas más tardías cuando se produce la especificación de linaje celular.

Finalmente, después de analizar los resultados obtenidos sobre los cambios epigenéticos en el genoma adyacente se propone el siguiente modelo. Durante el desarrollo embrionario temprano, una inserción *de novo* de LINE-1 podría silenciar un promotor o un enhancer activo cercano específico de células pluripotentes (<1,5Kb) mediante la generación de cambios de enriquecimiento de H3K9me3 en el genoma adyacente. De esta forma, la expresión del gen controlado por dicho promotor o enhancer se vería afectada durante las primeras etapas del desarrollo embrionario (Figura 58A). Del mismo modo, debido al posible mantenimiento de los cambios de H3K9me3 durante el proceso de diferenciación celular, una nueva inserción de LINE-1 podría silenciar un promotor o enhancer cercano específico de células diferenciadas evitando o dificultando así su activación. De esta forma, se evitaría o se afectaría la activación del gen controlado por dicho promotor o enhancer en el momento en el que fuera requerido durante el proceso de diferenciación celular o en un tejido diferenciado (Figura 58B).

En resumen, las inserciones *de novo* de LINE-1 son epigenéticamente silenciadas mediante H3K9me3 y probablemente H3K27me3 después de su integración en un genoma pluripotente. Además, las inserciones artificiales *de novo* pueden generar cambios en el enriquecimiento de H3K9me3 en el genoma adyacente alcanzando una longitud entre 1-1,5kb de distancia. Este proceso parece ser dependiente de la longitud de la inserción así como de la existencia de un contexto transcripcional muy represivo. Asimismo, estos cambios son generados en un contexto de pluripotencia que es semejante al desarrollo embrionario temprano periodo en el cual se acumulan la mayoría de las nuevas inserciones de LINE-1 [77], [78]. Por tanto, estos cambios epigenéticos podrían afectar a elementos reguladores de genes importantes durante esta fase temprana desarrollo embrionario.

Además, estos cambios podrían tener un impacto funcional sobre genes requeridos durante fases más tardías del desarrollo embrionario y alterar la diferenciación celular, ya



**Figura 58: Modelo del efecto de los cambios epigenéticos generados por las inserciones de novo de LINE-1.** A) Una nueva inserción de LINE-1 generada durante el desarrollo embrionario temprano mediante la expansión de H3K9me3 podría silenciar un promotor o enhancer activo cercano afectando la expresión del gen controlado por este promotor o enhancer. B) Una nueva inserción de LINE-1 que se produce en etapas tempranas del desarrollo produce cambios de enriquecimiento de H3K9me3 que posiblemente se mantengan a lo largo del proceso de diferenciación. De esta forma, podría afectar la expresión génica en etapas posteriores del proceso de desarrollo silenciando y evitando la activación de un promotor o enhancer específico de células diferenciadas. Las marcas rojas representan H3K9me3 y las marcas verdes representan modificaciones de actividad propias de promotores y enhancers activos. Las abreviaturas utilizadas son las siguientes: PRO, Promotor; ENH, Enhancer.

que, en principio, estos cambios parecen mantenerse a lo largo del proceso de diferenciación. Contrariamente a lo que se podía pensar, una nueva inserción endógena de LINE-1 después de su integración puede reclutar modificaciones de actividad como H3K4me3. Esto constituye un indicio de que dicha inserción podría ser transcrita de tal forma que pudiera iniciar un nuevo ciclo de retrotransposición. Estos resultados se pueden interpretar desde la perspectiva evolutiva de LINE-1 en la población humana. Según los datos de esta tesis, la mayoría de las inserciones de LINE-1 son silenciadas mediante H3K9me3 y en determinados contextos transcripcionales se producen cambios de H3K9me3 en el genoma adyacente seguramente como mecanismo para reforzar el silenciamiento de la inserción *de novo* de LINE-1. Este mecanismo que implica cambios en

el genoma adyacente puede ser permitido en determinados contextos genómicos donde el efecto y el coste del silenciamiento sobre el genoma hospedador probablemente sea mínimo. Esto es lo que ocurre con las inserciones de las líneas sensibles donde el contexto era altamente represivo. De esta forma, el retroelemento es silenciado pero estos cambios seguramente no tengan un impacto sobre el genoma hospedador manteniéndose el fitness del individuo. De este modo, esta nueva inserción puede ser transmitida a la nueva generación y puede ser fijada en la población. Por otro lado, en algunas ocasiones la localización de la inserción provoca que el coste asociado al silenciamiento y la generación de cambios en el genoma adyacente pueda ser mayor. De este modo, determinados contextos genómicos pueden favorecer que la inserción permanezca activa en lugar de ser silenciada como ocurre con la inserción 1-927 que acumula H3K4me3 sobre su promotor. En este caso, se asume el coste de que pueda volver a retrotransponerse a cambio de que el impacto sobre la viabilidad del individuo sea mínimo. Por tanto, bien siendo silenciada en un contexto donde el impacto sea mínimo para el genoma o bien permaneciendo activa, la inserción se asegura su transmisión a la siguiente generación ya que no se produce efecto sobre el fitness del individuo. En resumen, inserciones con un bajo impacto epigenético serían aquellas que podrían transmitirse a la siguiente generación y podrían finalmente fijarse en la población aumentando el nº de copias de LINE-1 en el genoma humano. Finalmente, los resultados obtenidos pueden explicar parcialmente como las nuevas inserciones de LINE-1 pueden transmitirse a través de las generaciones.



# CONCLUSIONES



## 6. Conclusiones

1. El microprocesador regula negativamente la retrotransposición de LINE-1 a través de la interacción con la región 5' no traducida del ARNm de LINE-1.
2. Las inserciones artificiales *de novo* de LINE-1 acumulan modificaciones de histonas asociadas con represión transcripcional, fundamentalmente H3K9me3 y probablemente H3K27me3. Además, las inserciones endógenas *de novo* de LINE-1 están enriquecidas en H3K9me3.
3. Las inserciones endógenas *de novo* de LINE-1 después de su integración, pueden reclutar modificaciones ligadas a actividad transcripcional como H3K4me3 lo cual sugiere que estarían predispuestas para un nuevo ciclo de retrotransposición en un genoma pluripotente.
4. La regulación epigenética de las inserciones *de novo* de LINE-1 podría depender de la longitud de la inserción y también del contexto transcripcional donde se han integrado.
5. Las inserciones artificiales *de novo* de LINE-1 producen cambios en el enriquecimiento de H3K9me3 en la cromatina adyacente expandiendo el enriquecimiento que existe sobre la propia inserción. Estos cambios pueden expandirse una distancia de 1-1,5 kb desde el sitio aguas arriba y aguas abajo desde el sitio de inserción. Estos cambios son generados por inserciones de copia completa y en un contexto de represión transcripcional.
6. Los cambios epigenéticos inducidos por las inserciones *de novo* de LINE-1 en un genoma pluripotente podrían mantenerse durante la diferenciación celular, lo cual implicaría que inserciones *de novo* de LINE-1 producidas durante el desarrollo embrionario temprano podrían tener un impacto funcional en etapas posteriores o incluso en células diferenciadas.





# BIBLIOGRAFÍA



## 7. Bibliografía

- [1] E. S. Lander *et al.*, “Initial sequencing and analysis of the human genome,” *Nature*, vol. 409, no. 6822, pp. 860–921, 2001.
- [2] C. R. Beck, J. L. Garcia-Perez, R. M. Badge, and J. V Moran, “LINE-1 Elements in Structural Variation and Disease,” *Annu Rev Genomics Hum Genet*, vol. 12, no. 60, pp. 187–215, 2011.
- [3] D. L. Mager and J. P. Stoye, “Mammalian Endogenous Retroviruses,” *Microbiol. Spectr.*, vol. 3, no. 1, pp. 1–20, 2015.
- [4] J. L. Garcia-Perez, A. J. Doucet, H. C. Kopera, S. R. Richardson, J. B. Moldovan, and J. V. Moran, “The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes,” *Microbiol. Spectr.*, vol. 3, no. 2, pp. 1–63, 2015.
- [5] A. F. A. Smit, “The origin of interspersed repeats in the human genome,” *Curr. Opin. Genet. Dev.*, vol. 6, no. 6, pp. 743–748, 1996.
- [6] J. H. Wildschutte, Z. H. Williams, M. Montesion, R. P. Subramanian, J. M. Kidd, and J. M. Coffin, “Discovery of unfixed endogenous retrovirus insertions in diverse human populations,” *Proc. Natl. Acad. Sci.*, vol. 113, no. 16, pp. E2326–E2334, 2016.
- [7] C. Stocking and C. A. Kozak, “Endogenous retroviruses: Murine endogenous retroviruses,” *Cell. Mol. Life Sci.*, vol. 65, no. 21, pp. 3383–3398, 2008.
- [8] A. Katzourakis, A. Rambaut, and O. G. Pybus, “The evolutionary dynamics of endogenous retroviruses,” *Trends Microbiol.*, vol. 13, no. 10, pp. 463–468, 2005.
- [9] O. Hohn, K. Hanke, and N. Bannert, “HERV-K(HML-2), the Best Preserved Family of HERVs: Endogenization, Expression, and Implications in Health and Disease,” *Front. Oncol.*, vol. 3, no. September, pp. 1–12, 2013.
- [10] D. C. Hancks and H. H. Kazazian, “Roles for retrotransposon insertions in human disease,” *Mob. DNA*, vol. 7, no. 1, p. 9, 2016.
- [11] B. Brouha *et al.*, “Evidence Consistent with Human L1 Retrotransposition in Maternal Meiosis I,” *Am. J. Hum. Genet.*, vol. 71, no. 2, pp. 327–336, 2002.
- [12] J. V Moran, S. E. Holmes, T. P. Naas, R. J. DeBerardinis, J. D. Boeke, and H. H. Kazazian, “High frequency retrotransposition in cultured mammalian cells,” *Cell*, vol. 87, no. 5, pp. 917–27, Nov. 1996.
- [13] C. R. Beck *et al.*, “LINE-1 retrotransposition activity in human genomes,” *Cell*, vol. 141, no. 7, pp. 1159–1170, 2010.

- [14] S. Solyom *et al.*, “Extensive somatic L1 retrotransposition in colorectal tumors,” *Genome Res.*, vol. 22, no. 12, pp. 2328–2338, 2012.
- [15] R. C. Iskow *et al.*, “Natural mutagenesis of human genomes by endogenous retrotransposons,” *Cell*, vol. 141, no. 7, pp. 1253–1261, 2010.
- [16] R. Shukla *et al.*, “Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma,” *Cell*, vol. 153, no. 1, pp. 101–111, 2013.
- [17] S. L. Martin and F. D. Bushman, “Martin&Bushman 2001 Nucleic Acid Chaperone Activity of the ORF1 Protein from the Mouse LINE-1.pdf,” vol. 21, no. 2, pp. 467–475, 2001.
- [18] Q. Feng, J. V. Moran, H. H. Kazazian, and J. D. Boeke, “Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition,” *Cell*, vol. 87, no. 5, pp. 905–916, 1996.
- [19] S. L. Mathias, a F. Scott, H. H. Kazazian, J. D. Boeke, and a Gabriel, “Reverse transcriptase encoded by a human transposable element.,” *Science*, vol. 254, no. 5039, pp. 1808–1810, 1991.
- [20] A. J. Doucet, J. E. Wilusz, T. Miyoshi, Y. Liu, and J. V. Moran, “A 3’ Poly(A) Tract Is Required for LINE-1 Retrotransposition,” *Mol. Cell*, vol. 60, no. 5, pp. 728–741, 2015.
- [21] J. V. Moran, “Exon Shuffling by L1 Retrotransposition,” *Science (80-. )*, vol. 283, no. 5407, pp. 1530–1534, 1999.
- [22] R. E. Mills, E. A. Bennett, R. C. Iskow, and S. E. Devine, “Which transposable elements are active in the human genome?,” *Trends Genet.*, vol. 23, no. 4, pp. 183–91, Apr. 2007.
- [23] C. Ade, A. M. Roy-Engel, and P. L. Deininger, “Alu elements: An intrinsic source of human genome instability,” *Curr. Opin. Virol.*, vol. 3, no. 6, pp. 639–645, 2013.
- [24] M. A. Batzer and P. L. Deininger, “Alu Repeats and Human Genomic Diversity,” *Nat. Rev. Genet.*, vol. 3, no. 5, pp. 370–379, 2002.
- [25] P. Deininger *et al.*, “Alu elements: know the SINES,” *Genome Biol.*, vol. 12, no. 12, p. 236, 2011.
- [26] J. Raiz *et al.*, “The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery,” *Nucleic Acids Res.*, vol. 40, no. 4, pp. 1666–1683, 2012.
- [27] J. Rohrer, Y. Minegishi, D. Richter, J. Eguiguren, and M. E. Conley, “Unusual

mutations in Btk: an insertion, a duplication, an inversion, and four large deletions.," *Clin. Immunol.*, vol. 90, no. 1, pp. 28–37, 1999.

- [28] M. E. Conley, J. D. Partain, S. M. Norland, S. A. Shurtleff, and H. H. Kazazian, "Two independent retrotransposon insertions at the same site within the coding region of BTK," *Hum. Mutat.*, vol. 25, no. 3, pp. 324–325, 2005.
- [29] E. M. Ostertag, J. L. Goodier, Y. Zhang, and H. H. Kazazian, "SVA Elements Are Nonautonomous Retrotransposons that Cause Disease in Humans," *Am. J. Hum. Genet.*, vol. 73, no. 6, pp. 1444–1451, 2003.
- [30] D. C. Hancks and H. H. Kazazian, "SVA retrotransposons: Evolution and genetic instability," *Semin. Cancer Biol.*, vol. 20, no. 4, pp. 234–245, 2010.
- [31] M. de Boer *et al.*, "Primary Immunodeficiency Caused by an Exonized Retroposed Gene Copy Inserted in the CYBB Gene," *Hum. Mutat.*, vol. 35, no. 4, pp. 486–496, 2014.
- [32] B. Brouha *et al.*, "Hot L1s account for the bulk of retrotransposition in the human population.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 9, pp. 5280–5285, 2003.
- [33] S. Boissinot, P. Chevret, and A. V Furano, "L1 (LINE-1) retrotransposon evolution and amplification in recent human history.," *Mol. Biol. Evol.*, vol. 17, no. 6, pp. 915–928, 2000.
- [34] A. F. Scott *et al.*, "Origin of the human L1 elements: Proposed progenitor genes deduced from a consensus DNA sequence," *Genomics*, vol. 1, no. 2, pp. 113–125, 1987.
- [35] D. Bourc'his and T. H. Bestor, "Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L," *Nature*, vol. 431, no. 7004, pp. 96–99, 2004.
- [36] S. Wissing *et al.*, "Reprogramming somatic cells into ips cells activates line-1 retroelement mobility," *Hum. Mol. Genet.*, vol. 21, no. 1, pp. 208–218, 2012.
- [37] G. D. Swergold, "Identification, characterization, and cell specificity of a human LINE-1 promoter.," *Mol. Cell. Biol.*, vol. 10, no. 12, pp. 6718–29, 1990.
- [38] R. Minakami, K. Kurose, K. Etoh, Y. Furuhashi, M. Hattori, and Y. Sakaki, "Identification of an internal cis-element essential for the human L1 transcription and a nuclear factor(s) binding to the element," *Nucleic Acids Res.*, vol. 20, no. 12, pp. 3139–3145, 1992.
- [39] J. N. Athanikar, R. M. Badge, and J. V. Moran, "A YY1-binding site is required for accurate human LINE-1 transcription initiation," *Nucleic Acids Res.*, vol. 32, no. 13,

pp. 3846–3855, 2004.

- [40] K. G. Becker, G. Swergold, K. Ozato, and R. E. Thayer, “Binding of the ubiquitous nuclear transcription factor YY1 to a cisregulatory sequence in the human LINE-1 transposable element,” *Hum. Mol. Genet.*, vol. 2, no. 10, pp. 1697–1702, 1993.
- [41] T. Tchénio, J. F. Casella, and T. Heidmann, “Members of the SRY family regulate the human LINE retrotransposons,” *Nucleic Acids Res.*, vol. 28, no. 2, pp. 411–5, 2000.
- [42] N. Yang, L. Zhang, Y. Zhang, and H. H. Kazazian, “An important role for RUNX3 in human L1 transcription and retrotransposition,” *Nucleic Acids Res.*, vol. 31, no. 16, pp. 4929–4940, 2003.
- [43] M. Speek, “Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes,” *Mol. Cell. Biol.*, vol. 21, no. 6, pp. 1973–85, 2001.
- [44] A. Macia *et al.*, “Epigenetic Control of Retrotransposon Expression in Human Embryonic Stem Cells,” *Mol. Cell. Biol.*, vol. 31, no. 2, pp. 300–316, 2011.
- [45] P. Nigumann, K. Redik, K. Mätlik, and M. Speek, “Many Human Genes Are Transcribed from the Antisense Promoter of L1 Retrotransposon,” *Genomics*, vol. 79, no. 5, pp. 628–634, 2002.
- [46] N. Yang and H. H. Kazazian, “L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells,” *Nat. Struct. Mol. Biol.*, vol. 13, no. 9, pp. 763–771, 2006.
- [47] M. Bodak, D. Cirera-Salinas, J. Yu, R. P. Ngondo, and C. Ciaudo, “Dicer, a new regulator of pluripotency exit and LINE-1 elements in mouse embryonic stem cells,” *FEBS Open Bio*, vol. 7, no. 2, pp. 204–220, 2017.
- [48] A. M. Denli *et al.*, “Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity,” *Cell*, vol. 163, no. 3, pp. 583–893, 2015.
- [49] S. L. Martin, D. Branciforte, D. Keller, and D. L. Bain, “Trimeric structure for an essential protein in L1 retrotransposition,” *Proc. Natl. Acad. Sci.*, vol. 100, no. 24, pp. 13815–13820, 2003.
- [50] S. E. H. Yi, M. F. Singersii, and G. D. Swergold, “Studies on,” pp. 19765–19768, 1992.
- [51] E. Khazina and O. Weichenrieder, “Non-LTR retrotransposons encode noncanonical RRM domains in their first open reading frame,” *Proc. Natl. Acad. Sci.*, vol. 106, no. 3, pp. 731–736, 2009.
- [52] S. L. Martin, “The ORF1 protein encoded by LINE-1: Structure and function during L1 retrotransposition,” *J. Biomed. Biotechnol.*, vol. 2006, pp. 1–6, 2006.

- [53] D. A. Kulpa and J. V Moran, "Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles," *Nat. Struct. Mol. Biol.*, vol. 13, no. 7, pp. 655–660, 2006.
- [54] C. Monot *et al.*, "The Specificity and Flexibility of L1 Reverse Transcription Priming at Imperfect T-Tracts," *PLoS Genet.*, vol. 9, no. 5, 2013.
- [55] A. J. Doucet *et al.*, "Characterization of LINE-1 Ribonucleoprotein Particles.," *PLoS Genet.*, vol. 6, no. 10, Jan. 2010.
- [56] K. Usdin and A. V Furano, "The Structure of the Guanine-rich Polyuracil Sequence at the Right End of the Rat L1 ( LINE ) Element \*," vol. 264, no. 26, pp. 15681–15687, 1989.
- [57] R. S. Alisch, J. L. Garcia-Perez, A. R. Muotri, F. H. Gage, and J. V. Moran, "Unconventional translation of mammalian LINE-1 retrotransposons," *Genes Dev.*, vol. 20, no. 2, pp. 210–224, 2006.
- [58] W. E. I. Wei *et al.*, "Human L1 Retrotransposition.," *Society*, vol. 21, no. 4, pp. 1429–1439, 2001.
- [59] J. L. Garcia-Perez *et al.*, "LINE-1 retrotransposition in human embryonic stem cells," *Hum. Mol. Genet.*, vol. 16, no. 13, pp. 1569–1577, 2007.
- [60] J. L. Garcia-Perez *et al.*, "Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells," *Nature*, vol. 466, no. 7307, pp. 769–773, 2010.
- [61] S. L. Martin, "Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells.," *Mol. Cell. Biol.*, vol. 11, no. 9, pp. 4804–7, 1991.
- [62] M. S. Taylor *et al.*, "Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition," *Cell*, vol. 155, no. 5, pp. 1034–1048, 2013.
- [63] J. L. Goodier, L. E. Cheung, and H. H. Kazazian, "Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition," *Nucleic Acids Res.*, vol. 41, no. 15, pp. 7401–7419, 2013.
- [64] P. K. Mandal, A. D. Ewing, D. C. Hancks, and H. H. Kazazian, "Enrichment of processed pseudogene transcripts in L1-ribonucleoprotein particles," *Hum. Mol. Genet.*, vol. 22, no. 18, pp. 3730–3748, 2013.
- [65] D. D. Luan, M. H. Korman, J. L. Jakubczak, and T. H. Eickbush, "Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition," *Cell*, vol. 72, no. 4, pp. 595–605, 1993.

- [66] G. J. Cost, Q. Feng, A. Jacquier, and J. D. Boeke, "Human L1 element target-primed reverse transcription in vitro," *EMBO J.*, vol. 21, no. 21, pp. 5899–5910, 2002.
- [67] J. Jurka, "Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons," *Proc. Natl. Acad. Sci.*, vol. 94, no. 5, pp. 1872–1877, 1997.
- [68] S. L. Gasior, T. P. Wakeman, B. Xu, and P. L. Deininger, "The human LINE-1 retrotransposon creates DNA double-strand breaks," *J. Mol. Biol.*, vol. 357, no. 5, pp. 1383–93, Apr. 2006.
- [69] S. L. Gasior, A. M. Roy-Engel, and P. L. Deininger, "ERCC1/XPF limits L1 retrotransposition," *DNA Repair (Amst.)*, vol. 7, no. 6, pp. 983–989, 2008.
- [70] N. G. Coufal *et al.*, "Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells," *Proc. Natl. Acad. Sci.*, vol. 108, no. 51, pp. 20382–20387, 2011.
- [71] C. Philippe *et al.*, "Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci," *Elife*, vol. 5, no. MARCH2016, pp. 1–30, 2016.
- [72] M. Friedli *et al.*, "Retroelements During Reprogramming To Pluripotency," vol. 1, no. 1, pp. 1251–1259, 2014.
- [73] S. Klawitter *et al.*, "Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells," *Nat. Commun.*, vol. 7, no. JANUARY, p. 10286, 2016.
- [74] S. A. Trelogan and S. L. Martin, "Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis," *Proc. Natl. Acad. Sci.*, vol. 92, no. 5, pp. 1520–1524, 1995.
- [75] D. Branciforte and S. L. Martin, "Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition.," *Mol. Cell. Biol.*, vol. 14, no. 4, pp. 2584–2592, 1994.
- [76] I. Georgiou *et al.*, "Retrotransposon RNA expression and evidence for retrotransposition events in human oocytes," *Hum. Mol. Genet.*, vol. 18, no. 7, pp. 1221–1228, 2009.
- [77] H. Kano *et al.*, "L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism," *Genes Dev.*, vol. 23, no. 11, pp. 1303–1312, 2009.
- [78] J. A. J. M. van den Hurk *et al.*, "L1 retrotransposition can occur early in human embryonic development," *Hum. Mol. Genet.*, vol. 16, no. 13, pp. 1587–1592, 2007.



- [79] A. Macla *et al.*, “Engineered LINE-1 retrotransposition in nondividing human neurons,” *Genome Res.*, vol. 27, no. 3, pp. 335–348, 2017.
- [80] N. G. Coufal *et al.*, “L1 retrotransposition in human neural progenitor cells,” *Nature*, vol. 460, no. 7259, pp. 1127–1131, 2009.
- [81] G. D. Evrony *et al.*, “Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain,” *Cell*, vol. 151, no. 3, pp. 483–496, 2012.
- [82] J. K. Baillie *et al.*, “Somatic retrotransposition alters the genetic landscape of the human brain,” *Nature*, vol. 479, no. 7374, pp. 534–537, 2011.
- [83] K. R. Upton *et al.*, “Ubiquitous L1 mosaicism in hippocampal neurons,” *Cell*, vol. 161, no. 2, pp. 228–239, 2015.
- [84] N. Rodić *et al.*, “Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma,” no. August, 2015.
- [85] J. M. C. Tubio and J. M. C. Tubio, “Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes,” 2014.
- [86] H. H. Kazazian, C. Wong, H. Youssofian, A. F. Scott, D. G. Phillips, and S. E. Antonarakis, “Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man,” *Nature*, vol. 332, no. 6160, pp. 164–166, 1988.
- [87] V. P. Belancio, D. J. Hedges, and P. Deininger, “LINE-1 RNA splicing and influences on mammalian gene expression,” *Nucleic Acids Res.*, vol. 34, no. 5, pp. 1512–1521, 2006.
- [88] C. Rodríguez-Martín *et al.*, “Familial retinoblastoma due to intronic LINE-1 insertion causes aberrant and noncanonical mRNA splicing of the RB1 gene,” *J. Hum. Genet.*, vol. 61, no. 5, pp. 463–466, 2016.
- [89] J. A. J. M. Van Den Hurk *et al.*, “Novel types of mutation in the choroideremia (CHM) gene: A full-length L1 insertion and an intronic mutation activating a cryptic exon,” *Hum. Genet.*, vol. 113, no. 3, pp. 268–275, 2003.
- [90] E. Kondo-Lida *et al.*, “Novel mutations and genotype-phenotype relationships in 107 families with Fukuyama-type congenital muscular dystrophy (FCMD),” *Hum. Mol. Genet.*, vol. 8, no. 12, pp. 2303–2309, 1999.
- [91] J. L. Goodier and H. H. Kazazian, “Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites,” *Cell*, vol. 135, no. 1, pp. 23–35, 2008.
- [92] N. Gilbert, S. Lutz-Prigge, and J. V Moran, “Genomic deletions created upon LINE-1

- retrotransposition.,” *Cell*, vol. 110, no. 3, pp. 315–25, Aug. 2002.
- [93] N. Gilbert, S. Lutz, T. A. Morrish, and J. V. Moran, “Multiple Fates of L1 Retrotransposition Intermediates in Cultured Human Cells,” *Mol. Cell. Biol.*, vol. 25, no. 17, pp. 7780–7795, 2005.
- [94] D. E. Symer *et al.*, “Human l1 retrotransposition is associated with genetic instability in vivo.,” *Cell*, vol. 110, no. 3, pp. 327–38, Aug. 2002.
- [95] T. Arnedo *et al.*, “Mutant GlialCAM causes megalencephalic leukoencephalopathy with subcortical cysts, benign familial macrocephaly, and macrocephaly with retardation and autism,” *Ann. Neurol.*, vol. 88, no. 6, pp. 41–48, 2014.
- [96] T. A. Morrish *et al.*, “DNA repair mediated by endonuclease-independent LINE-1 retrotransposition,” *Nat. Genet.*, vol. 31, no. 2, pp. 159–165, 2002.
- [97] S. K. Sen, C. T. Huang, K. Han, and M. A. Batzer, “Endonuclease-independent insertion provides an alternative pathway for L1 retrotransposition in the human genome,” *Nucleic Acids Res.*, vol. 35, no. 11, pp. 3741–3751, 2007.
- [98] N. Morisada *et al.*, “Branchio-oto-renal syndrome caused by partial EYA1 deletion due to LINE-1 insertion,” *Pediatr. Nephrol.*, vol. 25, no. 7, pp. 1343–1348, 2010.
- [99] J. L. Goodier, “Transduction of 3’-flanking sequences is common in L1 retrotransposition,” *Hum. Mol. Genet.*, vol. 9, no. 4, pp. 653–657, 2000.
- [100] O. K. Pickeral, W. Makalowski, M. S. Boguski, and J. D. Boeke, “Frequent human genomic DNA transduction driven by line-1 retrotransposition,” *Genome Res.*, vol. 10, no. 4, pp. 411–415, 2000.
- [101] S. E. Holmes, B. A. Dombroski, C. M. Krebs, C. D. Boehm, and H. H. Kazazian, “A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion,” *Nat. Genet.*, vol. 7, no. 2, pp. 143–148, 1994.
- [102] J. Chen, A. Rattner, and J. Nathans, “Effects of L1 retrotransposon insertion on transcript processing, localization and accumulation: Lessons from the retinal degeneration 7 mouse and implications for the genomic ecology of L1 elements,” *Hum. Mol. Genet.*, vol. 15, no. 13, pp. 2146–2156, 2006.
- [103] A. Kapusta *et al.*, “Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs,” *PLoS Genet.*, vol. 9, no. 4, 2013.
- [104] T. R. Mercer, M. E. Dinger, and J. S. Mattick, “Insights Into Functions,” *Nat. Rev. Genet.*, vol. 10, no. February, pp. 155–159, 2009.

- [105] V. A. Moran, R. J. Perera, and A. M. Khalil, "Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs," *Nucleic Acids Res.*, vol. 40, no. 14, pp. 6391–6400, 2012.
- [106] X. Lu *et al.*, "The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity," *Nat. Struct. Mol. Biol.*, vol. 21, no. 4, pp. 423–425, 2014.
- [107] A. Fort *et al.*, "Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance," *Nat. Genet.*, vol. 46, no. 6, pp. 558–566, 2014.
- [108] S. J. Wheelan, Y. Aizawa, J. S. Han, and J. D. Boeke, "Gene-breaking: A new paradigm for human retrotransposon-mediated gene evolution," *Genome Res.*, vol. 15, no. 8, pp. 1073–1078, 2005.
- [109] D. M. Sayah, E. Sokolskaja, L. Berthoux, and J. Luban, "Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1," *Nature*, vol. 430, no. 6999, pp. 569–573, 2004.
- [110] D. V. Babushok *et al.*, "A novel testis ubiquitin-binding protein gene arose by exon shuffling in hominoids," *Genome Res.*, vol. 17, no. 8, pp. 1129–1138, 2007.
- [111] K. Ohshima and K. Igarashi, "Inference for the initial stage of domain shuffling: Tracing the evolutionary fate of the PIPSL retrogene in hominoids," *Mol. Biol. Evol.*, vol. 27, no. 11, pp. 2522–2533, 2010.
- [112] B. Dombroski, S. Mathias, E. Nanthakumar, A. Scott, and H. Kazazian, "Isolation of an active human transposable element," *Science (80-. )*, vol. 254, no. 5039, pp. 1805–1808, 1991.
- [113] J. B. Moldovan and J. V. Moran, "The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition," *PLoS Genet.*, vol. 11, no. 5, pp. 1–34, 2015.
- [114] A. R. Muotri, V. T. Chu, M. C. N. Marchetto, W. Deng, J. V. Moran, and F. H. Gage, "Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition," *Nature*, vol. 435, no. 7044, pp. 903–910, 2005.
- [115] E. M. Ostertag, E. T. Prak, R. J. DeBerardinis, J. V. Moran, and H. H. Kazazian Jr., "Determination of L1 retrotransposition kinetics in cultured cells," *Nucleic Acids Res.*, vol. 28, no. 6, pp. 1418–1423, 2000.
- [116] Y. Xie, J. M. Rosser, T. L. Thompson, J. D. Boeke, and W. An, "Characterization of L1 retrotransposition with high-throughput dual-luciferase assays," *Nucleic Acids Res.*, vol. 39, no. 3, pp. 1–11, 2011.

- [117] A. Bulut-Karslioglu *et al.*, “Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells,” *Mol. Cell*, vol. 55, no. 2, pp. 277–290, 2014.
- [118] J. L. Goodier, “Restricting retrotransposons: a review,” *Mob. DNA*, vol. 7, no. 1, p. 16, 2016.
- [119] S. R. Heras, S. Macias, and J. L. Garcia-perez, “activities,” pp. 1–6, 2014.
- [120] M. Hamdorf *et al.*, “miR-128 represses L1 retrotransposition by binding directly to L1 RNA,” *Nat. Struct. Mol. Biol.*, vol. 22, no. 10, pp. 824–831, 2015.
- [121] H. S. Soifer, A. Zaragoza, M. Peyvan, M. A. Behlke, and J. J. Rossi, “A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon,” *Nucleic Acids Res.*, vol. 33, no. 3, pp. 846–856, 2005.
- [122] J. C. Chow *et al.*, “LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation,” *Cell*, vol. 141, no. 6, pp. 956–969, 2010.
- [123] V. N. Kim, J. Han, and M. C. Siomi, “Biogenesis of small RNAs in animals,” *Nat. Rev. Mol. Cell Biol.*, vol. 10, no. 2, pp. 126–139, 2009.
- [124] S. Macias, M. Plass, A. Stajuda, G. Michlewski, E. Eyraes, and J. F. Cáceres, “DGCR8 HITS-CLIP reveals novel functions for the Microprocessor,” *Nat. Struct. Mol. Biol.*, vol. 19, no. 8, pp. 760–766, 2012.
- [125] H. Muckenfuss *et al.*, “APOBEC3 proteins inhibit human LINE-1 retrotransposition,” *J. Biol. Chem.*, vol. 281, no. 31, pp. 22161–22172, 2006.
- [126] S. R. Richardson, I. Narvaiza, R. A. Planegger, M. D. Weitzman, and J. V. Moran, “APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition,” *Elife*, vol. 3, p. e02008, 2014.
- [127] S. Wissing, M. Montano, J. L. Garcia-Perez, J. V. Moran, and W. C. Greene, “Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells,” *J. Biol. Chem.*, vol. 286, no. 42, pp. 36427–36437, 2011.
- [128] K. Zhao *et al.*, “Modulation of LINE-1 and Alu/SVA Retrotransposition by Aicardi-Gouti Syndrome-Related SAMHD1,” *Cell Rep.*, vol. 4, no. 6, pp. 1108–1115, 2013.
- [129] S. Hu *et al.*, “SAMHD1 Inhibits LINE-1 Retrotransposition by Promoting Stress Granule Formation,” *PLoS Genet.*, vol. 11, no. 7, pp. 1–27, 2015.
- [130] D. B. Stetson, J. S. Ko, T. Heidmann, and R. Medzhitov, “Trex1 Prevents Cell-Intrinsic

- Initiation of Autoimmunity," *Cell*, vol. 134, no. 4, pp. 587–598, 2008.
- [131] T. Kouzarides, "Chromatin Modifications and Their Function," *Cell*, vol. 128, no. 4, pp. 693–705, 2007.
- [132] B. E. Bernstein, A. Meissner, and E. S. Lander, "The Mammalian Epigenome," *Cell*, vol. 128, no. 4, pp. 669–681, 2007.
- [133] A. J. Bannister and T. Kouzarides, "Regulation of chromatin by histone modifications," *Cell Res.*, vol. 21, no. 3, pp. 381–395, 2011.
- [134] B. E. Bernstein *et al.*, "A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells," *Cell*, vol. 125, no. 2, pp. 315–326, 2006.
- [135] N. D. Heintzman *et al.*, "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome," *Nat. Genet.*, vol. 39, no. 3, pp. 311–318, 2007.
- [136] M. G. Guenther, S. S. Levine, L. A. Boyer, R. Jaenisch, and R. A. Young, "A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells," *Cell*, vol. 130, no. 1, pp. 77–88, 2007.
- [137] A. Barski *et al.*, "High-Resolution Profiling of Histone Methylations in the Human Genome," *Cell*, vol. 129, no. 4, pp. 823–837, 2007.
- [138] G. Hon, W. Wang, and B. Ren, "Discovery and annotation of functional chromatin signatures in the human genome," *PLoS Comput. Biol.*, vol. 5, no. 11, 2009.
- [139] J. et al Ernst, "Mapping and Analysis of Chromatin State Dynamics in Nine Human Cell Types.," *Nature*, vol. 473, no. 7345, pp. 43–49, 2011.
- [140] Z. Wang *et al.*, "Combinatorial patterns of histone acetylations and methylations in the human genome," *Nat. Genet.*, vol. 40, no. 7, pp. 897–903, 2008.
- [141] T. S. Mikkelsen *et al.*, "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells," *Nature*, vol. 448, no. 7153, pp. 553–560, 2007.
- [142] J. Ernst and M. Kellis, "Discovery and characterization of chromatin states for systematic annotation of the human genome," *Nat. Biotechnol.*, vol. 28, no. 8, pp. 817–825, 2010.
- [143] V. W. Zhou, A. Goren, and B. E. Bernstein, "Charting histone modifications and the functional organization of mammalian genomes.," *Nat. Rev. Genet.*, vol. 12, no. 1, pp. 7–18, 2011.
- [144] K. Linblad-Toh *et al.*, "29 Mammals Used To Find Conserved Elements," *Nature*, vol. 478, no. 7370, pp. 476–482, 2011.

- [145] J. Ernst *et al.*, "Mapping and analysis of chromatin state dynamics in nine human cell types," *Nature*, vol. 473, no. 7345, pp. 43–49, 2011.
- [146] A. Rotem *et al.*, "Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state," *Nat. Biotechnol.*, vol. 33, no. 11, pp. 1165–1172, 2015.
- [147] J. S. Becker, D. Nicetto, and K. S. Zaret, "H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes," *Trends Genet.*, vol. 32, no. 1, pp. 29–41, 2016.
- [148] E. M. Blackwood, "Going the Distance: A Current View of Enhancer Action," *Science (80-. )*, vol. 281, no. 5373, pp. 60–63, 1998.
- [149] C. Koch *et al.*, "The landscape of histone modifications across 1% of the human genome in five human cell lines," *Genome Res.*, vol. 17, no. 6, pp. 691–707, 2007.
- [150] T. H. Kim *et al.*, "A high-resolution map of active promoters in the human genome," *Nature*, vol. 436, no. 7052, pp. 876–880, 2005.
- [151] K. Karmodiya, A. R. Krebs, M. Oulad-Abdelghani, H. Kimura, and L. Tora, "H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells.," *BMC Genomics*, vol. 13, no. 1, p. 424, 2012.
- [152] M. P. Creighton *et al.*, "Histone H3K27ac separates active from poised enhancers and predicts developmental state.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 50, pp. 21931–21936, 2010.
- [153] A. Rada-Iglesias, R. Bajpai, T. Swigut, S. A. Brugmann, R. A. Flynn, and J. Wysocka, "A unique chromatin signature uncovers early developmental enhancers in humans," *Nature*, vol. 470, no. 7333, pp. 279–283, 2011.
- [154] G. C. Hon, R. D. Hawkins, and B. Ren, "Predictive chromatin signatures in the mammalian genome," *Hum. Mol. Genet.*, vol. 18, no. R2, pp. 12–14, 2009.
- [155] S. Schwartz, E. Meshorer, and G. Ast, "Chromatin organization marks exon-intron structure.," *Nat. Struct. & Mol. Biol.*, vol. 16, no. 9, pp. 990–995, 2009.
- [156] P. Kolasinska-Zwierz, T. Down, I. Latorre, T. Liu, X. S. Liu, and J. Ahringer, "Differential chromatin marking of introns and expressed exons by H3K36me3," *Nat. Genet.*, vol. 41, no. 3, pp. 376–381, 2009.
- [157] K. Cui *et al.*, "Chromatin Signatures in Multipotent Human Hematopoietic Stem Cells Indicate the Fate of Bivalent Genes during Differentiation," *Cell Stem Cell*, vol. 4, no. 1, pp. 80–93, 2009.
- [158] V. Azuara *et al.*, "Chromatin signatures of pluripotent cell lines," *Nat. Cell Biol.*, vol.

8, no. 5, pp. 532–538, 2006.

- [159] G. Pan *et al.*, “Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells,” *Cell Stem Cell*, vol. 1, no. 3, pp. 299–312, 2007.
- [160] X. D. Zhao *et al.*, “Whole-Genome Mapping of Histone H3 Lys4 and 27 Trimethylations Reveals Distinct Genomic Compartments in Human Embryonic Stem Cells,” *Cell Stem Cell*, vol. 1, no. 3, pp. 286–298, 2007.
- [161] S. Bilodeau, M. H. Kagey, G. M. Frampton, P. B. Rahl, and R. A. Young, “SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state,” *Genes Dev.*, vol. 23, no. 21, pp. 2484–2489, 2009.
- [162] W. Xie *et al.*, “Epigenomic analysis of multilineage differentiation of human embryonic stem cells,” *Cell*, vol. 153, no. 5, pp. 1134–1148, 2013.
- [163] R. D. Hawkins *et al.*, “Distinct epigenomic landscapes of pluripotent and lineage-committed human cells,” *Cell Stem Cell*, vol. 6, no. 5, pp. 479–491, 2010.
- [164] F. M. Pauler *et al.*, “H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome,” *Genome Res.*, vol. 19, no. 2, pp. 221–233, 2009.
- [165] J. Zhu *et al.*, “Genome-wide chromatin state transitions associated with developmental and environmental cues,” *Cell*, vol. 152, no. 3, pp. 642–654, 2013.
- [166] B. Wen, H. Wu, and Y. Shinkai, “Large organized chromatin K9-modifications (LOCKS) distinguish differentiated from embryonic stem cells,” *Nat. Genet.*, vol. 41, no. 2, pp. 246–250, 2009.
- [167] B. Wen, H. Wu, Y. Shinkai, R. A. Irizarry, and A. P. Feinberg, “Reply\_Reassessing the abundance of H3K9me2 chromatin domains in embryonic stem cells,” *Nat. Genet.*, vol. 42, no. 1, pp. 9–10, 2010.
- [168] F. Ugarte *et al.*, “Progressive chromatin condensation and H3K9 methylation regulate the differentiation of embryonic and hematopoietic stem cells,” *Stem Cell Reports*, vol. 5, no. 5, pp. 728–740, 2015.
- [169] L. Guelen *et al.*, “Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions,” *Nature*, vol. 453, no. 7197, pp. 948–951, 2008.
- [170] A. H. F. M. Peters *et al.*, “Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin,” *Mol. Cell*, vol. 12, no. 6, pp. 1577–1589, 2003.

- [171] M. Piast, I. Kustrzeba-Wójcicka, M. Matusiewicz, and T. Banaś, "Molecular evolution of enolase," *Acta Biochim. Pol.*, vol. 52, no. 2, pp. 507–513, 2005.
- [172] G. H. Karpen and R. C. Allshire, "The case for epigenetic effects on centromere identity and function," *Trends Genet.*, vol. 13, no. 12, pp. 489–496, 1997.
- [173] J. A. Rosenfeld, Z. Wang, D. E. Schones, K. Zhao, R. DeSalle, and M. Q. Zhang, "Determination of enriched histone modifications in non-genic portions of the human genome," *BMC Genomics*, vol. 10, no. 1, p. 143, 2009.
- [174] G. Schotta *et al.*, "A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin," *Genes Dev.*, vol. 18, no. 11, pp. 1251–1262, 2004.
- [175] E. H. Blackburn, "Switching and signaling at the telomere," *Cell*, vol. 106, no. 6, pp. 661–673, 2001.
- [176] M. García-Cao, R. O'Sullivan, A. H. F. M. Peters, T. Jenuwein, and M. A. Blasco, "Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases," *Nat. Genet.*, vol. 36, no. 1, pp. 94–99, 2004.
- [177] H. Santos-Rosa *et al.*, "Active genes are tri-methylated at K4 of histone H3," *Nature*, vol. 419, no. 6905, pp. 407–411, 2002.
- [178] R. Schneider, A. J. Bannister, F. A. Myers, A. W. Thorne, C. Crane-Robinson, and T. Kouzarides, "Histone H3 lysine 4 methylation patterns in higher eukaryotic genes," *Nat. Cell Biol.*, vol. 6, no. 1, pp. 73–77, 2004.
- [179] T. Zhang, S. Cooper, and N. Brockdorff, "The interplay of histone modifications - writers that read," *EMBO Rep.*, vol. 16, no. 11, pp. 1467–1481, 2015.
- [180] D. Schübeler *et al.*, "The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote," *Genes Dev.*, vol. 18, no. 11, pp. 1263–1271, 2004.
- [181] T. Agalioti, G. Chen, and D. Thanos, "Deciphering the transcriptional histone acetylation code for a human gene," *Cell*, vol. 111, no. 3, pp. 381–392, 2002.
- [182] G. Taylor, R. Eskeland, and B. Hekimoglu-balkan, "H4K16 acetylation marks active genes and enhancers of embryonic stem cells , but does not alter chromatin compaction H4K16 acetylation marks active genes and enhancers of embryonic stem cells , but does not alter chromatin compaction," pp. 2053–2065, 2013.
- [183] K. Regha *et al.*, "Active and Repressive Chromatin Are Interspersed without Spreading in an Imprinted Gene Cluster in the Mammalian Genome," *Mol. Cell*, vol. 27, no. 3, pp. 353–366, 2007.



- [184] H. Li, T. Rauch, Z. X. Chen, P. E. Szabó, A. D. Riggs, and G. P. Pfeifer, "The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells," *J. Biol. Chem.*, vol. 281, no. 28, pp. 19489–19500, 2006.
- [185] M. M. Karimi *et al.*, "DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mescs," *Cell Stem Cell*, vol. 8, no. 6, pp. 676–687, 2011.
- [186] S. Ait-Si-Ali *et al.*, "A Suv39h-dependent mechanism for silencing S-phase genes in differentiating but not in cycling cells," *EMBO J.*, vol. 23, no. 3, pp. 605–615, 2004.
- [187] L. Vandel, E. Nicolas, O. Vaute, and R. Ferreira, "Transcriptional Repression by the Retinoblastoma Protein through the Recruitment of a Histone Methyltransferase Transcriptional Repression by the Retinoblastoma Protein through the Recruitment of a Histone Methyltransferase," *Society*, vol. 21, no. 19, pp. 6484–6494, 2001.
- [188] J. Lee *et al.*, "Monoallele deletion of CBP leads to pericentromeric heterochromatin condensation through ESET expression and histone H3 (K9) methylation," *Hum. Mol. Genet.*, vol. 17, no. 12, pp. 1774–1782, 2008.
- [189] L. Fritsch *et al.*, "A Subset of the Histone H3 Lysine 9 Methyltransferases Suv39h1, G9a, GLP, and SETDB1 Participate in a Multimeric Complex," *Mol. Cell*, vol. 37, no. 1, pp. 46–56, 2010.
- [190] X. Cui *et al.*, "Control of transposon activity by a histone H3K4 demethylase in rice.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, no. 5, pp. 1953–8, 2013.
- [191] L. A. Boyer *et al.*, "Polycomb complexes repress developmental regulators in murine embryonic stem cells," *Nature*, vol. 441, no. 7091, pp. 349–353, 2006.
- [192] T. I. Lee *et al.*, "Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells," *Cell*, vol. 125, no. 2, pp. 301–313, 2006.
- [193] a Kirmizis, B. Sm, a Kuzmichev, R. Margueron, and D. Reinberg, "Silencing of human polycomb target genes is associated with methylation of histone H3 Lys[thinsp]27," *Genes Dev.*, vol. 18, pp. 1592–1605, 2004.
- [194] A. Kuzmichev, K. Nishioka, H. Edrjument-Bromage, P. Tempst, and D. Reinberg, "Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein," *Genes ...*, pp. 2893–2905, 2002.
- [195] N. D. Montgomery *et al.*, "The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation," *Curr. Biol.*, vol. 15, no. 10, pp. 942–947, 2005.

- [196] R. Margueron and D. Reinberg, "The Polycomb complex PRC2 and its mark in life," *Nature*, vol. 469, no. 7330, pp. 343–349, 2011.
- [197] A. P. Bracken, N. Dietrich, D. Pasini, K. H. Hansen, and K. Helin, "roles in cell fate transitions Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions," *Genes Dev.*, pp. 1123–1136, 2006.
- [198] A. . Fallis, "An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells," *An epigenetic Switch ensures transposon repression upon Dyn. loss DNA methylation embryonic stem cells*, vol. 53, no. 9, pp. 1689–1699, 2013.
- [199] T. Matsui *et al.*, "Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET.," *Nature*, vol. 464, no. 7290, pp. 927–931, 2010.
- [200] L. K. Hutnick, X. Huang, T. C. Loo, Z. Ma, and G. Fan, "Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism," *J. Biol. Chem.*, vol. 285, no. 27, pp. 21082–21091, 2010.
- [201] D. Wolf and S. P. Goff, "TRIM28 Mediates Primer Binding Site-Targeted Silencing of Murine Leukemia Virus in Embryonic Cells," *Cell*, vol. 131, no. 1, pp. 46–57, 2007.
- [202] D. Wolf and S. P. Goff, "Embryonic stem cells use ZFP809 to silence retroviral DNAs," *Nature*, vol. 458, no. 7242, pp. 1201–1204, 2009.
- [203] D. C. Schultz, K. Ayyanathan, D. Negorev, G. G. Maul, and F. J. R. Iij, "SETDB1 : a novel KAP-1-associated histone H3 , lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins," pp. 919–932, 2002.
- [204] H. M. Rowe *et al.*, "KAP1 controls endogenous retroviruses in embryonic stem cells," *Nature*, vol. 463, no. 7278, pp. 237–240, 2010.
- [205] A. M. Silva *et al.*, "The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses," pp. 538–554, 2007.
- [206] M. Leeb, D. Pasini, M. Novatchkova, M. Jaritz, K. Helin, and A. Wutz, "Polycomb complexes act redundantly to repress genomic repeats and genes," *Genes Dev.*, vol. 24, no. 3, pp. 265–276, 2010.
- [207] S. Daujat *et al.*, "H3K64 trimethylation marks heterochromatin and is dynamically remodeled during developmental reprogramming," *Nat. Struct. Mol. Biol.*, vol. 16, no. 7, pp. 777–781, 2009.
- [208] T. S. Macfarlan *et al.*, "Endogenous retroviruses and neighboring genes are

- coordinately repressed by LSD1/KDM1A," *Genes Dev.*, vol. 25, no. 6, pp. 594–607, 2011.
- [209] H. M. Rowe *et al.*, "TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells," pp. 452–461, 2013.
- [210] N. Castro-Diaz *et al.*, "Evolutionally dynamic L1 regulation in embryonic stem cells Evolutionally dynamic L1 regulation in embryonic stem cells," *Genes Dev.*, vol. 28, no. 13, pp. 1397–1409, 2014.
- [211] J. H. A. Martens *et al.*, "The profile of repeat-associated histone lysine methylation states in the mouse epigenome.," *EMBO J.*, vol. 24, no. 4, pp. 800–12, 2005.
- [212] A. Vaheerap, "Characterization of a Human Ovarian Teratocarcinoma-Derived Cell Line," *Time*, vol. 32, pp. 19–32, 1980.
- [213] C. Gao, M. Miyazaki, J. W. Li, T. Tsuji, Y. Inoue, and M. Namba, "Cytogenetic characteristics and p53 gene status of human teratocarcinoma PA-1 cells in 407-445 passages.," *Int. J. Mol. Med.*, vol. 4, no. 6, pp. 597–600, 1999.
- [214] J. A. et. al. Thomson, "Embryonic Stem Cell Lines Derived from Human Blastocysts," *Science (80-. )*, vol. 282, no. 5391, pp. 1145–1147, 1998.
- [215] G. Chen *et al.*, "Chemically defined conditions for human iPS cell derivation and culture," *Nat. Methods*, vol. 8, no. 5, pp. 424–429, 2011.
- [216] F. Buchholz, P. O. Angrand, and a F. Stewart, "Improved properties of FLP recombinase evolved by cycling mutagenesis.," *Nat. Biotechnol.*, vol. 16, no. 7, pp. 657–662, 1998.
- [217] K. S. O. Shea, J. V Moran, H. P. Bogerd, H. L. Wiegand, A. E. Hulme, and B. R. Cullen, "Cellular inhibitors of long interspersed element 1 and Alu retrotransposition," 2006.
- [218] J. Han, Y. Lee, K. Yeom, Y. Kim, H. Jin, and V. N. Kim, "The Drosha – DGCR8 complex in primary microRNA processing," pp. 3016–3027, 2004.
- [219] J. C. Hogan Jr, "© 199 7 Nature Publishing Group <http://www.nature.com/naturebiotechnology>," *Nat. Biotechnol.*, vol. 15, pp. 328–330, 1997.
- [220] L. He, G. J. Hannon, and C. S. Harbor, "MicroRNAs : SMALL RNAs WITH A BIG ROLE IN GENE REGULATION," vol. 5, no. July, 2004.
- [221] J. M. Sperger *et al.*, "Gene expression patterns in human embryonic stem cells and

- human pluripotent germ cell tumors.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 23, pp. 13350–5, 2003.
- [222] K. Watanabe *et al.*, "A ROCK inhibitor permits survival of dissociated human embryonic stem cells," vol. 25, no. 6, pp. 681–686, 2007.
- [223] G. Meng, S. Liu, and D. E. Rancourt, "Synergistic Effect of Medium, Matrix, and Exogenous Factors on the Adhesion and Growth of Human Pluripotent Stem Cells Under Defined, Xeno-Free Conditions," *Stem Cells Dev.*, vol. 21, no. 11, pp. 2036–2048, 2012.
- [224] A. Huna *et al.*, "Role of stress-activated OCT4A in the cell fate decisions of embryonal carcinoma cells treated with etoposide," *Cell Cycle*, vol. 14, no. 18, pp. 2969–2984, 2015.
- [225] A. M. Denli, B. B. J. Tops, and R. H. A. Plasterk, "Processing of primary microRNAs by the Microprocessor complex," pp. 231–235, 2004.
- [226] R. I. Gregory, K. Yan, and G. Amuthan, "The Microprocessor complex mediates the genesis of microRNAs," vol. 208, no. 2003, pp. 2053–2056, 2004.
- [227] J. Han *et al.*, "Molecular Basis for the Recognition of Primary microRNAs by the Drosha-DGCR8 Complex," *Cell*, vol. 125, no. 5, pp. 887–901, 2006.
- [228] S. R. Heras *et al.*, "The Microprocessor controls the activity of mammalian retrotransposons," *Nat. Struct. Mol. Biol.*, vol. 20, no. 10, pp. 1173–1181, 2013.
- [229] Y. Wang, R. Medvid, C. Melton, R. Jaenisch, and R. Blelloch, "DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal," vol. 39, no. 3, pp. 380–385, 2007.
- [230] D. Cirera-Salinas, J. Yu, M. Bodak, R. P. Ngondo, K. M. Herbert, and C. Ciaudo, "Noncanonical function of DGCR8 controls mESC exit from pluripotency," *J. Cell Biol.*, vol. 216, no. 2, pp. 355–366, 2017.
- [231] D. C. Leung, K. B. Dong, I. A. Maksakova, P. Goyal, R. Appanah, and S. Lee, "Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment , but not the maintenance , of proviral silencing," vol. 108, no. 14, 2011.
- [232] S. Schlesinger and S. P. Goff, "scientific report," *EMBO Rep.*, vol. 14, no. 1, pp. 73–79, 2012.
- [233] S. Schlesinger, A. H. Lee, G. Z. Wang, L. Green, and S. P. Goff, "Report Proviral Silencing in Embryonic Cells Is Regulated by Yin Yang 1," *CellReports*, vol. 4, no. 1, pp. 50–58, 2013.

- [234] D. Greiner, T. Bonaldi, R. Eskeland, E. Roemer, and A. Imhof, "Identification of a specific inhibitor of the histone methyltransferase SU(VAR)3-9.," *Nat. Chem. Biol.*, vol. 1, no. 3, pp. 143–145, 2005.
- [235] K. H. Kim and C. W. M. Roberts, "Targeting EZH2 in cancer," *Nat. Med.*, vol. 22, no. 2, pp. 128–134, 2016.
- [236] R. Rebollo *et al.*, "Retrotransposon-Induced Heterochromatin Spreading in the Mouse Revealed by Insertional Polymorphisms," vol. 7, no. 9, 2011.
- [237] G. Ecco *et al.*, "Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Article Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues," pp. 611–623, 2016.

Las referencias bibliográficas señaladas con un \* presenta algún tipo de error producto de una incompatibilidad entre el documento PDF y el software utilizado para su citación. Aquí debajo, se puede encontrar dichas referencias bibliográficas correctamente.

#### **CORRECCIONES BIBLIOGRAFÍA:**

**Ref. 4.** S.R. Richardson, A.J. Doucet, H.C. Kopera, J.B. Moldovan, J.L. García-Pérez and J.V. Moran. "The influence of LINE-1 and SINE retrotransposons on mammalian genomes". *Microbiol. Spectr.*, vol 3, no. 2, MDNA3-0061-2014, 2015.

**Ref. 17.** S.L. Martin and F.D. Bushman. "Nucleic acid chaperone activity of the ORF1 protein from the mouse Line-1 retrotransposon". *Mol. Cell. Biol.*, vol. 21, no. 2, pp. 467-475, 2011.

**Ref. 37.** R. Minakami, K. Kurose, K. Etoh, Y. Furuhashi, M. Hatton and Y. Sakaki." Identification of an internal cis-element essential for the human L1 transcription and a nuclear factor(s) binding to element". *Nucleic Acids Res.*, vol. 20, no. 12, pp. 3139-3145, 1992.

**Ref. 49.** S. E. Holmes, M. F. Singer and G. D. Swergold. "Studies on p40, the leucine zipper motif containing protein encoded by the first open reading frame of an active human LINE-1 transposable element". *J. Biol. Chem.*, vol. 267, no. 28, pp. 19765-19768, 1992.

**Ref. 55.** K. Usdin and A.V. Furano. "The Structure of the guanine-rich polypurine:polypyrimidine sequence at the right end of the rat L1 (LINE) element". *J. Biol. Chem.*, vol. 264, no. 26, pp. 15681-15687, 1989.

- Ref. 57.** W. Wei *et al.*, “Human L1 retrotransposition: cis preference versus trans complementation”. *Mol. Cell. Biol.*, vol. 21, no. 4, pp. 1429-1439, 2001.
- Ref. 70.** C. Philippe *et al.*, “Activation of individual L1 retrotransposition instances is restricted to cell-type dependent permissive loci”. *Elife*, vol. 5, e13926, 2016.
- Ref. 71.** M. Friedli *et al.*, “Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency”. *Genome Res.*, vol. 24, no. 8, pp. 1251-1259, 2014.
- Ref. 72.** S. Klawitter *et al.*, “Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells”, *Nat. Commun.*, vol. 7: 10286, 2016.
- Ref. 83.** N. Rodic *et al.*, “Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma”. *Nat. Med.*, vol. 21, no. 9, pp. 1060-1064, 2015.
- Ref. 84.** J. M. C. Tubio *et al.*, “Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes”. *Science*, vol. 345, no. 6196, pp. 1251-1259, 2014.
- Ref. 94.** M. Miné, *et al.*, “A large genomic deletion in the PDHX Gene caused by the retrotranspositional insertion of a full-length line-1 element”. *Hum. Mutat.*, vol. 28, no. 2, pp. 137-142, 2007.
- Ref. 118.** S. R. Heras, S. Macías and J. L. García-Pérez. “Control of mammalian retrotransposons by cellular RNA processing activities”. *Mobile Genetic Elements*, vol. 6, no. 4, pp. e28439, 2014.
- Ref. 127.** K. Zhao *et al.*, “Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutières Syndrome-Related SAMHD1”. *Cell Rep.*, vol.4, no. 6, pp. 1108-1115, 2013.
- Ref. 143.** K. Linblad-Toh *et al.*, “A high-resolution map of human evolutionary constraint using 29 mammals”. *Nature*, vol. 478, no. 7370, pp. 476-482, 2011.
- Ref. 154.** S. Schwartz, E. Meshorer and G. Ast. “Chromatin organization marks exon-intron structure”. *Nat. Struct. Mol. Biol.*, vol. 16, no.9, pp. 990-995, 2009.
- Ref. 170.** B. Lehnertz *et al.*, “Suv39h-mediated histone h3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin”. *Current Biology*, vol. 13, no. 14, pp. 1192-2000, 2003.
- Ref. 181.** G. Taylor, R. Eskeland and B. Hekimoglu-Balkan, M. M. Pradeepa and W. A. Bickmore. “H4k16 acetylation marks active genes and enhancers of embryonic stem cells,

but does not alter chromatin compaction". *Genome Res.*, vol. 23, no. 12, pp. 2053-2065, 2013.

**Ref. 186.** L. Vandel, E. Nicolas, O. Vaute, R. Ferreira, S. Ait-Si-Ali and D. Trouche. "Transcriptional repression by the retinoblastoma protein through the recruitment of a histone methyltransferase". *Mol. Cell. Biol.*, vol. 21, no. 19, pp. 6484-6494, 2001.

**Ref. 192.** A. Kirmizis *et al.*, "Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27". *Genes Dev.*, vol. 18, no. 13, pp. 1592-1605, 2004.

**Ref. 193.** A. Kuzmichev, K. Nishioka, H. Erdjument-Bromage, P. Tempst and D. Reinberg. "Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein". *Genes Dev.*, vol. 16, no. 22, pp. 2893-2905, 2002.

**Ref. 196.** A. P. Bracken, N. Dietrich, D. Pasini, K. H. Hansen and K. Hellin. "Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions". *Genes. Dev.*, vol. 20, no. 9, pp. 1123-1136, 2006.

**Ref. 197.** M. Walter, A. Teissandier, R. Pérez-Palacios, D. Bourc'his. "An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells". *Elife*, vol. 5, e11418, 2016.

**Ref. 202.** D.C. Schultz, K. Ayyanathan, D. Negorev, G.G. Maul and F.J. Rauscher III. "SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contribute to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins". *Genes Dev.*, vol. 16, no. 8, pp. 919-932, 2002.

**Ref. 204.** G. Wolf *et al.*, "The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses" *Genes Dev.*, vol. 29, no. 5, pp. 538-554, 2015.

**Ref. 208.** H. M. Rowe *et al.*, " TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells". *Genome Res.*, vol. 23, no. 3, pp. 452-461, 2013.

**Ref. 209.** N. Castro-Diaz *et al.*, "Evolutionally dynamic L1 regulation in embryonic stem cells". *Genes Dev.*, vol. 28, no. 13, pp. 1397-1409, 2014.

**Ref. 211.** J. Zeuthen *et al.*, "Characterization of a human ovarian teratocarcinoma-derived cell line". *Int. J. Cancer*, vol. 25, pp.19-32, 1980.

- Ref. 217.** J. Han, Y. Lee, K. H. Yeom, Y. K. Kim, H. Jin and V. N. Kim. "The Drosha-DGCR8 complex in primary microRNA processing". *Genes Dev.*, vol. 18, no. 24, pp. 3016-3027, 2004.
- Ref. 218.** D. M. Sassaman *et al.*, "Many human L1 elements are capable of retrotransposition". *Nat. Genet.*, vol. 16, no. 1, pp. 37-43, 1997.
- Ref. 219.** L. He, G.J. Hannon and C. S. Harbor. "MicroRNAs: small RNAs with big role in gene regulation". *Nat. Rev. Genet.*, vol. 5, no. 7, pp. 522-531, 2004.
- Ref. 223.** A.M. Denli, B. B. J. Tops and R. H. A. Plasterk. "Processing of primary microRNAs by the Microprocessor complex". *Nature.*, vol. 432, no.7014, pp. 231-235, 2004.
- Ref. 224.** R. I. Gregory, K. Yan and G. Amuthan. "The Microprocessor complex mediates the genesis of microRNAs", *Nature*, vol. 432, no. 7014, pp. 235-240, 2004.
- Ref. 227.** Y.Wang, R. Medvid, C. Melton, R. Jaenisch and R. Blelloch. "DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal". *Nat. Genet.*, vol.39, no. 3, pp. 380-385, 2007.
- Ref. 229.** D. C. Leung, K. B. Dong, I. A. Maksakova, P. Goyal, R. Appanah and S. Lee. "Lysine methyltransferase G9a is required for *de novo* DNA methylation and the establishment, but not the maintenance, of proviral silencing. *Proc. Natl. Acad. Sci. USA.*, vol. 108, no. 14, pp. 5718-5723, 2011.
- Ref. 230.** S. Schlesinger and S. P. Goff."Silencing of proviruses in embryonic cells: efficiency, stability and chromatin modifications". *EMBO Rep.*, vol. 14, no. 1, pp. 73-79, 2012.
- Ref. 234.** R. Rebollo *et al.*, "Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms". *PLoS Genetics*, vol. 7, no. 9, pp. e1002301, 2011.
- Ref. 235.** G. Ecco *et al.*, "Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues". *Dev. Cell.*, vol. 36, no. 6, pp. 611-623, 2016.



# **AGRADECIMIENTOS**



## **8. Agradecimientos**

En primer lugar, tengo que agradecer a mi director de tesis José Luís García Pérez que me diera la oportunidad de formar parte de su laboratorio para la realización de esta tesis doctoral. Además le agradezco la plena confianza que mostró en mí desde el primer momento dándome una gran autonomía para la realización de este proyecto así como el respeto y el interés que siempre mostró por mis opiniones y comentarios científicos. También le agradezco la flexibilidad laboral de la que siempre dispuse desde el nacimiento de mi hijo, Elián, que vive en Almería junto con mi futura mujer, Lupe. Además, quiero agradecerle la sensibilidad que mostró con la enfermedad de Lupe intentando ayudarme en todo momento.

Quiero agradecer también a mi otro director de tesis David Landeira todo lo que me ha enseñado en el campo de la epigenética así como también todo el tiempo que ha invertido en el proyecto desde que comenzó la parte del estudio epigenético momento en comenzó a colaborar con nosotros. Gracias por estar siempre disponible para discutir, pensar y charlar sobre el proyecto. Además, gracias por el tiempo empleado en explicarme como escribir la tesis doctoral y por las tutorías telefónicas que han contribuido definitivamente a la escritura de la tesis doctoral.

Además, quiero agradecer a todos los miembros del laboratorio 8, tanto aquellos que están como aquellos que estuvieron, toda la ayuda científica que en algún momento me habéis dado a lo largo de estos 6 años. Gracias, también, por conseguir que me haya sentido siempre respetado y valorado profesional y personalmente en el laboratorio. Gracias a todos por ser unos compañeros ejemplares siempre dispuestos a ayudar al otro sin esperar nada a cambio. Gracias por estar ahí en los momentos buenos, que los hay, pero también en los momentos de frustración y de hartazgo que son mayoritarios en esta profesión. Sobre todo, gracias por ser más amigos que compañeros, por perder un rato de vuestro trabajo para escuchar tanto mis problemas personales como los profesionales. Y por supuesto, gracias por vuestro apoyo, por vuestra preocupación y también por vuestra empatía durante la enfermedad de Lupe. El laboratorio 8 tiene muchas cosas buenas, pero sobre todo, yo me quedo con una, la calidad humana. No existe tanta bondad por metro cuadrado en otro lugar del mundo. Por toda vuestra ayuda científica y personal, podéis sentir esta tesis como vuestra. Gracias chicas y chicos.

También, quiero agradecer a mi padres, Juan Tomás y M<sup>a</sup> Luisa todo lo que han hecho por mí para que yo haya podido llegar hasta aquí. Gracias Papá y Mamá por todos los valores que me habéis transmitido con vuestro ejemplo durante todos estos años. Sin ellos, esta tesis doctoral no habría sido posible. Gracias por enseñarme a trabajar sin

descanso, a intentar mejorar cada día, a levantarme después de caerme y a seguir intentándolo hasta el final. Por supuesto, gracias por no haber escatimado nunca en esfuerzos para mi formación académica. Gracias por haber confiado en mí en un momento difícil como fue el cambio de carrera universitaria, por confiar en que yo podía estudiar y que solo necesitaba encontrar la carrera adecuada. Gracias a aquel gesto, hoy día he llegado hasta aquí. Aunque nunca os lo diga, os quiero y sois el mejor ejemplo que yo, como hijo he podido tener.

Por último, quiero agradecer a mi pareja y futura mujer, Lupe, su apoyo incondicional y su ayuda para que esta tesis haya podido salir adelante. Gracias, Lupe, por estar ahí siempre escuchando mis problemas, por haberme apoyado desde el primer instante y por haber hecho un esfuerzo monumental intentando comprender el mundo científico y sus particularidades. Gracias, cariño, por haber hecho un esfuerzo sobrehumano siendo madre y padre durante la semana desde el nacimiento de Elián. Gracias por sacar fuerzas de flaqueza durante tu enfermedad para estar ahí siempre dispuesta a interesarte por mis minúsculos problemas. Gracias, sobre todo, por la comprensión y la entereza que has mostrado durante el último tramo de la tesis doctoral donde la tesis se ha antepuesto a la familia teniendo que estar separados más tiempo de lo previsto. Gracias también a tus padres y tu hermana que nos han ayudado mucho con Elián permitiendo que yo pudiera dedicarme tranquilamente a la tesis sabiendo que mi hijo estaba en las mejores manos posibles. Por último, sólo quiero agradecer a Elián, mi hijo, su simple existencia. Es lo mejor que me ha pasado en la vida. Gracias por enseñarme que la familia es lo mejor que tienen las personas. Gracias por hacer que ser padre sea una tarea exigente y dura pero a la vez divertida y emocionante. Gracias por ser tan alegre, por transmitir felicidad desde el primer hasta el último minuto del día. Gracias por ayudarme a ponerle una fecha de caducidad a esta tesis por la necesidad que tenemos los tres de estar definitivamente juntos. Os quiero.

**9. Anexo V: Artículo**