



ugr

Universidad
de Granada

Programa oficial de Doctorado en Biomedicina

TESIS DOCTORAL

IMPACTO DE LA RETROTRANSPOSICIÓN DEL ELEMENTO LINE-1 EN
GENOMAS PLURIPOTENTES

Jairo Andrés Pulgarín Rocha

Director de la tesis:
Dr. José Luís García Pérez

GRANADA, 2015



CENTRO PFIZER-UNIVERSIDAD DE GRANADA-JUNTA DE ANDALUCÍA
DE GENÓMICA E INVESTIGACIÓN ONCOLÓGICA

Editor: Universidad de Granada. Tesis Doctorales
Autor: Jairo Andrés Pulgarín Rocha
ISBN: 978-84-9125-848-3
URI: <http://hdl.handle.net/10481/43639>

RESUMEN

Long interspersed element 1 (LINE-1) ó L1 es el retrotransposón autónomo más eficiente encontrado en mamíferos. Constituye el 17% del ADN humano y, dada su condición proliferativa y mutagénica puede causar daños o beneficios en el ADN de la célula hospedadora ya sea por su capacidad generadora de rearrreglos cromosómicos, e por efectos sutiles y específicos en los que puede afectar la expresión génica de una célula hospedera. Aunque el L1 hace parte constitutiva de la mayoría de las unidades de transcripción en los genomas mamíferos, los efectos de sus inserciones que principalmente son intrónicas, aún son objeto de investigación. En general se ha observado que la retrotransposición del L1 tiende a darse en células germinales o en fases tempranas del desarrollo, datos que indican que nuestro genoma pluripotente es un sustrato natural para el proceso de retrotransposición. Por tanto es relevante la determinación del impacto causado por las nuevas inserciones del L1 en los genomas pluripotentes y su papel en el proceso de diferenciación celular. Los eventos de retrotransposición del L1, en los que impacta unidades codificantes de genomas pluripotentes, traen como consecuencias cambios en la estructura molecular del gen que pueden interferir o no en su expresión en el estado pluripotente, aunque su efecto en el proceso de diferenciación celular no se ha observado. El conocimiento de estos eventos contribuiría a la etiología y epidemiología molecular, de las afecciones humanas causadas por inserciones del L1 llevadas a cabo en células madre en el desarrollo temprano de un individuo, además contribuiría aportando datos con respecto a las orientaciones funcionales desde el punto de vista evolutivo. El presente estudio busca determinar el impacto a nivel molecular causado por los eventos de retrotransposición del elemento L1 en secuencias codificantes de genomas pluripotentes y su repercusión en el estado diferenciado. Para esto se generaron a partir de un novedoso sistema de rescate en plásmido (*Recovery*) líneas celulares transgénicas pluripotentes a partir de células de carcinoma embrionario humano (hECs) PA1 con inserciones únicas del L1 dentro de unidades génicas identificadas. Mediante el uso del vector pRAM/LRE3 usado en la generación de las inserciones fue posible escindir todo el ADN "*artificial*" convencionalmente presente en los ensayos de retrotransposición *in vitro*, logrando de esta forma obtener inserciones del L1 en un estado "*naturalizado*", imitando de esta forma y de manera controlada un ambiente natural "*silvestre*" en el que se encuentra normalmente una nueva inserción del L1. De esta forma las líneas obtenidas con una inserción *de novo* del L1 son aptas para múltiples análisis moleculares, tales como evaluaciones de expresión génica entre otros. Aproximadamente el 40% de las inserciones obtenidas fueron identificadas en regiones intrónicas de unidades codificantes. Las variaciones en la expresión de los genes impactados por el L1 fueron diversos proponiendo panoramas individuales e intervenciones del ADN huésped que quedan por aclarar. En ciertas ocasiones se pudo determinar que la presencia de un L1 manipulado genéticamente influía de una manera diferencial en los resultados de la expresión génica de las células pluripotentes, validando de esta forma el sesgo

experimental al que están sometidos los ensayos de retrotransposición *in vitro*. interesantemente se identificaron varios casos atípicos de sobre-regulación de la expresión génica tanto en inserciones sentido como antisentido, desmitificando los efectos negativos sobre la expresión de los transcritos por efectos de las inserciones del L1 hasta ahora observados. Tres de los eventos de retrotransposición fueron identificados como agentes moduladores de la expresión génica en células una vez diferenciadas, apreciándose efectos desreguladores y sobrerreguladores igualmente observados en los estados pluripotentes, efectos que no se encontraron relación con el sentido de la orientación de las inserciones. De esta forma queda en evidencia la influencia del elemento L1 en la regulación de la expresión génica en las células pluripotentes (EC), convirtiéndose así mismo en un agente moderador de los niveles de expresión en células diferenciadas, y de esta forma contribuir en el desarrollo de variabilidad somática en el que en términos evolutivos podría desencadenar en ciertos casos procesos adaptación.

Tabla de contenido

RESUMEN	2
PARTE A:	7
Rol de la Retrotransposición del Elemento LINE-1 en la Modulación de la Expresión Génica Durante el Proceso de la Diferenciación Celular	7
CAPITULO 1: INTRODUCCIÓN	7
EL GENOMA HUMANO.....	8
REPETICIONES DE SECUENCIA SIMPLE (SSRs).....	10
ADN REPETIDO DISPERSO.....	11
DIVERSIDAD Y CLASIFICACIÓN DE LOS EMS.....	11
Clase II (Transposones).....	11
Mecanismo de movilización.....	13
Clase I (Retrotransposones).....	14
Retrotransposones LTR.....	14
RETROTRANSPOSONES NO-LTR.....	16
Elementos LINE.....	17
Estructura Molecular del LINE-1.....	17
Mecanismo de Movilización del LINE-1.....	18
Elementos SINE.....	22
Elementos Alu.....	22
Elementos SVA.....	23
Pseudogenes Procesados.....	24
RETROTRANSPOSICIÓN EN EL GENOMA HUMANO.....	25
CONSECUENCIAS DE LA RETROTRANSPOSICIÓN DEL L1 EN EL GENOMA HUMANO.....	27
Aspecto generales.....	27
Alteración de la Expresión Génica por la Inserción del L1.....	29
REGULACIÓN DE LOS ELEMENTOS MÓVILES.....	33
Regulación Epigenéticas.....	34
Control por RNAs.....	36
Control Post-transcripcional.....	39
OBJETIVOS	42
GENERAL	42
ESPECIFICOS:	42
Capitulo 2: MATERIALES Y METODOS	43
Líneas celulares y condiciones de cultivo.....	43
HeLa.....	43
Células de carcinoma embrionario humano (hEC) PA-1.....	43
Diferenciación celular.....	43
Subcultivos Celulares.....	44
ANTIBIÓTICOS	44
Ampicilina.....	44
Kanamicina.....	44
G418.....	44
PLASMIDOS	45
pRAM/LRE3.....	45
pCEP4-eGFP.....	45
CMV-Flp.....	45
MEDIOS DE CULTIVO BACTERIANO	46
Luria-Bertani (LB), caldo lisogénico.....	46

Luria-Bertani (LB), Agar	46
Medio SOB (SuperOptimalBroth) con Glicina	46
Bacterias competentes y Transformación Bacteriana.....	47
DH5 α	47
XL10-GOLD.....	47
TRANSFORMACIÓN	47
AISLAMIENTO Y PURIFICACIÓN DE ADN	48
ADN plasmídico	48
Aislamiento de ADN genómico	48
Aislamiento ARN total	49
Cuantificación de los Ácidos Nucleicos	49
TRANSFECCIÓN DE VECTORES.....	50
ENSAYO DE RETROTRANSPOSICIÓN	50
Fijación y Tinción Celular.....	51
Aislamiento clonal (Colony Lifting).....	51
Eficiencia de Transfección	51
Expansión y congelación-descongelación celular.	52
SISTEMA DE RESCATE DE INSERCIÓN (RECOVERY).....	52
Análisis de las Secuencias de Inserción	54
Recombinación FLP	54
Dilución Clonal.....	55
OLIGONUCLEÓTIDOS.....	56
Oligonucleótidos ó Primers para Identificación del casete de Neo.....	56
Primers para Recovery.....	56
Primers para Identificación de la recombinación <i>Flp</i>	59
Primers para RT-qPCR	59
REACCIÓN EN CADENA DE LA POLIMERASA (PCR).....	63
PCR NEO	63
PCR FLP.....	63
PCR Para la identificación del extremo 5' del L1.....	64
Análisis de la expresión génica RT-qPCR	64
Estadística.....	65
HERRAMIENTAS INFORMÁTICAS.....	66
Serial Cloner.....	66
BLAT	66
GeneCard	66
The Human Protein Atlas	66
Primer-BLAST	66
Universal Probe Library Assay Design Center	66
OligoCalc y Oligoevaluator.....	67
Capítulo 3: RESULTADOS	68
EVALUACIÓN DEL VECTOR PRAM/LRE3	68
Ensayo de RTN en células HeLa	68
Capacidad generadora de colonias transgénicas	68
GENERACIÓN DE LÍNEAS CLONALES EN PA1	69
Identificación de las inserciones PCR-NEO	70
Mapeo de las inserciones	70
Distribución de las inserciones.....	72
Sitios de corte de la endonucleasa (EN)	73
TSD y Cola de poli A.....	74
Mapeo intragénico de las inserciones.....	74
pRAMPA1_6-CPNE4.....	75
pRAMPA1_12-DDAH.....	76
pRAMPA1NR_2-TRPM3	77
pRAMPA1NR_9- ARHGAP42.....	77
pRAMPA1NR_10- FGGY.....	78
pRAMPA1NR_13- CASC15.....	79

p RAMPA1NR_25-PMM2.....	80
pRAMPA1D_4- OSBPL6	81
pRAMPA1D_6-FAM65B	82
pRAMPA1D_8- LPHN2	82
pRAMPA1D_15- RAD51B	83
pRAMPA1D41_ MAPK10	84
pRAMPA1D44_ LOC100287944.....	85
pRAMPA1D45_ FRAS1	86
RECOMBINACIÓN FLP: ANÁLISIS DE LA ESCISIÓN DEL CASETE DE NEO.....	86
Determinación Molecular por PCR-NEO.....	87
Subclones celulares sensibles	89
PCR-FLP	89
ANÁLISIS DE LA EXPRESIÓN GÉNICA	90
Variación de la Expresión Génica en los Subclones Celulares Flipped en Estado	
Pluripotente	90
Análisis de la expresión del gen CPNE4	91
Análisis de la expresión del gen TRPM3	92
Análisis de la expresión del gen ARHGAP42	93
Análisis de la expresión del gen FGGY	94
Análisis de la expresión del gen PMM2	95
Análisis de la expresión del gen OSBPL6.....	96
Análisis de la expresión del gen FAM65B	97
Análisis de la expresión del gen LPHN2.....	98
Análisis de la expresión del gen RAD51B.....	99
Análisis de la expresión del gen DDAH.....	100
Análisis de la expresión del gen MAPK10	101
Análisis de la expresión del gen FRAS1.....	102
Variación de la Expresión Génica Posterior al Proceso de Diferenciación Celular	103
Análisis de la expresión del gen CPNE4	103
Análisis de la expresión del gen TRPM3	104
Análisis de la expresión del gen ARHGAP42	105
Análisis de la expresión del gen FGGY	106
Análisis de la expresión del gen OSBPL6.....	107
Análisis de la expresión del gen FAM65B	108
Análisis de la expresión del gen LPHN2.....	109
Capítulo 4: DISCUSIÓN	111
ACTIVIDAD DEL VECTOR PRAM EN CÉLULAS TUMORALES	111
USO DEL PRAM COMO SISTEMA GENERADOS DE LÍNEAS CLONALES	111
ACTIVIDAD DEL VECTOR PRAM EN CÉLULAS PA-1	111
SITIO DE CORTE NO CARACTERÍSTICO DE LA EN DEL L1.....	112
LAS INSERCIONES PRODUCIDAS POR EL PRAM TIENEN UNA ALTA INCIDENCIA EN	
REGIONES CODIFICANTES	113
L1 “NATURALIZADOS”	113
EFECTO DE LAS INSERCIONES GÉNICAS EN LA EXPRESIÓN DEL GEN.	114
EFECTO DEL CASETE REPORTERO EN LA EXPRESIÓN DE LOS GENES.....	116
PAPEL DEL ELEMENTO L1 EN LA DIFERENCIACIÓN CELULAR	117
CAPITULO 5: CONCLUSIONES.....	119
PARTE B: ARTICULO ANEXO	121
Reprogramming triggers endogenous L1, <i>Alu</i> and SVA retrotransposition in	
human induced pluripotent stem cells	121
REFERENCIAS BIBLIOGRÁFICAS	151

PARTE A:

Rol de la Retrotransposición del Elemento LINE-1 en la Modulación de la Expresión Génica Durante el Proceso de la Diferenciación Celular

CAPITULO 1: INTRODUCCIÓN

EL GENOMA HUMANO

A comienzos del siglo XXI, los biólogos moleculares habían completado la secuenciación de los genomas completos de cientos de virus, decenas de bacterias, así como el de la levadura *S. cerevisiae*, como un organismo eucariota unicelular. Además, la gran cantidad de la secuencia del genoma es ahora conocida gracias a varios eucariotas multicelulares, dentro de los cuales se encuentran organismos como *C. elegans*, la mosca de la fruta *D. melanogaster*, y los seres humanos¹. El análisis detallado de estos datos de secuenciación ha revelado que una gran parte de los genomas de eucariotas superiores no codifica ARNm o cualquier otro ARN requerido por el organismo. Sorprendentemente, dicho ADN no codificante constituye más del 95 por ciento de ADN cromosómico humano².

En eucariotas superiores, las regiones de ADN codificantes, es decir los genes, se encuentran en medio de esta extensión de ADN aparentemente no funcional. Además del ADN no funcional existente entre genes, secuencias no codificantes denominadas **intrones** son comunes dentro de los genes de plantas y animales multicelulares. Los intrones son menos comunes pero a en ocasiones presentes en eucariotas unicelulares y raras veces en bacterias. La secuenciación del mismo gen codificante de una proteína en una variedad de especies eucariotas ha demostrado que la presión evolutiva selecciona el mantenimiento de las secuencias relativamente similares en las regiones codificantes, o **exones**. Tanto los intrones como los exones constituyen los genes que también podríamos denominar como **unidades génicas**, es decir el conjunto de secuencias de ADN que son necesarias para la síntesis de un producto génico funcional, ya sea una proteína o una molécula de ARN. Aunque los intrones son eliminados durante la transcripción y, en general no parecen desempeñar un papel crítico en la función de los genes, sus posiciones están muy conservadas³.

El genoma humano presenta una densidad de genes muy inferior a la que inicialmente se había predicho, representando aproximadamente menos del 1.5% de su longitud que constituyen los exones codificantes de proteínas. Un 70% está compuesto por ADN extragénico y un 30% por secuencias relacionadas con genes. Del total de ADN extragénico, aproximadamente un 70% corresponde a repeticiones dispersas, de manera que, más o menos, la mitad del genoma humano corresponde a secuencias repetitivas de ADN. Por su parte, del total de ADN relacionado con genes se estima que el 95% corresponde a ADN no codificante, que corresponde básicamente a los pseudogenes, fragmentos de genes, intrones o secuencias sin traducir UTR, entre otros².

En el caso de los genes eucariotas, existen distintas regiones que controlan su transcripción y procesamiento. Por ejemplo, regiones que se encuentran en zonas alejadas de los genes que controlan su transcripción a 50kb o más de distancia y son conocidas como *enhancers*. Otras regiones no codificantes presentes en los genes eucariotas son las secuencias que especifican un corte en el extremo 3' produciendo poliadenilaciones conocidos como sitios poli A, además de la presencia de sitios de *splicing* en el procesamiento del transcrito

primario de ARN. Dicho esto, mutaciones en estas señales del procesamiento del ARN pueden impedir la expresión funcional de un ARN mensajero (ARNm) y por lo tanto la de la proteína que codificada¹.

Además de los genes codificantes de proteínas, el genoma humano contiene varios miles de genes de ARN, cuya transcripción reproduce ARNs de transferencia (ARNt), ARN ribosómico (ARNr), microARN (miARN), u otros genes ARN no codificantes. Los ARNs ribosómicos y de transferencia son esenciales en la constitución de los ribosomas y en la traducción de las proteínas. Por su parte, los microARN tienen gran importancia en la regulación de la expresión génica, estimándose que hasta un 20-30 % de los genes del genoma humano puede estar regulado por el mecanismo de interferencia por miARN. Hasta el momento se han identificado más de 300 genes de miARN y se estima que pueden existir unos 500².

La densidad de los genes varía ampliamente dependiendo de las regiones del cromosoma, desde regiones altamente ricas en genes (eucromatina), hasta regiones pobres en secuencias codificantes (heterocromatina). En el genoma humano, menos del 2% corresponde a secuencias codificantes de proteínas (exones)⁴. La mayoría de los exones humanos que constituyen estas secuencias poseen una longitud entre 50 a 200pb, aunque muchos genes en sus extremos 3' son muchos más extensos².

En el caso de los intrones humanos, éstos varían de longitud considerablemente presentando una longitud media aproximada de 3.3kb. La totalidad de los intrones parecen representar alrededor del 26% de la secuencia del genoma humano². Aproximadamente un tercio del ADN genómico humano se puede transcribir en pre ARNm, en donde un 95% de estas secuencias las constituyen los intrones que serán eliminados posteriormente mediante el mecanismo de *splicing*².

Las regiones intergénicas ó extragénicas comprenden la mayor parte de la secuencia del genoma humano, y su función es generalmente desconocida. Buena parte de estas regiones está compuesta por elementos repetitivos, clasificados como repeticiones en tándem o repeticiones dispersas, aunque el resto de la secuencia no responde a un patrón definido y clasificable. Gran parte del ADN intergénico puede ser un artefacto evolutivo sin una función determinada en el genoma actual, por lo que tradicionalmente estas regiones han sido denominadas ADN "basura" (Junk DNA), denominación que incluye también las secuencias intrónicas y pseudogenes. No obstante, esta denominación no es la más acertada dado el papel regulador conocido de muchas de estas secuencias, representando actualmente un verdadero tesoro de información acerca de los procesos biológicos². Además el notable grado de conservación evolutiva de algunas de estas secuencias parece indicar que poseen otras funciones esenciales aún desconocidas o poco conocidas. Por lo tanto, algunos prefieren denominarlo "ADN no codificante" (a pesar de que el llamado "ADN basura" incluye también transposones codificantes) o "ADN repetitivo". Algunas de estas regiones constituyen en realidad genes precursores para la síntesis de microARN que como se mencionó

anteriormente son reguladores de la expresión génica y del silenciamiento génico².

Como agentes activos dentro del genoma, las secuencias repetidas han reconfigurado el genoma causando reordenamientos ectópicos, creando genes totalmente nuevos, modificación y reorganización de los ya existentes, y la modulación del contenido general de regiones ricas en GC. También arrojan información sobre la estructura cromosómica y su dinámica, además proporcionan herramientas en la medicina genética así como para los estudios de genética de poblaciones².

El ADN no codificante en los organismos multicelulares contiene muchas regiones que son similares pero no idénticas. Las variaciones dentro de algunos tramos de este ADN repetitivo son tan grandes que cada persona solo puede ser distinguida por una huella genética ó "*fingerprint*" de ADN con base a estas variaciones de la secuencia, a este tipo de secuencias se les ha denominado **repeticiones de secuencia simple (SSRs)**. Además, algunas secuencias de ADN repetitivas no se encuentran en posiciones constantes en el ADN de individuos de la misma especie. Tales elementos de ADN "móviles", que están presentes en ambos organismos procariontes y eucariotes, pueden causar mutaciones cuando se trasladan a nuevos sitios en el genoma. A este tipo de secuencias se les ha denominado **ADN repetido disperso** y a pesar de que en general no tienen ninguna función en el ciclo de vida de un organismo individual, los elementos móviles probablemente han jugado un papel importante en la evolución¹.

REPETICIONES DE SECUENCIA SIMPLE (SSRs).

Dentro del ADN repetitivo, este tipo de secuencias son las menos prevalentes en el genoma humano, representando aproximadamente el 3% de este, y están conformadas por repeticiones perfectas o casi perfectas de secuencias relativamente cortas². El conjunto de repeticiones en tándem más común son los denominados **satélites de ADN**, el cual comprenden un total de 250 Mb del genoma humano. Son secuencias de entre 5 y varios cientos de nucleótidos que se repiten en tándem miles de veces generando regiones repetidas con tamaños que oscilan entre 100kb hasta varias megabases. Las SSRs con una unidad de repetición mediana que contienen desde 14 hasta 500pb, son frecuentemente llamadas **minisatélites** mientras que aquellas que poseen unidades de repetición mas cortas aproximadamente de 1 a 13pb son frecuentemente llamadas como **microsatélites**².

Dentro de una especie, las secuencias de nucleótidos de las unidades repetidas que constituyen series en tándem de secuencias simples de ADN son altamente conservadas entre individuos. En contraste, diferencias entre el número de repeticiones, y por tanto en la longitud de las series en tándem de las secuencias simples de la misma unidad repetida, son muy comunes entre individuos. Estas diferencias en longitud se piensan son el resultado de un entrecruzamiento desigual entre regiones de ADN de secuencia simple durante el proceso de meiosis. Como consecuencia de este entrecruzamiento desigual,

la longitud de algunas series en tándem son únicas en cada uno de los individuos¹.

ADN REPETIDO DISPERSO

Son secuencias de ADN que se repiten de modo disperso por todo el genoma, constituyendo aproximadamente el 45% del genoma humano. Estas secuencias tienen la potencialidad de propagarse por el genoma directamente a partir de su secuencia de ADN o bien al transcribirse en una molécula de ARNm intermediario, para luego retrotranscribirse e insertarse en otro punto del genoma. A este fenómeno molecular se le ha denominado **transposición**.

El ADN repetido disperso está constituido por las secuencias de ADN móvil que a su vez está compuesto de un gran número de copias comprendidas en un reducido número de familias. Estas secuencias se encuentran dispersas por todo el genoma mamífero constituyendo alrededor de un 25 a 50% del mismo (aproximadamente 45% del genoma humano)². Debido a su habilidad para moverse por todo el genoma, por medio de la transposición, a este tipo de secuencias se les denomina **elementos móviles del ADN** ó simplemente *elementos móviles (EMs)*.

DIVERSIDAD Y CLASIFICACIÓN DE LOS EMS.

Como se mencionó anteriormente los EMs son secuencias discretas de DNA que pueden cambiar de posición en el genoma de un organismo. Fueron descubiertos en maíz en 1951 por Bárbara McClintock⁵. En función del mecanismo de transposición se clasifican en: clase I (aumentan en el número de copias) y clase II (no aumenta el número de copias).

Clase II (Transposones)

También denominados **transposones de ADN**, generalmente se movilizan por medio de un mecanismo de "*cortar y pegar*" en el que el transposón se escinde de un lugar y reintegrado en otros lugares. La mayoría de los transposones de ADN se mueven a través de un mecanismo no replicativo⁶.

Este tipo de elementos móviles consisten en un gen que codifica para una transposasa y está flanqueado por dos repeticiones terminales invertidas (TIRs) (figura 1-3). La transposasa reconoce estas TIRs para llevar a cabo la escisión del cuerpo del transposón de ADN, que será insertado en una nueva localización genómica⁷. Luego de la movilización, el sitio de inserción del ADN es duplicado, formando las *duplicaciones del sitio diana* (TDS), que representan una marca única para cada transposón de ADN⁷.

En general los transposones de ADN son clasificados en las subclases I y II, formadas a su vez por diferentes familias ubicadas de acuerdo a las repeticiones invertidas terminales TIR ó las duplicaciones de ADN en el sitio diana de la inserción TDS características de la transposición. De acuerdo a las

semejanzas existentes entre transposones, la mayoría de los transposones de ADN eucariotas están agrupados en dos subclases pertenecientes a la subclase I⁶: la case *Ac/hobo* caracterizada por duplicaciones del sitio de inserción de 8pb; y la clase *Tcl/mariner* formada por un dímero de duplicación TA⁸. Dentro de la subclase II encontramos los elementos *Maverick* y los *Helitron* debido a que son replicados y no realizan roturas en el ADN de cadena doble durante su inserción⁶ (Figura 1-1).

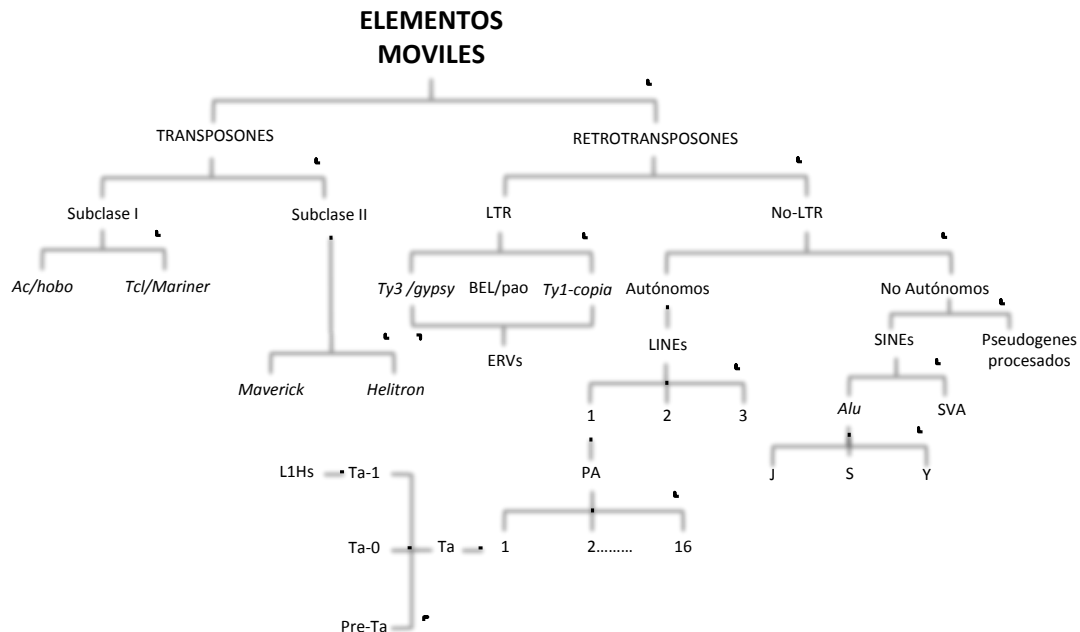


Figura 1-1. Diagrama general de la clasificación de los elementos móviles. (Adaptado de Smit, AF. 1995 y 1996; Muñoz-López & García-Pérez 2010).

Los transposones de ADN comprenden alrededor del 3% del genoma humano, y los miembros más relevantes son los elementos *mariner*⁴ (figura 1-2). A pesar de que no se ha confirmado que los elementos *mariner* sean activos en humanos, un elemento activo de este tipo ha sido reconstruido a partir de secuencias consenso de salmón y ha sido denominado “Sleeping Beauty”. Este elemento se ha visto que puede movilizar por el mecanismo de cortar y pegar en ensayos in vitro en células HeLa, planteando la posibilidad de encontrar este tipo de elementos activos en el genoma humano⁹.

Inicialmente se asumió que los transposones de ADN no eran activos en los genomas de mamíferos, pero estudios posteriores identificaron nuevos transposones de ADN potencialmente activos en genomas de ciertas especies de murciélagos, indicando que en dichas especies estos elementos podría estar contribuyendo a la diversidad genética de estos organismos^{10; 11}. En el caso de humanos los transposones de ADN se encuentran mutados e inactivos y por tanto se consideran fósiles moleculares^{2; 12}.

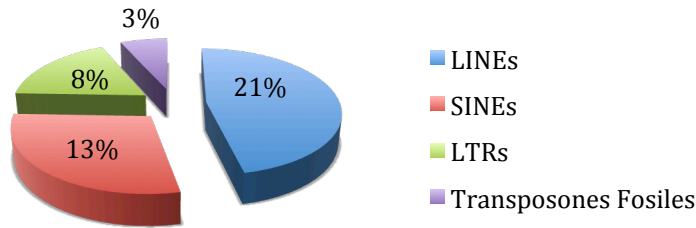


Figura 1-2. Porcentaje de abundancia relativa de los elementos móviles (EM) en humanos (adaptado de Lander et al; 2001).

Mecanismo de movilización

El mecanismo de transposición como se mencionó anteriormente, en un proceso que consiste en cortar y pegar la secuencia del transposón en un nuevo sitio genómico, por lo que en términos generales el ADN no es replicado.

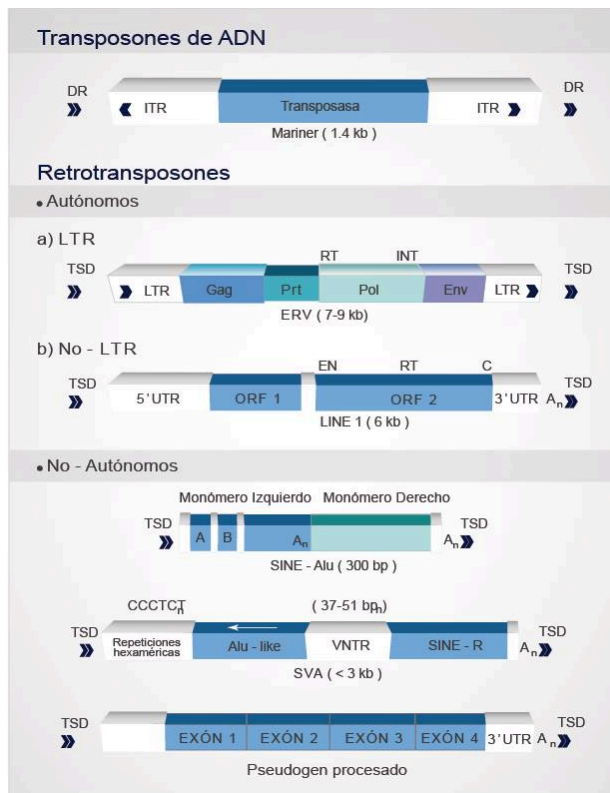


Figura 1-3. Elementos móviles en mamíferos. Esquema de los elementos móviles identificados en el genoma humano. DR: repeticiones directas; ITR: repeticiones terminales invertidas; Gag: antígeno grupo-específico; Prt: proteasa; Pol: polimerasa; Env: envoltura; RT: reverso transcriptasa; INT: dominio integrasa; TSD: duplicación en el sitio diana; LTR: repetición terminales largas; EN: dominio endonucleasa; C: dominio rico en cisteína; A_n: cola de poli Adeninas; A y B: Regiones promotoras internas de la ARN polimerasa III; SVA (SINE-R, VNTR, Alu); VNTR: numero variable de repeticiones en tándem. (Adaptado de Goodier y Kazazian, 2008)

Dos moléculas de la enzima transposasa reconocen las TIRs y se unen a ellas por medio del motivo HTH formando el complejo denominado SEC (*Single-End*

complex). Ambas transposasas cortan el enlace fosfodiéster mediante hidrólisis de la cadena de ADN en el extremo 5' de las TIRs, generando de esta manera extremos cohesivos. Posteriormente las transposasas interactúan entre sí en la formación del denominado PEC (*Paired-End Complex*) que consiste un dímero de transposasa. En este punto los extremos 3' son hidrolizados y de esta manera obtener el fragmento de ADN que será trasladado. Finalmente el complejo PEC se une al sitio diana del ADN, formando el complejo de *Captura de la Secuencia (Diana target capture complex)* en donde se lleva a cabo la inserción. El extremo 5' de la secuencia diana sufre un ataque nucleofílico por parte del extremo 3' de las secuencias movilizadas. El espacio generado en el extremo 5' del transposón es completado por el hospedador, y de esta manera generando TSD canónicas flanqueando la nueva inserción del transposón⁶

Clase I (Retrotransposones)

Los miembros de este grupo también son llamados *transposones de ARN* ó *retrotransposones* ya que se movilizan a través de la transcripción inversa de un intermedio de ARN, mecanismo denominado *retrotransposición (RTN)* el cual representa un proceso de replicación del ADN⁶. Pueden ser subdivididos en dos grupos principales dependiendo de la presencia o ausencia de *Repeticiones Terminales Largas* o **LTRs** flanqueando la secuencia central de los retroelementos⁶ (figura 1-1).

A diferencia del mecanismo de transposición descrito en los transposones de la clase II, los elementos móviles de la clase I en general realizan un proceso de *copiar y pegar* la secuencia del transposón en un nuevo sitio genómico, por lo que en términos generales el ADN es replicado.

Retrotransposones LTR

Los retrotransposones LTR son más comunes en algunos organismos, como es el caso de las levaduras (elemento Ty)¹³ y *Drosophila* (elemento copia); aunque son menos abundantes en mamíferos comparados con los retrotransposones no LTR¹. En general los retrotransposones LTR son a su vez subdivididos en 3 grupos basados tanto en el grado de similitud entre las secuencias así como el orden de codificación de los genes, estos son los Ty1-*copia*, Ty3-*gypsy* y BEL-Pao (figura 1-1). Los grupos Ty1-*copia* y Ty3-*gypsy* se encuentran comúnmente en alto número de copias (hasta unos pocos millones de copias por núcleo haploide) en genomas animales, hongos, protistas, y plantas. Los elementos BEL-Pao han hasta ahora solo se han encontrado en animales^{14; 15}. Aunque los *Retrovirus* se clasifican a menudo por separado, comparten muchas características con retrotransposones LTR. Una de las diferencias más importantes con los retrotransposones Ty1-*copia* y Ty3-*gypsy* es que los retrovirus tienen una proteína de la envoltura (*env*). Un retrovirus se puede transformar en un retrotransposón LTR través de la inactivación o eliminación de los dominios que permiten la movilidad extracelular. Si tal retrovirus infecta y, posteriormente se inserta en el genoma de una células de

una línea germinal, puede ser transmitido verticalmente y es convertido en un retrovirus endógeno (ERV)¹⁵.

Los elementos LTR constituyen aproximadamente el 8% (figura 1-2) del genoma humano, y debido a sus similitudes con los retrovirus, algunas veces son llamados como retrotransposones virales o más comúnmente ERVs (del inglés endogenous retroviruses)^{2; 16}. Debido a mutaciones recientes en sus secuencias, los ERVs han mostrado ser inactivos en el genoma humano por lo que no pueden realizar una retrotransposición autónoma². Los retrovirus infecciosos de vertebrados son monofiléticos y probablemente evolucionaron a partir de un retrotransposón LTR por transmisión vertical, quizás se asemejan a un miembro de la familia relacionada más cercana Ty3-*gypsy*, o a la más lejana como la Ty1-*copi* de los elementos LTR⁸(figura 1-1).

En la estructura de los LTRs se observa que el cuerpo del elemento se encuentra flanqueado igualmente por secuencias TSDs semejante a los transposones de ADN. Como su nombre lo indica, estos retrotransposones se caracterizan por la presencia de secuencias repetidas terminales largas LTRs que oscilan entre 100pb y 5kb de longitud, flanqueando una región central codificante, constituidas por aproximadamente de 250 a 600pb. Estas secuencias igualmente son características del ADN retroviral integrado y resultan fundamentales para el ciclo de vida de los retrovirus debido a que contienen todos los elementos reguladores transcripcionales necesarios.

Además de compartir las LTRs con los retrovirus, los retrotransposones codifican todas las proteínas de los tipos de retrovirus más comunes, a excepción de la cápside proteica. Perdiendo esta cápside, los retrotransposones LTR no pueden propagarse por una célula huésped ni infectar otras células; sin embargo se pueden transponer a nuevos sitios del ADN hospedador^{17; 18}. Contienen además los genes *Gag* y *Pol*, lo cuales codifican una proteasa, reversotranscriptasa, RNasa H y una integrasa, proteínas necesarias para que se lleve a cabo correctamente su transposición²(figura 1-3). Los retrovirus exógenos parecen haber surgido de retrotransposones endógenos por la adquisición de un gen de envoltura celular (*env*)¹⁹. La transposición se produce a través del mecanismo retroviral con transcripción inversa y ocurre en una partícula citoplasmática semejante a los virus, preparado por un ARNt. Aunque existe una gran variedad de retrotransposones LTR, sólo los ERVs específicos de vertebrados parecen haber estado activos en el genoma de los mamíferos. Los retrovirus de mamíferos se dividen en tres clases (I - III), cada una esta formada por muchas familias con orígenes independientes. Aproximadamente el 85% de los "fósiles" derivados de retrotransposones LTRs consisten solamente de un LTR aislado, habiéndose perdido la secuencia interna por recombinación homóloga entre las secuencias LTRs flanqueantes².

Un factor fundamental en el ciclo de vida del retrovirus, es la formación del ARN genómico retroviral procedente del ADN retroviral integrado. Este proceso sirve como modelo para la generación del ARN intermediario durante la transposición de los retrotransposones LTRs. En este proceso, la LTR 5' retroviral funciona a modo de promotor de la ARN polimerasa II de la célula

huésped el cual guía dicha enzima para el inicio de la transcripción. Después de que se haya transcrito toda la secuencia del ADN retroviral, la secuencia de ARN que corresponde a la porción 3' del LTR, conduce a las enzimas procesadoras del ARN de la célula huésped a que corten el transcrito primario y sea unido una cola de poli A. El ARN genómico resultante, el cual ha perdido su LTR completamente es empacado en un virión que emerge de la célula huésped¹.

Una vez los retrovirus han infectado la célula huésped, se lleva a cabo el proceso de reverso transcripción de ARN genómico por la reverso transcriptasa codificada por ellos mismos, obteniendo al final un ADN de doble cadena que contiene la secuencia completa de los LTRs. La integrasa, otra enzima que es codificada por los retrovirus, y que está estrechamente relacionada con la transposasa de algunos transposones, usa un mecanismo similar para insertar la doble cadena de ADN retroviral en el genoma del ADN huésped. En este proceso, pequeñas repeticiones continuas en la secuencia diana son generadas en cada extremo del ADN vírico insertado¹

En el genoma humano y otros primates existen varias familias distintas de secuencias similares a retrovirus endógenos (HERV). Una de estas familias, el grupo HERV-K, contiene miembros que codifican proteínas funcionales y que han sido implicados en la etiología de la diabetes mellitus insulino dependiente (DMID). Debido al potencial funcional y la relevancia en enfermedades, es importante para determinar si existen diferencias genéticas asociadas a HERV-K entre individuos de la misma especie²⁰. Sólo una sola familia retrotransposones LTR (HERVK10) es conocida por haberse transpuesto desde nuestra divergencia con el chimpancé hace 7 Myrs, con sólo una copia conocida (en la región HLA) que no se comparte entre los humanos²⁰.

RETROTRANSPOSONES NO-LTR

Los elementos móviles más abundantes en mamíferos son los retrotransposones que han perdido las secuencias LTR, también son conocidos como retrotransposones no víricos²¹, y particularmente se encuentran ubicados en regiones del genoma ricas en AT²² Estas secuencias moderadamente repetidas de ADN presentes en los genomas de los mamíferos se dividen principalmente en dos clases: los elementos autónomos **LINES** (del inglés Long Interspersed Elements) que poseen la maquinaria necesaria para llevar a cabo su propia movilización, y los elementos no autónomos **SINEs** (del inglés Short Interspersed Elements)(figura 1-1) que usan los elementos producidos por los LINES para su movilización¹⁶. Un gran número de LINES y SINEs se han acumulado en los genomas de los eucariotas superiores durante la evolución, debido a la copia repetida de secuencias en algunas posiciones del genoma e inserciones de las copias en nuevas posiciones genómicas^{1; 23}. Aunque estos elementos no poseen secuencias LTRs, igualmente se ha encontrado que se movilizan a través de un ARN intermediario por el mecanismo de “copiar y pegar” que implica la replicación del ADN.

Elementos LINE

Los elementos LINEs constituyen alrededor del 21% del genoma humano, representados en aproximadamente 850000 copias (figura 1-2). Los LINEs humanos se agrupan en tres familias principales similares en sus mecanismos de transposición pero que se diferencian en sus secuencias: L1, L2 y L3 (figura 1-1). Sólo los miembros de la familia L1 (**LINE-1**) son actualmente activos. El genoma humano contiene aproximadamente 900000 copias de elementos LINE, representado la asombrosa cantidad del 21% del ADN total humano² (figura 1-2). El elemento LINE-1 o L1, es el principal constituyente de los LINE en el genoma humano representado en aproximadamente 500000 copias².

Un análisis comparativo entre secuencias genómicas individuales del L1 y una secuencia consenso derivada de un actual L1 activo, se realizó en la determinación de la edad de los L1s genómicos. Por medio de este análisis se pudieron identificar 16 subfamilias del L1 específicas de primates (denominadas PA1 hasta PA16)^{8; 24}. En el caso de humanos, se han identificado ciertos tipos de L1 que aún pueden movilizarse, denominados L1Hs que provienen de la subfamilia PA1^{25; 26}(figura 1-1).

Se ha podido estimar, con base a análisis computacionales y funcionales, que un genoma promedio humano contiene aproximadamente de 80 a 100 copias de LINE-1 que permanecen activas, es decir, son retrotransposones competentes para la RTN (**L1-RC**)^{26; 27}. De estos la mayor parte pertenece a un grupo denominados “hot L1s”, a los cuales se les atribuye la gran proporción de la actividad de RTN de un genoma determinado²⁷.

La mayoría de los elementos L1Hs activos pertenecen a una pequeña población de elementos denominados Ta (del inglés transcribe active)^{28; 29}. Notablemente, los L1s provenientes del subgrupo Ta son los responsables de la mayoría de la actividad de retrotransposición en los genomas humanos actuales (figura 1-1) (revisado en referencia 27)³⁰. El grupo Ta del LINE-1 a su vez puede ser subdividido aún más finamente, con base a diferencias sutiles entre secuencias. Las subdivisiones que componen este subgrupo son la Ta-1, Ta-0 y pre-Ta, de éstas la primera es la que contiene el mayor número de elementos LINE-1 activos seguida de la Ta-0 y la pre-Ta^{25; 26; 29}(figura 1-1).

Estructura Molecular del LINE-1

La estructura general de un LINE-1 humano competente ó L1-RC se muestra en la figura 1-4. Poseen un tamaño promedio de 6 kb de longitud^{31; 32} en el que inicialmente contienen una región que se había considerado que no se traducía por lo que se le había denominado extremo 5' (5'UTR), que entre otras cosas posee una isla CpG involucrada en funciones reguladoras³³, además de secuencias promotoras internas de la ARN polimerasa II en dirección sentido y antisentido (ASP)^{34; 35}. El promotor sentido es el encargado de guiar la transcripción del L1, mientras que ASP ha sido relacionado con la autorregulación del L1 mediante la generación de ARNs de interferencia (ARNi)^{36; 37}.

ELINE1 contiene tres marcos abiertos de lectura ORF0, ORF1 y ORF2 (del inglés Open Reading frames) que codifican para las proteínas necesarias para la retrotransposición^{38; 39}. ORF0 se encuentra en la región 5' UTR en dirección antisentido y codifica para un péptido de 71 aminoácidos que se ha visto que potencia la movilidad del L1 de primates⁴⁰. ORF1 es de aproximadamente 1 kb de longitud, y codifica para una proteína (p40) de unión a ARN, que se une específicamente al transcrito del L1 cerca del extremo 5' del ORF2p^{41; 42}. Además posee una actividad chaperona de ácidos nucleicos^{43; 44} (figura 1-4). Separado por 66pb de secuencia intergénica del ORF1 encontramos el ORF2, el cual tiene una longitud aproximada de 4kb, y codifica para una proteína de 150kDa (ORF2p)^{45; 46} con actividad endonucleasa (EN)³⁹ y reversotranscriptasa (RT)⁴⁷, con un dominio rico en cisteína en el C-terminal^{38; 48} (figura 1-4). Análisis comparativos han determinado que posee una larga región de homología con las reverso transcriptasas de los retrovirus y de los retrotransposones LTRs⁴⁷. La secuencia completa del L1 finaliza con una región sin traducir en el extremo 3' (UTR) de 205pb que parece no influir en la RTN del L1³⁸ el cual recientemente se le ha identificado una región promotora⁴⁹. Al final de la región 3' UTR se puede observar una señal de poliadenilación poco ortodoxa seguida de una cola de poli A^{31; 50} (figura 1-4).

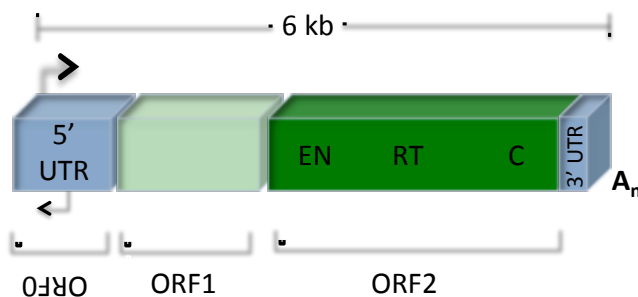


Figura 1-4. Elemento LINE-1. Representación gráfica de la estructura de un L1 de copia completa.

Mecanismo de Movilización del LINE-1

Para la movilización del L1 las proteínas ORF1p y ORF2p son traducidas, ya que representan la maquinaria necesaria para que la RTN se pueda llevar a cabo. Inicialmente la secuencia promotora sentido ubicada en el extremo 5' guía la transcripción dependiente de la polimerasa II^{34; 51-54}. Posteriormente el transcrito primario de ARN es poliadenilado (adición de una cola de poli A) por los mismos mecanismos post-transcripcionales de la célula huésped. Una vez maduro el transcrito, es transportado al citoplasma donde se lleva a cabo la traducción de las proteínas ORF1p y ORF2p (figura 1-5). La traducción iniciada a partir del extremo 5' UTR del L1 humano es dependiente de caperuza⁵⁵. Inicialmente múltiples copias de ORF1p se unen a su propio ARN, y la proteína ORF2p se une directamente a la cola de poli A, proceso denominado preferencia en *Cis*⁵⁰. El nivel de traducción de ORF2 es dependiente de la traducción de ORF1^{56; 57}.

Posteriormente el mismo ARNm es transportado de nuevo al núcleo para su integración, junto con las proteínas ORF1p y ORF2p en forma de un complejo ribonucleoproteico, el cual se ha considerado como un intermediario de la retrotransposición, y se le ha dado el nombre de ribonucleopartícula (RNP)^{41; 58-60} (figura 1-5).

El mecanismo por el cual el L1 se integra en el cromosoma se denomina *TPRT* (del inglés Target Prime Reverse Transcription). El proceso se inicia cuando la proteína ORF2p con su dominio endonucleasa (EN) se encarga de producir un corte en una de las cadenas de DNA cromosomal en sitios con una secuencia consenso degenerada 5'-A/TTTT-3'³⁹ (figura 1-5 y 1-6). Posteriormente el ARN del L1 es reversotranscrito por el dominio reversotranscriptasa (RT) de ORF2p, el cual usa el extremo hidroxilo 3' liberado en el corte, y usando este extremo libre rico en poli T como cebador uniéndolo a la cola de poli A del ARN del L1, empieza la síntesis de una nueva molécula de ADN complementario (ADNc) de cadena sencilla usando el ARN como molde (figura 1-6). El corte de la segunda hebra puede ocurrir aguas arriba, aguas abajo o en línea del corte inicial, y de esta forma crear secuencias duplicadas del sitio de origen TSDs, deleciones de sitio de origen o inserciones romas respectivamente. Se presume que el dominio EN de ORF2p también es el que realiza el corte en la segunda hebra de ADN, sugiriéndose que ORF2p con sus dominios RT y EN trabaja en forma de dímero, en donde cada subunidad cataliza la rotura de cada una de las hebras^{61; 62}. Sin embargo aún queda mucho por conocer con respecto al mecanismo de corte de la segunda hebra, aunque hay evidencias de que este corte está dirigido por pequeñas micro-homologías con la nueva hebra sintetizada en la hebra del inferior^{63; 64}.

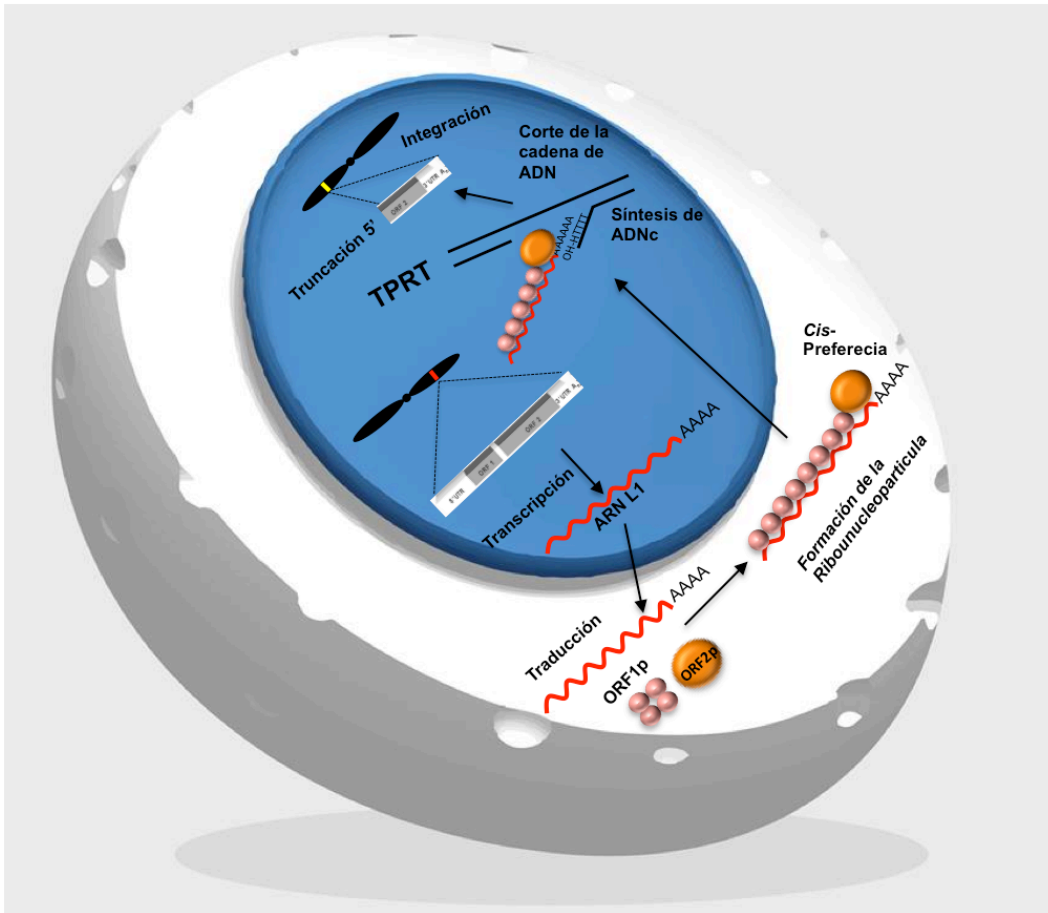


Figura 1-5. Ciclo de la retrotransposición. En el esquema el L1 se muestra en gris y los cromosomas en negro. El ARNm está representado en rojo y las proteínas ORF1p y ORF2p en rosa y naranja respectivamente. Las inserciones del L1 se indican por una línea roja y amarilla dentro de los cromosomas (Adaptado de Kazazian;1998).

Algunos tipos de no-LTRs han adquirido un dominio RNasa H, el cual posiblemente elimine el ARNm molde tan pronto como la hebra complementaria se haya sintetizado^{65; 66}. De otros no hay evidencia que posean éste dominio, y el ARNm posiblemente pueda ser desplazado por la síntesis de la segunda hebra⁶⁷, o simplemente hidrolizado por las enzimas celulares del hospedador¹.

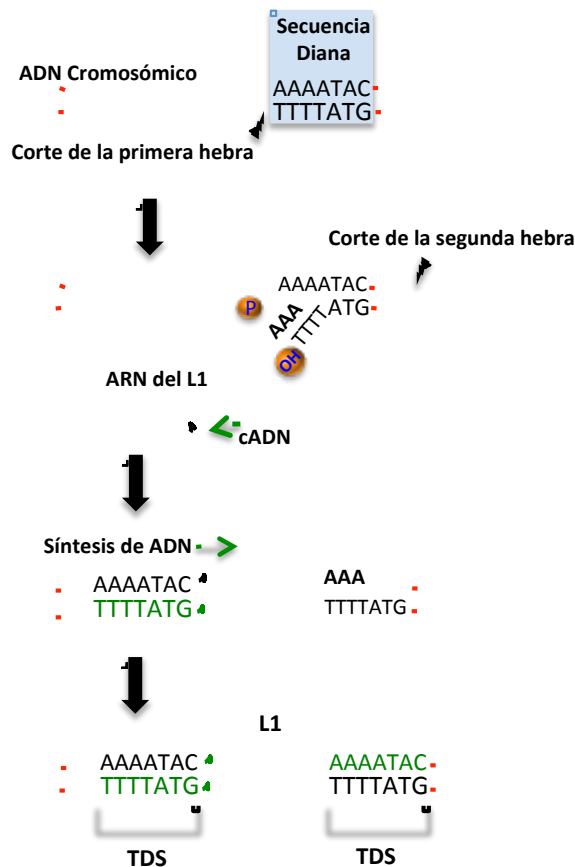


Figura 1-6. Mecanismo molecular del TPRT. En el esquema se representa el mecanismo de integración de la secuencia del L1 en un nuevo loci.

Finalmente los extremos 5' y 3' de la nueva cadena de ADN son ligados, y de esta manera se completa el ciclo de la inserción (1-6). Éste último paso posiblemente sea catalizado por las mismas enzimas celulares que intervienen en el proceso de replicación del ADN las cuales están encargadas de eliminar los cebadores o primers de ARN y de ligar los fragmentos de Okazaki (revisado en referencia 66)⁶⁸. El fin del proceso, trae como consecuencia la inserción de una copia de LINE en un nuevo sitio del ADN genómico (figura 1-5).

Un 95% de los LINE humanos están truncados en su extremo 5'⁶⁹, sugiriendo que la reverso-transcripción termina antes de completar la totalidad de la secuencia del L1 probablemente a la baja procesividad del dominio EN de ORF2p, por lo que es posible encontrar fragmentos de longitud variable desde la cola de poli A hasta la truncación⁵⁰. Debido a esto es que el tamaño promedio de un elemento LINE es de solo 900pb, teniendo en cuenta que la longitud completa es de aproximadamente 6 kb². Además, casi toda la copia completa del elemento L1 contiene codones de parada además de marcos de lecturas mutados tanto en ORF1 como en ORF2^{27; 48; 50}. Estas mutaciones posiblemente han sido acumuladas en las secuencias del L1 a lo largo de la evolución. Como resultado de la truncación y las mutaciones, solo aproximadamente el 0,01% de las secuencias del LINE humano poseen una copia completa con sus marcos abiertos de lectura intactos tanto para ORF1 y ORF2, de estos, solo de 80 a 100 copias se estima son elementos putativos activos para la retrotransposición²⁷.

Elementos SINE

Los elementos SINEs se encuentran principalmente en los genomas mamíferos aunque también se han encontrado en plantas, hongos e insectos¹⁶. El genoma humano contiene aproximadamente 1.6 millones de copias de elementos SINEs representando el 13% del ADN genómico humano (figura 1-2), con respecto a esto aproximadamente un 70% está constituido por elementos *Alu*² (figura 1-1). De esta forma los elementos *Alu* que se convierten en los mayores constituyentes de las secuencias SINE en el genoma humano². El otro miembro mayoritario de los SINEs son las secuencias SVA que representan el 1% de nuestro genoma³⁰ (figura 1-1).

Las proteínas codificadas por los elementos LINE-1 son usadas por los elementos no autónomos SINE para su movilización^{70; 71}, hecho al que se le denomina preferencia en *trans*¹⁶. Los elementos movilizados por las proteínas sintetizadas a partir del L1 poseen igualmente los mismos rasgos estructurales característicos producidos en la RTN de los elementos autónomos; tales como la presencia de una cola de poli adeninas además de secuencias TSDs semejantes a las encontradas en los LINE-1. Igualmente la integración se realiza en una secuencia consenso del ADN genómico reconocida por la endonucleasa del L1⁷². Los elementos SINE al igual que los otros elementos móviles, en su mayoría han acumulado mutaciones desde el momento de su inserción en la línea germinal de un antiguo ancestro humano. Semejante a lo encontrado en los elementos LINEs, muchos SINEs también presentan un extremo 5' truncado.

Elementos Alu

Los elementos *Alu* son los elementos SINEs más abundantes e inidentificados en el genoma humano, comprendiendo alrededor de 1.1 millones de copias dispersas por todo el genoma². Son llamados así debido a que muchos de ellos contienen un sitio de reconocimiento único para la enzima de restricción *Alu*.

Análisis de la organización de los elementos *Alu* explica el gran misterio de su distribución genómica sorprendente, y sugiere que puede haber una fuerte selección en favor de la retención preferencial de elementos *Alu* en regiones ricas en GC y que estos elementos “egoístas”, pueden ser benéficos para sus hospedadores humanos². Sorprendentemente, los *Alus* más recientes muestran una preferencia por regiones de ADN ricas en AT semejante a lo visto en los LINEs, mientras que los *Alus* progresivamente mayores muestran un sesgo cada vez más fuerte hacia ADN rico en GC².

Los elementos *Alu*, se clasifican en tres grandes subfamilias: J, S e Y (figura 1-1), y estas a su vez están divididas en subgrupos más específicos (revisado en referencia 71)⁷³. Los miembros de la familia J representan las versiones más antiguas de las secuencias *Alu*, seguidos de la subfamilia S, por lo que la subfamilia Y se ha identificado como la más joven del linaje *Alu*⁷⁴. Igualmente esta última es la responsable de la mayoría de eventos de retrotransposición de los *Alu* presentes en el genoma humano; se ha demostrado que un genoma

promedio humano puede contener miles de núcleos de elementos Alu activos provenientes tanto de la subfamilia Y como de la S⁷⁵.

La longitud aproximada de las secuencias Alu es de 300 pb⁷⁶ y son derivadas del gen 7SL ARN que forma parte del complejo ribosomal⁷⁷ (figura 1-3). Un elemento Alu completo consta de una estructura dimérica de dos monómeros altamente semejantes unidos por una secuencia ligadora rica en adenina. Poseen una secuencia promotora tipo 2 de la polimerasa III finalizando con una secuencia o señal de poli A. (revisado en referencia 71)⁷³. En el extremo 3' los elementos Alu activos contienen una secuencia terminadora de la RNA polimerasa III usualmente constituida por cuatro nucleótidos de timina⁷⁴.

Los elementos Alu han sido unos mejores colonizadores genómicos que los L1s. Se ha sugerido que las actividades endonucleasa (EN) como reversotranscriptasa (RT) del ORF2p del L1 están disponibles para los Alus por proximidad celular¹⁶. Además se ha observado que en los eventos de retrotransposición son influenciados por la interacción de un heterodímero formado por dos partículas proteicas de reconocimientos de señal (SRP9 y SRP14) que se unen al ARN de los elementos Alu^{75; 78; 79}. Este complejo posiblemente se une a la subunidad grande ribosomal, dejando la cola de poli A del elemento Alu próxima a la proteína ORF2p del L1⁵⁰. Es posible que la estructura del complejo ARN y proteínas juegue un papel determinante en la efectividad de la retrotransposición (RTN) además de la especificidad del sitio de inserción⁵⁰. El elemento Alu con mayor tasa de movilización in vitro es el elemento AluYa5⁸⁰.

Elementos SVA

Otra familia de retrotransposones no autónomos y evolutivamente jóvenes son los elementos SVA (SINE-R/VNTR/Alu) (figura 1-1), que surgieron en los linajes de primates hace aproximadamente 25 millones años^{81; 82}. Representan aproximadamente 0,2% con alrededor de 8,000–15,000 copias en genoma humano^{30; 73}, y a pesar de ello se piensa que probablemente son muy activos, siendo altamente polimórficos y la causa de varias disfunciones humanas^{81; 83}

Los elementos SVA están compuestos básicamente por: repeticiones de un hexanucleótido (CCCTCT) de longitud variable, dos fragmentos semejantes a los Alu invertidos, un número variable de repeticiones en tándem ricos en G-C, y una envoltura parcial con una secuencia LTR derivada de HERV-K10 extinto (figura 1-3). Los SVAs normalmente terminan en su propia señal de poli A, con inserciones genómicas que normalmente contiene un número de adeninas en el extremo 3'⁸⁴. Igualmente presentan secuencias TSDs flanqueantes semejantes a las presentadas por otros elementos conducidos por el L1⁸¹ (figura 1-3).

El ARN de los SVAs se ha propuesto que se transcribe por la polimerasa II aunque no se conoce ningún promotor interno para esta enzima. Además de ser ARNs no codificantes específicos de homínidos, poseen una longitud de entre 700 a 4000bp y son posiblemente movilizados por el LINE-1 humano en *trans* al igual que los elementos *Alu*^{81; 83; 85; 86} (figura 1-3)

Una de las características que ha salido a la luz recientemente, es la capacidad de ciertos elementos SVA de influenciar la expresión de genes reporteros en ensayos *in vitro*, por lo que en teoría estos elementos podrían tener la capacidad de alterar la expresión de genes que se encuentren próximos a ellos^{73; 87; 88}.

Pseudogenes Procesados

Además de los elementos móviles, copias de ADN de una amplia variedad de ARNm parece haberse integrado en ADN genómico de forma no autónoma⁸⁹ (figura 1-1). Debido a que estas secuencias carecen de intrones y no tienen secuencias flanqueantes similares a las copias de los genes funcionales, es evidente que no son simplemente genes duplicados que se han desplazado a una no-funcionalidad y convertirse en pseudogenes¹ (figura 1-3). Hay aproximadamente de 8,000 a 15,000 copias de pseudogenes procesados en el genoma humano, y la mayoría se derivan de genes que son altamente expresados en la línea germinal, como genes *housekeeping* y genes de proteínas ribosomales (revisado en referencia 27)³⁰.

Estos segmentos de ADN son copias retrotranspuestas en “*trans*” de ARNm maduros por las proteínas codificadas por el LINE-1^{90; 91}. Comparados con genes normales que codifican para un ARNm, estos segmentos insertados generalmente contienen múltiples mutaciones, las cuales se piensa han sido acumuladas desde que su ARNm fue reversotranscrito por primera vez e integrado al azar en el genoma de una célula germinal ancestral. Estas copias de ADN genómico no funcionales derivadas de ARNm se les denomina pseudogenes procesados^{90; 92}.

La mayoría pseudogenes procesados no son operativos ya que carecen de un promotor funcional. Por lo tanto, pueden ser utilizados como relojes moleculares para estimar las tasas de mutación entre especies. Sin embargo, algunos pseudogenes procesados humanos pueden expresarse, y un reducido número pueden codificar genes funcionales o servir como fuentes de pequeños RNAs de interferencia (siRNAs) con funciones de regulación de genes (revisado en referencia 30)³⁰.

Hasta la fecha, no existen ejemplos de eventos de retrotransposición *de novo* con pseudogenes procesados que causen enfermedades humanas. Sin embargo, la expresión de un pseudogen procesado ha sido implicado en pacientes con distrofia facioescapulohumeral humana⁹³. Del mismo modo, la expresión del pseudogen procesado FGF4 se asocia con condrodisplasia en 19 razas de perros, en concordancia con la idea de que la cría selectiva puede enriquecer alelos raros mutagénicos mediados por la inserción del L1⁹⁴.

Las TSDs que flanquean los pseudogenes procesados soportan la idea de que estas secuencias han sido generadas por eventos de retrotransposición extraños que involucran ARNs celulares. Las proteínas codificadas por el LINE-1 igualmente están implicadas en la movilidad de otras secuencias que son

moderadamente repetidas, que corresponden a copias parciales o mutadas de genes que codifican ARNs pequeños nucleares (snRNA) y de transferencia (ARnt) de los genomas mamíferos. Al igual que los genes procesados, estas copias no funcionales son flanqueadas por pequeñas secuencias repetidas y probablemente sean resultado de un evento de retrotransposición raro que se ha acumulado a través de la evolución^{95; 96}.

RETROTRANSPOSICIÓN EN EL GENOMA HUMANO

Desde la perspectiva de un transposón, tiene más sentido ser más activo en líneas germinales o en fases sexuales del ciclo de vida de un organismo huésped, que en células somáticas, si se tiene en cuenta que de estas células derivarán las nuevas generaciones y así se aseguraría su transmisión. En contraste, la RTN en células somáticas será evolutivamente no productiva, debido a que los nuevos eventos de transposición se perderán cuando el organismo muera. Teniendo en cuenta lo anterior es de esperar que los no-LTR retrotransposones sean expresados y retrotranspuestos en mayor medida en células germinales o durante el desarrollo embrionario temprano⁹⁷⁻⁹⁹.

Aunque se estima que cerca del 5% de los nuevos nacimientos pueden contener un evento de RTN *de novo* del L1^{100; 101}, se sabe poco acerca del momento en el desarrollo o el tipo celular que acomodan la RTN del L1 en humanos. Estudios *in vivo* en modelos de ratón indican que la expresión y/o la RTN del L1 puede ocurrir en células germinales masculinas o femeninas, durante etapas tempranas del desarrollo y en tejidos somáticos seleccionados¹⁰²⁻¹⁰⁵. Ejemplos de esto han sido los observados en células progenitoras neurales de rata (mNPCs) y en cerebro de ratones transgénicos los cuáles han soportado eventos de RTN usando L1 humanos¹⁰⁵. Así estos resultados indican que nuestro genoma pluripotente es un sustrato natural para el proceso de RTN, aunque en las células germinales las inserciones del L1 parecen ser menos frecuentes^{106; 107}. Sin embargo, recientemente la retrotransposición del L1 ha sido asociada con la viabilidad de los oocitos fetales en ratones¹⁰⁸. Igualmente se ha demostrado que las células madre embrionarias humanas (hESCs) expresan elementos L1 endógenos y pueden acomodar *in vitro* la RTN de L1s modificados genéticamente al igual que en células pluripotentes de carcinoma embrionario humano (hECs)^{109; 110}. Estos eventos de RTN pueden ocurrir entre genes y pueden resultar en deleciones del ADN genómico en el sitio de inserción. Datos que refuerzan la idea de que la RTN del elemento L1 puede ocurrir durante estadios tempranos del desarrollo humano, por lo que algunos individuos en la población pueden ser mosaicos genéticos con respecto a su contenido de L1¹⁰⁹. Además mediante el uso del promotor antisentido conservado (ASP) localizado en el extremo 5' del LINE-1, se pudo observar que varias subfamilias del L1 pueden estar expresándose en las células madre embrionarias humanas (hESCs)¹¹¹.

Más recientemente, y aprovechando las células madre humana pluripotentes inducidas (hiPSCs) producidas por la expresión ectópica de algunos factores de transcripción¹¹², se ha demostrado que al igual que las hESCs, las estas células se caracterizan por la sobreexpresión de L1-RNP debido a cambios en

los patrones de metilación en la región “5’ UTR” del L1¹¹³. Además, el desarrollo de hiPSCs a partir de células somáticas diferenciadas permitió evaluar si el estado pluripotente de una célula, que contiene el mismo fondo genómico, podría influir en la tasa global de RTN de un L1 modificado genéticamente¹¹⁴. La caracterización de un pequeño número de inserciones de L1s modificados genéticamente en hESCs e hiPSCs reveló características canónicas de la RTN del L1, incluyendo inserciones dentro de genes y alteraciones menores en el sitio de inserción^{109; 113}. Recientemente se ha observado movilización del L1 endogenamente en estas células, datos que dan soporte a las observaciones hasta ahora realizadas *in vitro*¹¹⁵.

Por lo tanto, y aunque los datos *in vitro* sugieren que la movilización del L1 puede afectar el genoma de estas células de importancia biomédica, se requieren estudios futuros para demostrar definitivamente que los L1s se mueven endógenamente en estas células, además que se hace relevante definir el impacto de las inserciones *de novo* de los L1 en el funcionamiento y regulación en los genomas pluripotentes. Teniendo en cuenta lo anterior tanto las hESCs como las hiPSCs son consideradas modelos del desarrollo embrionario debido a que imitan las primeras etapas del desarrollo humano y por tanto un escenario natural para las investigaciones relacionadas con las biología del L1. En el caso de las hESCs, son células pluripotentes derivadas de la masa celular interna de los blastocistos humanos¹¹⁶ que pueden auto-renovarse indefinidamente y someterse a diferenciaciones para generar tipos celulares de cada una de las tres capas germinales embrionarias: endodermo, mesodermo y ectodermo.

La identificación de la inserción de un L1 en el exón 6 del gen CHM (Choroideremia (RabEscortProtein 1), que desencadenó un caso de coroideremia (pérdida progresiva de la visión), debido a alteraciones en el *splicing* del ARNm del CHM¹¹⁷, permitió determinar de manera natural que dicha inserción se realizó durante la embriogénesis temprana de la madre del paciente, antes de la segregación de la línea germinal de los linajes somáticos.

Como se mencionó anteriormente, sorprendentemente y aunque en niveles menores, la transposición del L1 también puede ocurrir en células somáticas¹¹⁸⁻¹²¹, en este caso la secuencia transpuesta solo es heredada por la células hijas de dichas células. En algunas ocasiones este tipo de movimiento puede desencadenar una mutación en la célula somática, causando defectos fenotípicos perjudiciales para el organismo, como por ejemplo inactivaciones de genes supresores de tumores¹¹⁸. La reestructuración del genoma mediada por la actividad del L1 puede jugar un papel importante en la respuesta del hospedador, facilitando la adaptación de las poblaciones y especies a afrontar los cambios ambientales. Esto se ha podido observar en análisis de los EMS cuando los organismos hospedadores han sido sometidos a diferentes tipos de estrés, dando como resultado un incremento en la actividad de transposición¹²².

En resumen, la retrotransposición en humanos tiene lugar al menos a dos niveles diferentes: bien sea en el genoma de la línea germinal, en forma de nuevas inserciones acumuladas durante la embriogénesis, y/o en los tejidos

somáticos seleccionados (sanos o no) ¹¹⁴. Por tanto, estos datos revelan que los requerimientos mínimos necesarios del huésped para llevar a cabo la movilización del L1 están presentes dentro de estas células.

CONSECUENCIAS DE LA RETROTRANSPOSICIÓN DEL L1 EN EL GENOMA HUMANO

Aspecto generales

Dada las razones anteriormente expuestas, la movilidad de los L1-RC tanto en células germinales como somáticas ha resultado en una variedad de desórdenes genéticos, como es el caso de la hemofilia A, la distrofia muscular y el cáncer ^{100; 123}.

A pesar que los elementos móviles no parecen tener una función directa más que la de mantener su propia existencia, su presencia posiblemente tuvo un profundo impacto en la evolución de los organismos que actualmente conocemos. En mamíferos, sin embargo, los elementos móviles causan una proporción de mutaciones espontáneas aproximadamente del 10% en ratones y sólo del 0,1 al 0,2 % en humanos¹. Aun así, los elementos móviles han sido encontrados en alelos mutantes asociados a diferentes tipos de enfermedades humanas¹. Es por esto que la actividad mutagénica de los EMs ha contribuido a la diversidad genética de los organismos promoviendo diferentes mecanismos que con el paso del tiempo han sido modeladores de genomas. Por tanto, los EMs han jugado un papel importante en la evolución de los diferentes genomas y son considerados como una de las mayores fuerzas evolutivas ^{124; 125}.

En los linajes principales de los eucariotas superiores, la recombinación homóloga entre elementos móviles dispersos a lo largo de los genomas ancestrales puede haber generado duplicaciones de genes, generación de nuevos genes a partir de exones existentes y otros tipos de reordenamientos del ADN durante la evolución. Entrecruzamientos desiguales entre elementos móviles localizados en intrones de un gen particular, podría causar la duplicación de exones del mismo gen. Este proceso probablemente fue el que influyó en la evolución de los genes que contienen múltiples copias de exones similares, que codifican para dominios de proteínas similares, tal es el caso del gen de la fibronectina¹. Además las recombinaciones ectópicas entre múltiples copias de elementos móviles pueden ser desencadenantes de reajustes cromosómicos con efectos mayores.

Tanto en los ADN transposones como los LINE retrotransposones se ha demostrado que ocasionalmente llevan consigo secuencias flanqueantes no relacionadas cuando ellos se insertan en sitios nuevos y este tipo de mecanismos ha contribuido igualmente al exón *shuffling* durante la evolución de genes contemporáneos ¹²⁶,

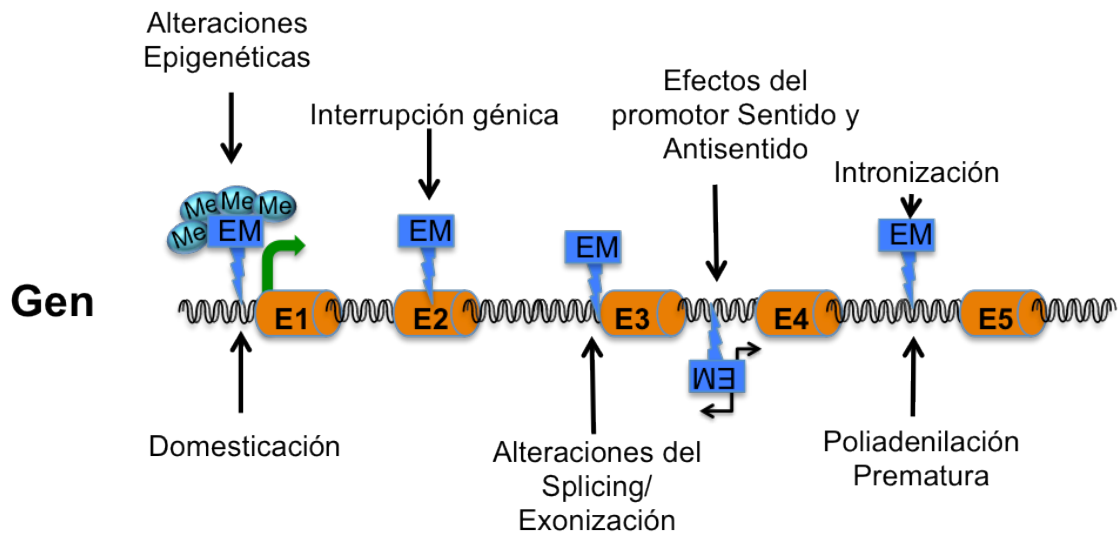


Figura 1-7. Efectos de los elementos móviles (EMs). En el esquema se muestran los diferentes efectos que un elemento móvil puede causar tras una inserción de novo en un gen. Los rectángulos azul oscuro representan las inserciones de los EMs, los cilindros en naranja representan exones unidos por intrones en negro. Las esferas azul claro representan grupos metilo unidos al ADN del EM (adaptado de Hulme et al., 2006; Goodier y Kazazian, 2008; Beck et al., 2011; Hancks y Kazazian, 2012).

Igualmente las inserciones de los elementos móviles actúan promoviendo la formación de regiones reguladoras más complejas aportando un control más polifacético de la expresión génica (domesticación) (figura 1-7). Aunque la expresión génica es igualmente influenciada por el tejido y los factores de transcripción específicos del estado del desarrollo, los mecanismos epigenéticos también juegan un importantes rol en la expresión de los genes diana además de aquellos que se encuentran en su vecindad¹²⁷. Estas consideraciones sugieren que el concepto inicial de los elementos móviles el cual los trataba como parásitos moleculares totalmente egoístas pierde mucho peso (figura 1-7).

Además de causar cambios en secuencias codificantes en el genoma, la recombinación entre elementos móviles, y la transposición de ADN adyacente a un transposón o retrotransposón, probablemente jugó un papel significativo igualmente en la evolución de secuencias reguladoras que controlan la expresión génica. Como se mencionó anteriormente, los genes eucariotas poseen regiones que controlan la transcripción llamadas potenciadoras o “*enhancers*”, que pueden actuar desde decenas de kilobases de distancia. La transcripción de muchos genes es controlada a través de efectos combinados de varios *enhancers*. Inserciones de los elementos móviles cerca de estas regiones controladoras de la transcripción posiblemente contribuyeron a la evolución de nuevas combinaciones de secuencias reguladoras. Estos a su vez, controlan qué genes específicos son expresados en un tipo celular particular y qué cantidad de proteína codificada se produce en los organismos modernos¹.

Aproximadamente el 25% de las regiones promotoras humanas contiene secuencias derivadas de EMs además de numerosos factores de transcripción

correspondientes a motivos de unión a ADN (revisado en referencia 119)¹²². Además estas regiones pueden estar afectadas por cambios epigenéticos causados por inserciones de los EMs debido a desbalances en los estados de metilación presentes en el ADN hospedador¹²⁸ (figura 1-7). La inserción de EMs en regiones reguladoras puede ser la causa de diferentes enfermedades humanas tales como cáncer o la talasemia¹²².

Otra forma por la que los EMs pueden modificar la expresión génica es proporcionando elementos en *cis* a las secuencias promotoras y regiones enhancers; secuencias derivadas de la domesticación de los EMs o por exaptación^{125; 129} (figura 1-7). Cada vez es más clara la contribución de los EMs en la evolución de las regiones codificadoras de proteínas, por lo que se estima que la contribución de los EMs a los genes codificantes de proteínas por medio de la domesticación de secuencias es alrededor del 4%¹²².

Más allá de los efectos citogenéticos causados por los elementos transponibles en el tamaño del genoma y los reajustes cromosómicos, las inserciones del L1 pueden tener efectos más sutiles afectando sólo a la expresión de genes. La alteración génica es quizás el efecto más evidente de las inserciones de los EMs, estas incorporaciones pueden resultar frecuentemente en desordenes genéticos pudiendo causar terminaciones prematuras deletéreas de una secuencia peptídica. Sin embargo los EMs también pueden regular la expresión génica de otras formas diferentes.

Alteración de la Expresión Génica por la Inserción del L1

En efecto, estos elementos tienen la capacidad de insertarse entre genes o en elementos reguladores de genes, hechos que podrían causar la interrupción de la unidad génica y/o el mal funcionamiento de la misma (figura 1-3).

La inserción del L1 en cualquier orientación podría atenuar la expresión de los genes disminuyendo la producción del ARNm¹³⁰. Debido a que las largas inserciones tienen efectos inmediatos en la expresión de los genes, muy probablemente por ser deletéreas no sea seleccionado en la evolución¹³⁰. Con respecto a las inserciones del L1 entre regiones codificantes, se sabe que regiones libres de EM son significativamente asociadas a genes que codifican para productos reguladores del desarrollo o factores de transcripción¹³¹. Igualmente genes que codifican para proteínas estructurales y ribosomales son desprovistos de EMs tanto en regiones intrónicas como en espacios intergenicos¹³², aunque el efecto de estas secuencias insertadas dentro de los genes, principalmente intrónicas, no ha sido estudiado cuidadosamente, ya que inicialmente se había pensado que estas secuencias eran eliminadas por splicing sin llegar a interferir en el funcionamiento del gen. La inserción de los elementos móviles y la fijación en los intrones mamíferos son influenciadas por la función génica, los niveles de expresión y la conservación de la secuencias¹³².

Aunque se sabe ampliamente que las inserciones del L1 están involucradas en el moldeamiento y funcionamiento de los genomas mamíferos, el

entendimiento del impacto del L1 sobre la expresión de los genes aun esta lejos de ser comprendido, dado a los múltiples factores que intervienen en el momento de una nueva inserción¹³³. Brevemente se analizaran algunos conceptos acerca de los posibles escenarios en lo que el L1 interactúa a la hora de interferir en el funcionamiento y/o expresión génica una vez la nueva inserción toma lugar en el genoma.

Movilización de ADN genómico (transducción)

Por medio de la RTN se pueden movilizar secuencias en 5' o 3' de ADN genómico que no pertenecen a los retroelementos, bien sea por el uso del promotor de la ARN pol II ó por hacer una extensión de la transcripción haciendo puente con la señal débil de la cola de poli A del L1 hasta encontrar una señal fuerte del ADN genómico (figura 1-8A). En el primer caso puede haber dos rondas sucesivas de retrotransposición y en ambos casos se produce un reajuste o reordenamiento debido a una inserción de un fragmento de ADN genómico en una nueva ubicación genómica. A este mecanismo se le conoce con el nombre de *transducción*^{134; 135}.

Eventos de transducción en 3' del L1 ocurren de forma frecuente y podrían en principio movilizar exones y promotores a una nueva localización cromosómica, proporcionando un posible mecanismo de exón *shuffling*^{126; 136}. Se piensa que alrededor del 25% de los eventos de retrotransposición están acompañados de secuencias transducidas en 3'¹³⁷.

Formación de exones, intrones y formas alternativas de los transcritos

El LINE-1 forma un componente de la mayoría de las unidades de transcripción de mamíferos (aproximadamente 79%) además estos segmentos ya sean derivados de inserciones de L1s pre-existentes o recientes usualmente contienen secuencias del tercer marco abierto de lectura de L1 (ORF2)^{29; 130}. Igualmente se ha observado que la mayoría de los exones humanos recientes son semejantes y han sido originados a partir de secuencias altamente repetidas como los EMs¹³⁸ por un proceso denominado *exonización* (figura 1-8B)¹³⁹. Esto es, un exón interno en el que un L1 constituye una fracción o la secuencia completa del mismo¹⁴⁰. Igualmente las inserciones del L1 en una secuencia intrónica, puede dar lugar a su activación en forma de exón, modificando las formas alternativas de splicing y de esta manera constituir un nuevo evento de exonización¹⁴¹. Si un L1 se inserta en una secuencia codificante, tiene el potencial de producir un ARN truncado y de esta forma una nueva isoforma para un transcrito de ARNm¹⁴².

Las inserciones del L1 también pueden modificar el ARNm, al insertarse dentro del primer o último exón de un gen. Esta inserción provoca un alargamiento de los exones (primeros o últimos) que suelen ser parte de una región UTR, o una activación de un intrón alternativo, proceso que se ha denominado *intronización*¹⁴⁰. Lo anterior es de especial importancia si se tiene en cuenta que en humanos el 95% de los genes son multiexónicos y que producen al menos 2 isoformas por inclusión diferencial de exones¹²⁷.

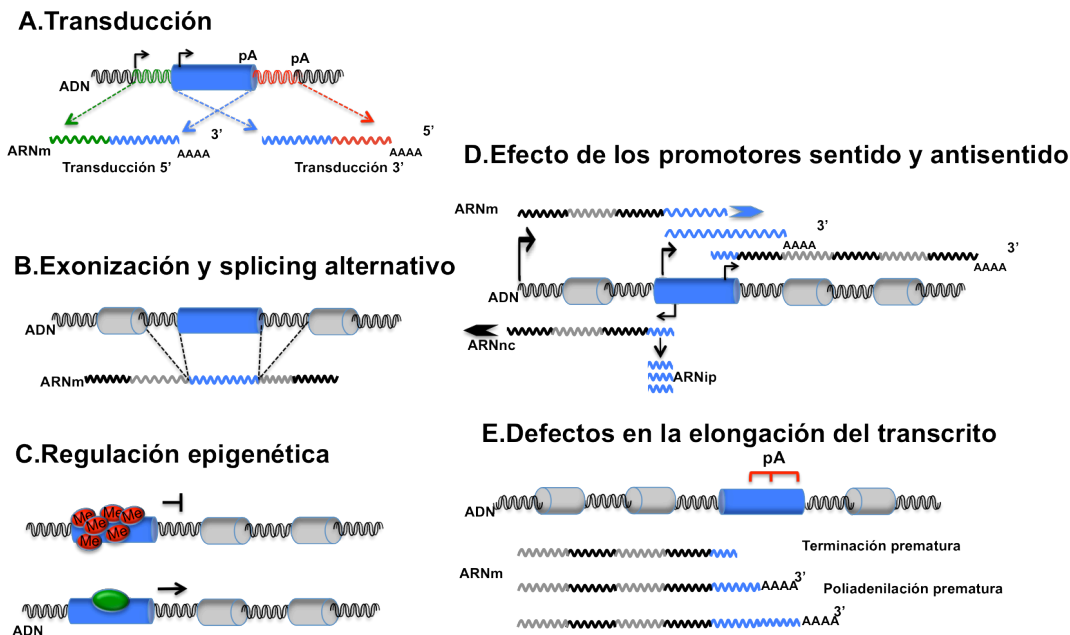


Figura 1-8. Efectos de inserciones *de novo* del L1 sobre el ADN genómico. En el esquema se muestran algunos ejemplos de los posibles efectos que puede causar la retrotransposición del L1 (azul) en el transcriptoma celular (exones en gris e intrones en negro). A) Transducción en 5' cuando se transcribe el L1 a partir de un promotor endógeno aguas arriba de la inserción, dando como resultado una fracción de ADN genómico (verde) seguido de las secuencias del L1. Transducción en 3', un transcrito es formado a partir del promotor del L1 finalizando con una señal de poliadenilación más fuerte aguas abajo del L1 generando un transcrito con el L1 seguido de un fragmento de ADN genómico (rojo). B) Transcritos aberrantes causados por los sitios de splicing presentes en el L1 cuando este es insertado dentro de un intrón dando como resultado la creación de exones por transcritos alternativos del ARNm. C) Efectos de silenciamientos y activaciones debido a la isla CpG presente en la región promotora del L1 que es susceptible de metilaciones y desmetilaciones. D) Cambio en el nivel de expresión de los genes debido a la acción de los promotores sentido y antisentido del L1. E) Formación de ARNm alternativos ó prematuras terminaciones debido a las señales internas de poliadenilación del L1. (Adaptado de Singer et al; 2010).

Modificaciones Epigenéticas

Una activación ectópica de un L1 que ha sido previamente silenciado puede causar cambios en la expresión de los genes colindantes alterando la sincronización o la especificidad de tejidos de sus expresiones génicas que posiblemente pueda resultar en una variabilidad somática importante (figura 1-8C)¹⁴³. Por la presencia de islas CpG en la fracción 5' del retroelemento L1, estas son dianas para la metilación de ADN, por lo que pueden inducir una estructura cromatínica restrictiva, que puede extenderse a las secuencias codificantes cercanas, trayendo como consecuencia su silenciamiento (figura 1-8C).

Teniendo en cuenta las señales de poliadenilación presentes en la secuencia del L1, estas pueden ser usadas por el hospedador alterando la regulación epigenética del gen. Las poliadenilaciones alternativas incrementan la diversidad del transcriptoma generando múltiples isoformas de transcripción para un simple gen. Este proceso se ha pensado que puede estar sujeto a regulaciones epigenéticas con base en observaciones en cerebro de ratón en

donde alteración de las marcas epigenéticas de un gen incrementaba el uso de sitios de poliadenilación alternativos, alterando de esta manera el splicing de los exones del gen¹²⁷.

Sin embargo la importancia de las marcas epigenéticas en la regulación de la poliadenilación alternativas permaneces ampliamente inexploradas en contraste con la evidencia relativamente abundante para el control epigenético de la elección de promotores alternativos y mas recientemente de splicing alternativo¹²⁷.

Efectos de las regiones promotoras

Como se mencionó anteriormente, dentro de las secuencias del L1 poseen secuencias promotoras de la ARN pol II en dirección sentido y antisentido del L1. La transcripción bidireccional a partir de los promotores en la región 5' puede producir transcritos quiméricos, ARNs no codificantes, ARNm antisentido o ARNs de doble cadena, afectando esto a la expresión de los genes en diferentes formas (figura 1-8D)^{130; 144}. Debido a que la mayoría de las inserciones de los retroelementos están truncadas en la región 5', muchas de estas inserciones carecerán de las regiones promotoras presentes en 5', lo que en teoría impediría una alteración de la expresión génica por medio de los promotores en 5'. A pesar de esto, por medio del promotor sentido presente en la región 3' del L1 es posible el inicio de la transcripción corriente abajo de la secuencia. Por tanto este promotor alternativo tiene la capacidad de aumentar la transcripción de una secuencia codificante cercana (figura 1-8D)¹⁴⁵.

El uso de ambas regiones promotoras cuando el L1 es insertado dentro de regiones codificantes, podría dar lugar a la individualización de una misma proteína, en el que un transcrito inicial sea dividido en partes individuales, realizándose las transcripciones individuales en cada región promotora del L1, a este fenómeno se le ha dado el nombre de *Gene-Breaking*¹⁴⁶.

Defectos en la Elongación

Los retroelementos humanos tanto L1 como Alu, pueden introducir señales de poliadenilación intragenicas¹⁴⁷. De hecho los elementos L1s incluyen varias señales de poliadenilación dentro de su propia secuencia¹³⁰, además que en la colas de poli A ya sea en los LINEs como en los SINEs es creada una secuencia común AATAAA para ambas¹²². La presencia de esta señal de poliadenilación puede causar en determinado casos la terminación prematura de la transcripción modificando de esta manera la longitud del transcrito primario de un gen, al igual de los efectos anteriormente mencionados (figura 1-8E).

Ensayos in vitro pudieron determinar que los L1 ubicados en dirección antisentido inhiben la transcripción a causa de la poliadenilación prematura, dato que es relevante si se tiene en cuenta que la mayoría de las inserciones del L1 ubicadas en regiones codificantes están en orientación antisentido con respecto al gen hospedero¹³⁰. A pesar de esto se sabe que las inserciones del L1 independiente de la orientación, puede afectar la producción de ARN de los

genes endógenos tanto cualitativa como cuantitativamente ya sea por la inhibición de la transcripción en dirección sentido, o por poliadenilaciones prematuras causadas por las inserciones antisentido¹³⁰.

Teniendo en cuenta lo planteado anteriormente con respecto a los efectos moleculares que desencadenan las inserciones del L1 en el genoma humano, se plantean muchas preguntas acerca de la naturaleza de la relación entre elemento móvil y hospedero que aun permaneces sin ser resueltas, debido básicamente, a que el escenario donde se integran todas las fuerzas de selección que están afectando la fijación de los elementos móviles en los genomas mamíferos aún permanece ausente¹³².

REGULACIÓN DE LOS ELEMENTOS MÓVILES

Mientras que el éxito de los EMs depende de la transposición en una línea germinal, el daño genómico de un exceso de transposición podría traer consecuencias negativas sobre el huésped, por lo que las células hospedadoras generalmente están constantemente reprimiendo su actividad. Esto implica que las células germinales han sido evolutivamente un sitio hostil en la lucha entre el EM y el huésped, en donde la pérdida de la represión de la movilidad de los EMs puede traer consecuencias dramáticas para cualquier tipo de célula. La sobre-regulación de un elemento transponible se asocia con defectos en el desarrollo de la célula germinal, evidenciando la importancia del silenciamiento de los EMs^{148; 149}. Esto fue observado en ensayos en células germinales de ratón donde la desregulación del L1 mostró un defecto en la progresión meiótica en la profase temprana de la meiosis I, además una pérdida marcada y progresiva de las células germinales con la edad¹⁴⁸.

Con el fin de controlar los efectos deletéreos que pueden ser causados por la RTN del L1, las células huésped han desarrollado diferentes mecanismos que les ha permitido reducir su tasa de RTN y de esta manera aumentar las probabilidades de supervivencia. Según el proceso involucrado, de una manera muy general, los mecanismos de control actúan por medio de modificaciones epigenéticas y regulaciones de los transcritos producidos por el L1, incluyendo procesos metabólicos relacionados con el ARN(figura 1-9).

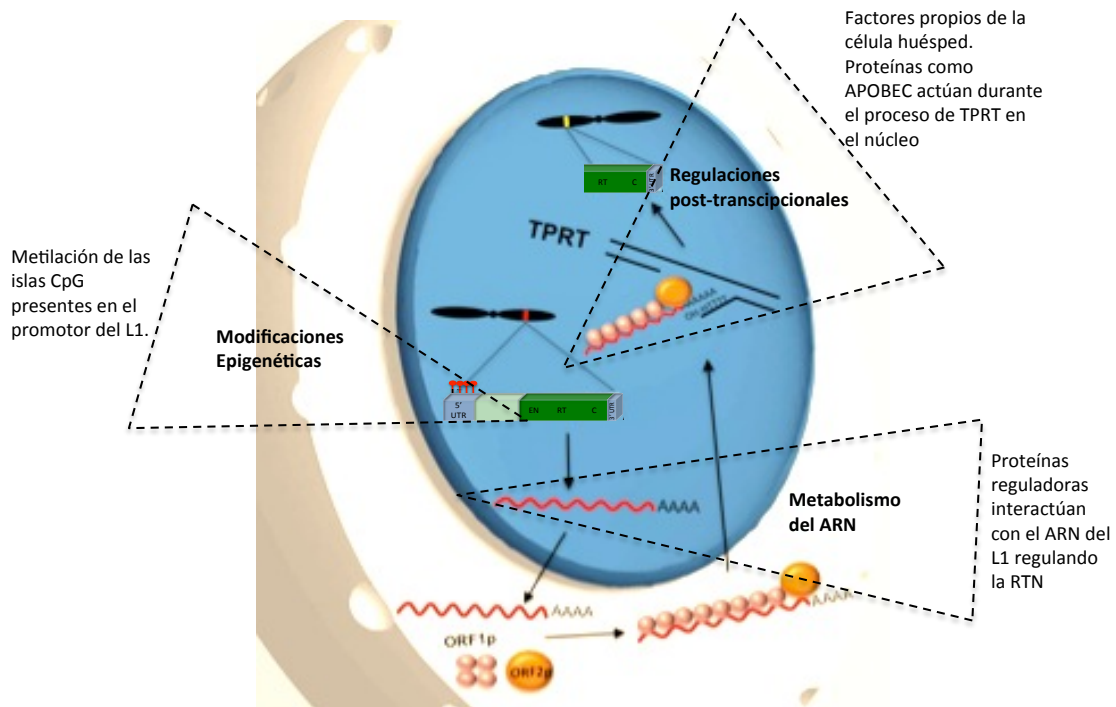


Figura 1-9. Control de la retrotransposición del L1 por el huésped. El L1 es susceptible de ser inhibido en todas las etapas de su ciclo de movilidad en la célula hospedadora, en la gráfica se esquematizan en que puntos del ciclo de retrotransposición puede ser atenuado.

Regulación Epigenéticas

Es bien sabido que los elementos L1 pueden ser controlados por mecanismos epigenéticos. Estos son silenciados ya sea por metilación del ADN ó por modificaciones de las histonas presentes en el nucleosoma, e igualmente son activados cuando dichas acciones o modificaciones desaparecen¹⁵⁰.

Para la movilización de un L1-RC es necesaria la transcripción de la copia completa del ARNm del L1 a partir de su promotor interno ubicado en la región 5'. Teniendo en cuenta esto, el control de la transcripción del L1 representa una de las formas más efectivas de regular la continua movilización de los retrotransposones LINE activos en humanos, y por tanto el control de la movilización de los elementos no autónomos SINE.

Es bien conocido que los transposones están asociados con la metilación del ADN, y que esa metilación juega un papel importante en el silenciamiento del transposón a largo plazo^{33; 151}. La metilación de las islas CpG ubicadas en el extremo 5' UTR, está asociada con la supresión de la expresión del L1 en varios tipos celulares¹⁵¹. Es importante anotar que la metilación del ADN que se establece en las células germinales primordiales, se mantiene durante toda la vida de un organismo, y se cree que controla la expresión del LINE-1 en los tejidos somáticos^{152; 153}. Estudios adicionales han mostrado que la metilación de los EMs varía entre tipos celulares, siendo estas modificaciones epigenéticas relacionadas con el patrón global de la regulación génica^{49; 154} (figura 1-10).

Debido a las posibles alteraciones que pueden causar la RTN de los EMs, tanto las células germinales como las células madre adultas, estas mantienen la homeostasis del tejido a través de un fino programa de respuestas a señales fisiológicas y relacionados con el estrés. El caso de PLZF (Promyelocytic Leukemia Zinc Finger protein), un miembro de la familia de factores de transcripción POK, actúa como un regulador epigenético de mantenimiento de células madre, en células germinales y células madre hematopoyéticas, se ha identificado que su diana principal es el L1 provocando su metilación induciendo el silenciamiento de los genes del L1 de copia completa y de esta forma inhibiendo la RTN¹⁵⁵

Actualmente se piensa que la serie de hipometilaciones comienza durante la embriogénesis temprana, hecho que representaría una gran oportunidad para la acumulación de nuevas copias del L1¹⁵⁶. De acuerdo a esto, se pudo determinar que las células madre embrionarias humanas (hESCs) y las células madre inducidas pluripotentes (iPSc) poseen hipometilado la región del promotor de LINE-1^{109; 113; 156; 157}; igualmente y en particular, ambos tipos celulares se caracterizan por la sobre expresión de la ribonucleopartícula del L1^{109; 113}.

Con respecto a la metilación de las células madre, se ha encontrado en estudios en ratones que el ADN metiltransferasa-3 (DNMT3L) media la metilación selectiva del ADN de los retroelementos en las células germinales primordiales³³. Efectos semejantes se observado cuando se inhibe la expresión de la proteína Tex19.1 el cual ha visto anomalías en el crecimiento de la placenta en ratones a causa de la activación de retroelementos¹⁵⁸. Igualmente, en células madre somáticas como las células madre neurales (NSCs), la proteína MeCP2 está involucrada en la metilación global del ADN y en enfermedades del desarrollo neural humano; junto a factores de transcripción como SOX2 y HDAC1 forman un complejo represor sobre la región promotora del L1 controlando su transcripción neuronal y por ende la RTN^{151; 159; 160}.

En ensayos realizados en líneas embrionarias derivadas de células de carcinoma humano (hECs) PA-1 demostraron que inserciones *de novo* del L1 eran silenciadas rápidamente en el transcurso de su integración¹¹⁰. En este estudio se pudo demostrar que en dicho silenciamiento estaban involucrado cambios en el estado de la cromatina en el sitio de integración del L1, posiblemente causados por variaciones en los perfiles de acetilaciones de la cromatina, hecho que se puso en evidencia con el uso un inhibidor de la histona desacetilasa el cual pudo bloquear la inhibición de la RTN¹¹⁰. En el mismo estudio igualmente se observó una disminución del silenciamiento de la RTN del L1 cuando las células hEC eran sometidas a un proceso de diferenciación¹¹⁰.

Hay que tener en cuenta que las modificaciones epigenéticas del ADN e histonas pueden influir la transcripción en su iniciación alterando la accesibilidad a las secuencias promotoras. Además, regiones de heterocromatina pueden sutilmente impedir el progreso del complejo de la polimerasa reduciendo el resultado transcripcional¹⁶¹. Por tanto, las

modificaciones de histonas y sus factores asociados son los principales reguladores de la expresión del L1 especialmente en células pluripotentes con un contenido alto de hipometilación, tal y como sucede en las hESCs y iPSCs. De hecho la regulación del L1 durante la embriogénesis es un proceso dinámico donde múltiples factores pueden actuar y contribuir en la regulación de la RTN en humanos¹⁶².

En ESCs y en células primordiales germinales los estados epigenéticos son reajustados requiriendo que los patrones de metilación se restablezcan de una forma específica para que se pueda dar la diferenciación. En este tipo de células se han identificado varias sitios dianas de metilación entre los que se incluyen los elementos móviles, promotores, repeticiones pericentroméricas, y regiones controladas por *imprinting*¹⁶³. Con respecto a las modificaciones de histonas, en ambas células contribuyen al entorno característico transcripcionalmente permisivo de estos tipos celulares y a las reducciones en la metilación del ADN que experimentan. Por ejemplo, la pérdida global de las marcas de metilación represivas (H3K9), son un paso esencial en la reprogramación epigenética tanto en células PGC como en iPS^{164; 165}.

Es asumido que el complejo efector de ARNs asociados a *piwi* (piARNs) presentes en las células pluripotentes de ratón, de alguna manera recluta las ADN metil-transferasas DMNT3A, DMNT3B *de novo*, y su proteína asociada DMNT3L para metilar el locus del transposón^{166; 167}. Aunque los mecanismos mencionados no son específicos de los retrotransposones no-LTR, se aplican para todos los elementos móviles (figura 1-10).

Otra forma de regulación epigenética observada ha sido la relacionada con los micro ARNs (miARNs) celulares de mamíferos. Los miARNs mamíferos asociados a secuencias repetidas, son producto del procesamiento de ARN de doble cadena (dsARN) por medio de la proteína DICER, que son a su vez cargados en el interior de la proteína Argonauta 2 (AGO2)¹⁶⁸ (figura 1-10). En el caso de células madre embrionarias de ratón (mESCs), *knockouts* para DICER mostraron un estado de hipometilación del promotor del elemento L1 mientras que su transcripción, traducción y el número de copia se vio incrementado. Lo anterior deja en evidencia que los componentes de la biogénesis de los miARNs en mamíferos están ligados al silenciamiento transcripcional de los elementos móviles como es el caso del L1^{37; 169}.

Cabe destacar que estos mecanismos pueden operar durante la embriogénesis, pero sugieren que la severa hipometilación de los genomas después de la fecundación puede producir un exceso de transcripciones de los EMs y por lo tanto dar lugar a la saturación de los mecanismos de restricción, permitiendo la movilización permanente de los EMs en los seres humanos¹¹⁴.

Control por RNAs

El descubrimiento del ARN interferencia (iARN)¹⁷⁰ llevó a la hipótesis de que los ARNs pequeños están involucrados en el silenciamiento de los transposones por la vía de la metilación de ADN y la degradación del ARN del

transposón⁶⁸. Hecho que fue reforzado cuando fue observado que cepas mutantes de *C. Elegans* deficientes en iARN aumentaban la movilidad de los transposones endógenos¹⁷¹. Evaluaciones posteriores revelaron que mutantes en iARN podían de hecho conducir a la inhibición de la represión general del transposon^{171; 172}. Ahora se sabe que las células expresan varias clases de ARNs pequeños como son los: micro ARNs y los ARNs asociados a proteínas *piwi* anteriormente mencionados, además de los ARNs endógenos pequeños de interferencia (endo-siARN). Una fracción significativa de endo-siARNs y piARNs corresponden a secuencias de transposones, lo que implica un reconocimiento en base a su secuencia y un mecanismo de represión^{173; 174}.

La principal función de los piARN es la del silenciamiento de los EMs en la línea germinal, rol que está altamente conservado a través de las especies animales. Estos ARNs interactúan con miembros de la familia de proteínas Argonautas (PIWI) con el fin de llevar a cabo su función regulatoria¹⁷⁵. Los piARNs son específicos de las células germinales, mientras que los endo-siARNs son ubicuos¹⁷⁶.

Los piARN son generados por transcripción en ARN de largas regiones de EMs, resultando en la acumulación de piARN maduros cortos en el citoplasma como consecuencia de su amplificación por medio de un mecanismo denominado ping-pong¹⁷⁶. Los fragmentos generados luego actúan como guías en la destrucción de los transcritos complementarios del transposón por ruptura endonucleolítica¹⁷⁶ (figura 1-10).

En las células germinales la amplificación en ping-pong, se presume que no solo actúa cortando el ARNm guiándolo a una represión post-transcripcional, sino también sirve como generador de sustratos de secuencias específicas conduciendo a la metilación del propio transposón, proporcionando un vínculo genético entre la producción de piARNs y la eventual metilación del ADN del transposón^{177; 178} (figura 1-10). Lo anterior deja claro que existe una conexión entre sistemas de regulación, en donde se integra una respuesta transcripcional a un sistema de regulación epigenética¹⁶⁶.

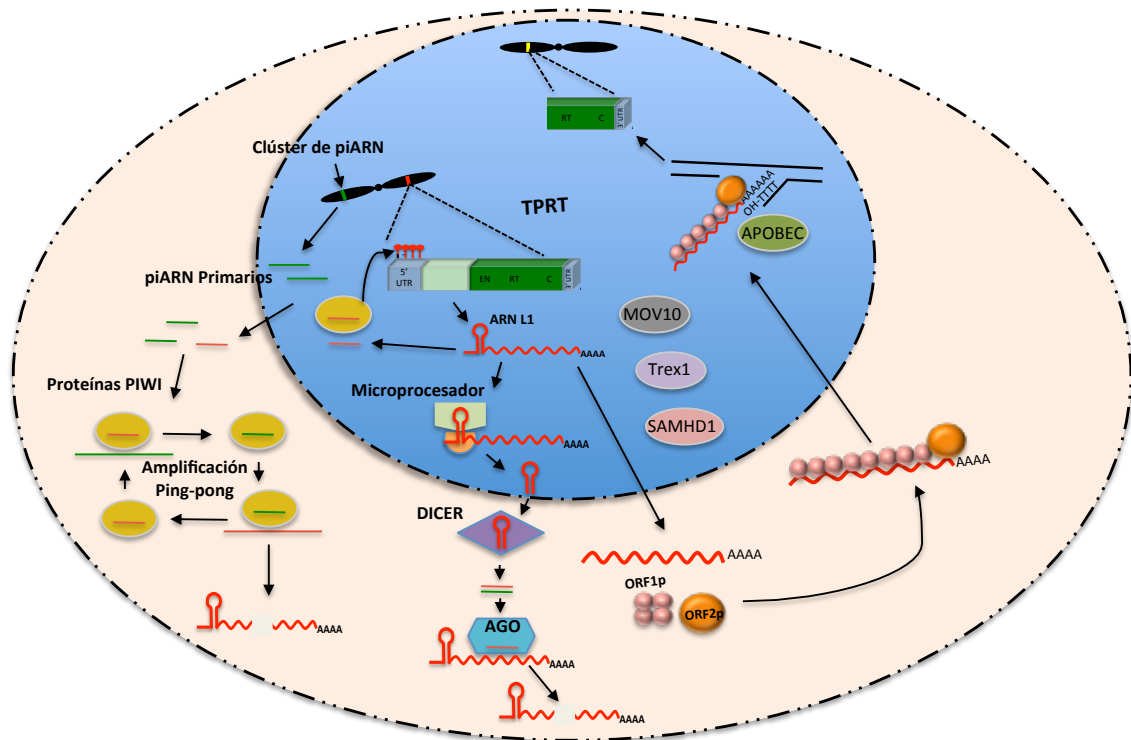


Figura 1-10. Mecanismos de control de la retrotransposición del L1 por el huésped. La célula huésped controla la RTN del L1 por diferentes medios: metilación del promotor, procesamiento de su transcrito además de interferir para que se lleve a cabo procesos como la traducción proteica al igual que el TPRT.

Recientes experimentos han demostrado otra forma de regulación complementaria en el control de la RTN del L1, en el que interviene la ruta de regulación mediada por los miARNs¹⁷⁹. El complejo encargado en la generación de los miARNs denominado *microprocesador* (figura 1-10), está conformado principalmente por dos proteínas: DGCR8 que se encarga del reconocimiento de los microARNs primarios (pri-microARNs) con una estructura en horquilla, y *Drosha*, que tiene como función el procesamiento de los pri-miARNs en la generación de los pre-miARNs. Estos últimos son exportados al citoplasma en donde la proteína DICER se encarga de generar los miARNs maduros (figura 1-10). El *microprocesador* es capaz de reconocer dominios de RNAs presentes en los L1 y posiblemente controle los niveles de expresión *in vitro* del L1 por procesamientos de sus transcritos¹⁷⁹.

Recientemente, varios estudios han definido la existencia de abundantes ARN largos no codificantes (lncRNAs) en los genomas de mamíferos. En general, los lncRNAs son un grupo heterogéneo de ARN recientemente catalogados, pero sus funciones biológicas no se habían determinado. Curiosamente, son más abundantes que los genes humanos conocidos, y la investigación está comenzando a revelar que estos ARNs participan en los procesos de regulación del genoma mucho más de lo previsto (revisado en referencia 112)¹¹⁴. Más recientemente, se ha demostrado que los ARNs producidos por un promotor antisentido activo situado en la secuencia del ORF1 de un L1-RC de ratón pueden controlar la retrotransposición de elementos L1 modificados genéticamente³⁶. Datos que revelan cómo los lncRNAs naturales producidos por el LINE-1s podrían ayudar a controlar su tasa de RTN.

Los lncRNAs se pueden clasificar en diferentes categorías, incluyendo: lincRNAs (del inglés long intervening lncRNAs, producidos a partir de sitios genómicos lejos de secuencias codificantes), NAT (Natural Antisense Transcripts, producidos a partir de la hebra antisentido de un gen de codificación) y lncRNAs intrónicos (transcritos a partir de intrones de genes codificantes)¹⁸⁰. Entre las funciones que se les ha atribuido está la del control de la expresión génica, el mantenimiento de la pluripotencia, la arquitectura nuclear¹⁸¹ y el splicing¹⁸⁰. Los lncRNAs son similares en algunos aspectos a los ARNm, ya que pueden, tener caperuza, someterse a *splicing*, y tener una cola de poli (A).

Además, es importante saber que las hESCs e hiPSCs recapitulan algunos aspectos de la regulación mediada por ARNs derivados de EMs; ejemplo de esto es la reciente descripción del control de la pluripotencia mediada por lncRNAs. De hecho, reportes anteriores demostraron que las hESCs expresan una gran cantidad de lncRNAs producidos por varios miembros de la subfamilia del LINE-1, probablemente debido a la hipometilación del promotor antisentido del L1^{113; 156}. Igualmente se ha sugerido, que los genes que codifican proteínas podrían utilizar secuencias de retrotransposones cercanas como promotores alternativos, generando ARNm alternativos y la producción de ARN no codificantes cortos y largos (ncRNAs) como se describió anteriormente¹⁸².

Control Post-transcripcional

El mecanismo de TPRT a menudo conduce a la retrotransposición de un ARNm de copia completa de un L1. Del mismo modo se ha encontrado que los ARNs humanos de copia completa contienen una secuencia donadora de splicing en la posición +98 de la región 5'¹⁸³ y se piensa que su uso puede estar guiando a la formación de ARNs más cortos que pueden estar comprometiendo la retrotransposición¹⁸³. Por tanto, el splicing del ARNm del L1 puede servir como mecanismo regulatorio en la restricción de la expresión del L1 y/o la retrotransposición de forma específica sobre un tejido.

Nuevas observaciones han demostrado que proteínas celulares de unión al ARN pueden tener un papel relevante como moduladoras post-transcripcionales de la RTN del L1. Tal es el caso de la ribonucleoproteína L heterogénea nuclear (del inglés hnRNPL), una proteína que facilita el splicing alternativo en la célula además de ser un agente modulador tanto del L1 humano como del de ratón^{184; 185}.

Como se ha mencionado anteriormente, un L1-RC contiene un potencial número de señales de poliadenilación en su secuencia, cuyo uso puede guiar la formación de ARNs más cortos que pueden comprometer la retrotransposición¹⁴⁷. De hecho se propone también que la naturaleza rica en adenina de la cadena sentido del transcrito del L1, al igual que las señales de poliadenilación mencionadas anteriormente, pueden actuar como reguladores moleculares para afinar la expresión de los genes que contienen L1s de copia completa¹³⁰.

La proteína C1 de unión a poli A (del inglés PABPC1) se ha visto que interacciona con el ARN del L1 además de ser requerida para la formación de la ribonucleopartícula (RNP) y concluir en la RTN¹⁸⁶. Otra proteína involucrada en la inhibición del L1 humano así como el elemento IAP en ratón es la RNasa L¹⁸⁷, aunque aún falta mucho por elucidar sobre el mecanismo de acción de estas proteínas de unión al ARN y su papel en RTN del LINE-1. Como en el caso de la metilación, otros factores se han identificado que inhiben la RTN en células pluripotentes, como es el caso de la familia de proteínas APOBEC3 la cual se ha identificado que limita la RTN del L1 tanto en células tumorales como en células pluripotentes¹⁸⁸. Esta familia de proteínas de desaminasas de citidina, estas involucradas en varios procesos de la defensa del genoma de las células hospedadoras contra los EMs. Actúan convirtiendo la citidina en uridina en la edición del ADN y ADNc del retrotransposón como una forma de defensa contra la retrotransposición¹⁸⁹.

Al igual que en el virus del VIH, la enzima APOBEC3G se ha observado que regula la RTN del L1 a nivel postranscripcional^{190; 191}. La desaminación de la citidina de la recién sintetizada hebra de ADNc del VIH por APOBEC3G resulta en una hipermutación G-A en la secuencia codificante viral. Dado que los retroelementos sintetizan una hebra de ADNc naciente similar, este hallazgo condujo a la hipótesis que los miembros de la familia APOBEC inhibían similarmente la RTN. En efecto, todos los miembros de la familia del APOBEC3 han mostrado en ensayos basados en líneas celulares que inhiben la actividad del L1^{192; 193}. Además, el gen AID (activation-induced deaminase), que comparte ancestro común con la proteína APOBEC3, también inhibe la retrotransposición del L1 en los ensayos con las mismas líneas celulares las cuales fue evidenciado la acción del APOBEC3¹⁹⁴(figura 1-10).

Probablemente existen al menos dos diferentes formas por las cuales los miembros de la familia del APOBEC inhiben los retrotransposones no LTR, bien sea dependiente o independiente de desaminasa. Tal es el caso del gen APOBEC3A (A3A) que se ha encontrado que inhibe fuertemente la RTN tanto del LINE-1 como del Alu en cultivos celulares¹⁹⁵. En el caso del L1 se ha demostrado que lo hace por la vía dependiente de deaminasa, además cumple la función de desaminar las cadenas sencillas de ADN que quedan expuestas transitoriamente durante el proceso del TPRT¹⁹⁶.

Con respecto a otras proteínas que interactúan con el ARN del L1, encontramos el caso de MOV10, una helicasa de ARN que actúan inhibiendo la acción de algunos retrovirus, también se ha encontrado que inhibe la RTN de los elementos LINE-1, Alu y SVA humanos y de IAP de ratón en cultivos celulares (revisado en referencia 73)⁷³. En el caso específico del L1, previene a la célula hospedadora de su RTN interactuando directamente con las ribonucleopartículas¹⁹⁷. MOV10 también es encontrada en complejos con miembros de la familia APOBEC3, y esta asociada a defensa contra retrovirus el cual tienen mecanismos de replicación similares a los retrotransposones^{198; 199}. Una proteína relacionada con MOV 10, conocida como MOV10L1 (MOV10 like 1), interactúa con las proteínas *piwi* en células germinales masculinas de

ratón, y esta involucrada en el silenciamiento transcripcional de ciertas familias de EMs^{200; 201}.

Igualmente, la sobre-expresión de la exonucleasa Trex1, y de la trifosfohidrolasa SAMHD1, semejante a MOV10 se ha visto que pueden inhibir tanto la RTN del elemento LINE-1 como del IAP en cultivos celulares^{202; 203} aunque el mecanismo de acción aún falta por ser elucidado (figura 1-10). Así como la proteína SETDB1 el cual esta involucrada en el control transcripcional de los EMs independientemente de la metilación del ADN²⁰⁴.

Sorprendentemente, tanto las hESCs como las hiPSCs recapitulan algunos aspectos de los mecanismos conocidos que afectan, tanto la expresión como la RTN del L1, incluyendo el control por metilación del ADN, proteínas APOBEC y las proteínas Piwi^{109; 111; 113; 188; 205}.

OBJETIVOS

GENERAL

Determinar el impacto causado por los eventos de retrotransposición del elemento LINE-1 dentro de genes de genomas pluripotentes y diferenciados.

ESPECIFICOS:

1. Generación de líneas celulares clonales a partir de células pluripotentes con al menos una inserción de novo del retroelemento LINE-1
2. Determinar el locus de cada inserción y selección de aquellos con impacto en regiones codificantes.
3. Eliminación del casete reportero de las inserciones del L1 para cada una de las líneas transgénicas generadas.
4. Analizar el nivel de expresión de los genes impactados por el L1 en las células pluripotentes antes y después de perder la secuencia del casete reportero unido al retrotransposón.
5. Analizar el nivel de expresión de los genes impactados por el L1 luego del proceso de diferenciación celular

Capítulo 2: MATERIALES Y METODOS

Líneas celulares y condiciones de cultivo

Para el mantenimiento de las líneas celulares, se siguieron las condiciones estándares de cultivos, esto es a 37°C en un atmosfera húmeda de presión y con una concentración constante de CO₂ al 5%.

HeLa

Línea celular epitelial e inmortal extraída de un carcinoma cervical. Se caracteriza por ser altamente proliferativa y por tener una versión activa de la telomerasa durante la división celular. En su cariotipo se observa un número modal de 82 con un rango de 70 a 164.

Las células se cultivan en Medio de Cultivo Eagle Modificado de Dulbecco DMEM con alta concentración de glucosa (4.5g/L) (Gibco) sin piruvato; suplementado con 10% de suero bovino fetal SBF (en adelante medio completo) sin descomplementar (Gibco), y 1% de penicilina/estreptomicina (100X, Gibco).

Células de carcinoma embrionario humano (hEC) PA-1

Línea establecida a partir de células tomadas del líquido ascítico proveniente de un teratocarcinoma de ovario. Las células forman colonias muy unidas, y se diferencian para formar cuerpos embrioides cuando se cultivan a baja concentración de suero, bajas densidades de plaqueo ó cuando son tratadas con 5´BrdU.

Las células se cultivan en medio mínimo esencial MEM + Glutamax (Gibco), suplementado con 10% suero bovino fetal SBF (Gibco) (descomplementado por temperatura), 1% de aminoácidos esenciales (Gibco) y 1% de penicilina/estreptomicina (100X, Gibco).

Diferenciación celular

Las células PA-1 fueron crecidas durante 14 días en un medio que promueve la diferenciación (MEM suplementado con 10% de suero de reemplazo (Knockout Serum Replacement Medium; KOSR), 2mM L-Glutamina, 1% de penicilina/estreptomicina (100X, Gibco), y 0.1mM de aminoácidos no esenciales, suplementado con 1µM de ácido retinoico)²⁰⁶.

Subcultivos Celulares

Las condiciones de subcultivo en ambas líneas fueron semejantes bajo tratamiento con tripsina/EDTA (Gibco) a una concentración de 0.05% seguido de un bloqueo enzimático, para el cual se usó medio completo respectivo de cada línea celular.

En el caso de células diferenciadas una vez desprendidas las células, estas fueron sembradas en placas de 6 pozos recubiertas con una matriz de Matrigel, que consiste en una mezcla proteica gelatinosa que se asemeja al entorno extracelular de muchos tejidos dándole soporte a las nuevas células diferenciadas.

ANTIBIÓTICOS

Todos los antibióticos usados son llevados a una solución stock en agua destilada y pasados por un filtro (Millipore) de 0,2µm.

Ampicilina

La ampicilina (SIGMA) es un antibiótico betalactámico que inhibe la síntesis de la pared celular bacteriana mediante la inactivación de transpeptidasas en la superficie interna de la membrana celular bacteriana. Las bacterias resistentes escinden el anillo betalactámico de la ampicilina por la presencia de betalactamasa inhibiendo su acción. Posee un espectro de acción amplio, el cual incluye bacterias gram-negativas (como en el caso de la tetraciclina y el cloranfenicol), y *gram* positivas (similar a la bencilpenicilina).

Se prepara un stock de 100 mg/mL y conservado a -20°C. Tanto en medio líquido como en agar se usa una solución de trabajo de 100 µg/mL.

Kanamicina

La Kanamicina sulfato (Amresco) es un antibiótico aminoglicosido que inhibe la síntesis proteica en bacterias *gram*-negativas y *gram*-positivas. Actúa sobre la subunidad 30S ribosomal impidiendo la translocación provocando errores de codificación. Las bacterias resistentes aportan enzimas aminoglicosido-modificadas previenen la acción del antibiótico impidiendo su interacción con el ribosoma.

El stock es preparado a 50mg/mL y conservado a -20°C. Tanto en medio líquido como en agar se usa una solución de trabajo de 50 µg/mL.

G418

El G418 disulfato (Gibco) es un antibiótico aminoglicosídico similar en estructura a la gentamicina usado para selección y mantenimiento de células procariotas y eucariotas. Cuando las células son cultivadas en medio en presencia de G418 como agente de selección, es posible la obtención de colonias expresando este marcador de resistencia en un promedio de 10 a 14 días. Este antibiótico actúa bloqueando la síntesis peptídica y la elongación proteica mediante la inhibición de la síntesis en los ribosomas 70S y 80S.

Las concentraciones óptimas varían dependiendo del tipo celular, y para nuestro caso específico usamos concentraciones de 400 µg/mL en el caso de las células HeLa JVM, y de 200µg/mL en el caso de las líneas PA1. Para efectos prácticos se preparó un stock a una concentración de 50 mg/mL y almacenado en ausencia de luz y a 4°C.

PLASMIDOS

pRAM/LRE3

Es un plásmido que contiene una copia completa de un RC-L1 (LRE3) dirigido por un promotor UBC (del inglés ubiquitin C promoter). En la región 3' del L1 se ha clonado un casete reportero que codifica para la proteína neomicintransferasa (neo). Dicho casete además de tener un origen de replicación bacteriano (ColE1) contiene un promotor procariota (EM7), y una secuencia Shine-Dalgarno, modificaciones genéticas que median la traducción del casete en *E. coli*. La traducción en células procariotas de este casete modificado (neo) confiere resistencia al antibiótico bacteriano kanamicina (kan). En células eucariotas es conducido por el promotor SV40 que permite su traducción confiriendo resistencia al antibiótico G418. En ambos extremos del neo se encuentran dos secuencias dianas de la recombinasa Flp llamadas FRT (del inglés Flp Recombinase Target) con la misma orientación transcripcional.

pCEP4-eGFP

El pCEP4 (Invitrogen) es un vector de aproximadamente 10 Kb al que se le ha clonado de forma continua la proteína verde fluorescente humanizada derivada del phrGFP-C (Stratagene). Es usualmente usado como control de la transfección en la determinación de la **eficiencia de transfección** del plásmido de interés.

CMV-Flp

Plásmido de expresión también conocido como pOG-Flpe6 descrito en Buchholz et al; en 1998²⁰⁷. Recombinasa de sitio específico testada con éxito en células madre embrionarias.

MEDIOS DE CULTIVO BACTERIANO

Luria-Bertani (LB), caldo lisogénico.

El LB (SIGMA) líquido es preparado según las indicaciones de la casa comercial. En términos generales su composición para 1 litro es:

Triptona	10 g
NaCl	10 g
Extracto de levadura	5 g

La mezcla es autoclavada y enfriada antes de añadirle los respectivos antibióticos según sea el caso.

Luria-Bertani (LB), Agar

El LB (SIGMA) Agar es preparado según las indicaciones de la casa comercial. En términos generales su composición para 1 litro es:

Agar	15g
Triptona	10 g
NaCl	5 g
Extracto de levadura	5 g

La mezcla es autoclavada y enfriada antes de añadirle los respectivos antibióticos según sea el caso. Posteriormente se preparan las cajas de Petri adicionándoles 10 mL de medio y dejándolas solidificar.

Medio SOB (SuperOptimalBroth) con Glicina

Este medio se usa para crecer las cepas de bacterias competentes. En general está compuesto por los siguientes reactivos:

Triptona	20g
Extracto de levadura	5g
NaCl	0,5g
KCl 4M	625 uL
Glicina	10g

Una vez mezclados los reactivos el pH es ajustado a 7 y autoclavado. 10 mL de MgCl₂ (previamente autoclavado) es adicionado y conservado a 4°C.

Medio TB (TerrificBroth).

Medio nutricionalmente rico para el óptimo crecimiento bacteriano, usado en la preparación de las bacterias competentes. Para 100mL sus componentes son:

H ₂ O	Hasta 100mL
PIPES(10mM)	0.30g
CaCl ₂ *2H ₂ O (15mM)	0.22g
KCl(250mM)	1.86g

Una vez mezclado los reactivos se ajusta el pH a 6,7 y adicionan 1,09g de MnCl₂*4H₂O y posteriormente filtrado en vacío y almacenado a 4°C.

Bacterias competentes y Transformación Bacteriana

En la transformación bacteriana se usaron las cepas bacterias competentes de tipo DH5α y XL-GOLD

DH5α

Es la cepa de *E. coli* más usada para la transformación y amplificación de vectores, valida tanto para la electroporación como para la transformación por choque térmico. Posee una eficiencia de transformación de alrededor de 10⁶ufc/μg de ADN. Esta cepa fue obtenida del Banco Andaluz de Células Madre.

XL10-GOLD

XL-GOLD (Stratagene) es una cepa de *E. coli* optimizada para presentar una alta eficiencia de transformación alcanzando 10⁹ufc/μg de ADN. Igualmente permite la transformación de vectores de tamaños considerables (≈10 kb) por choque térmico. Además permite la replicación del ADN humano metilado.

TRANSFORMACIÓN

La transformación bacteriana es el proceso por el cual una bacteria es alterada genéticamente por la incorporación de un ADN exógeno. Brevemente, las bacterias competentes son descongeladas en hielo, y una vez en estado líquido se añade de 30-50μL a un vial que contenga aproximadamente de 50 a 100ng del plásmido de interés. Se incuba a 4°C durante 30 minutos. Posteriormente se procede a realizar un choque térmico a 42°C durante 45 segundos (en el caso de las DH5α). Los viales si incuban en hielo nuevamente durante 3 minutos. Luego se adiciona de 800 μL de medio LB (37°C). En el caso de estar seleccionando con kanamicina, las bacterias se dejan recuperar al menos 1 hora a 37°C, con el fin de que estas expresen los genes de resistencia. Finalmente las bacterias son sembradas en una placa de LB agar que contiene el antibiótico requerido y se procede a sembrar las bacterias por agotamiento usando un asa de siembra bacteriana. Las bacterias son incubadas a 37°C por un periodo de 12 a 16 horas, con el fin de tener colonias resistentes bien definidas.

AISLAMIENTO Y PURIFICACIÓN DE ADN

ADN plasmídico

Los plásmidos bacterianos, así como los generados a partir de ADN genómico, son replicados en la cepa bacteriana DH5 α para su posterior aislamiento y purificación.

Los plásmidos generados para las transfecciones fueron adquiridos usando un kit de purificación de plásmido súper enrollado midiprep (Quiagen) según indicaciones del fabricante a partir de 50 mL de cultivo bacteriano. El plásmido es purificado mediante flujo por gravedad a través de columnas basadas en intercambios aniónicos. Las concentraciones de plásmidos alcanzadas estuvieron en un rango de 250 a 1000 ng/ μ L.

Para las técnicas de análisis molecular, los plásmidos fueron aislados por el proceso de miniprep (Promega) a partir de 2 mL de cultivo bacteriano. Este proceso es más rápido y a menor escala que la midiprep. El aislamiento se realizó según las indicaciones del fabricante para la obtención de concentraciones de plásmido en un rango de 50 a 100 ng/ μ L.-

Aislamiento de ADN genómico

Según el número de células usadas y la cantidad de ADN requerido, el ADN genómico fue extraído de dos formas diferentes. Para pocas densidades celulares (en el caso de cultivos procedentes de placas de 96/pozos), se empleó el sistema de extracción celular QuickExtract (Epicentre) el cual es un sistema rápido y sencillo en el que se obtiene ADN genómico eficientemente en aproximadamente 10 minutos. En la figura 2-1 se esquematiza el uso del sistema QuickExtract.

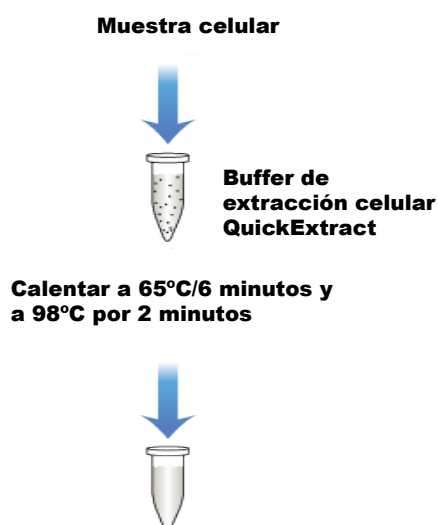


Figura 2-1. Aislamiento de ADN genómico. En el esquema se simplifica el proceso de extracción por medio de buffer QuickExtract (Epicentre).

Para densidades celulares mayores (es decir a partir de placas de 6/pozos en adelante), el procedimiento se realizó mediante el uso de un buffer de extracción de ADN genómico (200µg/mL ProtK; 10mM Tris pH=8.2; 10mM EDTA; 200mM NaCl; 0.5% SDS). Brevemente, las células son desprendidas y centrifugadas a 1200 rpm, luego son resuspendidas en un volumen (1V) de buffer de extracción de ADN genómico (300µL para un pozo de una placa de 6/pozos), e incubadas a 56°C/3h. Posteriormente se le agrega dos volúmenes (2V) de etanol absoluto para precipitar el ADN, seguido de un enfriamiento a -20°C “overnight”. Una vez precipitado el ADN, las muestras son centrifugadas y lavada con etanol al 70%. Finalmente es el etanol y la muestra se deja secar a 37°C; una vez seco el pellet, es resuspendido en 150 µL de buffer TE+RNasa pH= 8.2 e incubado a 37°C en agitación. Una vez disuelto el ADN se procede a su cuantificación.

Aislamiento ARN total

Para la extracción de ARN total se hizo por medio del método del Trizol® (Ambion), método por el cual se logra una lisis celular manteniendo la integridad del ARN. Brevemente, las células fueron plaqueadas en placas de 6 pozos hasta alcanzar un nivel de confluencia cercano al 90%. Por cada pozo se adicionó 1 mL de Trizol® y el lisado transferido a un vial del 1.5 mL, al que se le adicionará 0.2 mL de una mezcla de Cloroformo y Alcohol isoamílico (24:1) (Fluka). Posteriormente se centrifuga la muestra y se separa en un nuevo vial la fase acuosa de la fase orgánica y se le añade isopropanol (Sigma) en una relación 1:1 para la precipitación del ARN. Nuevamente se centrifuga la muestra seguida de una decantación. Al pellet formado se lava con 1 mL de Etanol (Merk) al 75% y nuevamente centrifugado. Finalmente se descarta el sobrenadante y se deja secar a temperatura ambiente. Una vez seco el ARN es diluido en de 20 a 30 µL de agua libre de RNasa y cuantificado posteriormente.

Cuantificación de los Ácidos Nucleicos

La cuantificación de los ácidos nucleicos se realizó mediante la medida de absorbancia usando un espectrofotómetro NanoDrop 2000c (ThermoScientific). Como principio básico, los ácidos nucleicos absorben a una longitud de onda de 260nm y de esta manera es posible cuantificar la cantidad de ADN en nanogramos por un microlitro (ng/µL).

Para la evaluación de la pureza se usó la relación de absorbancias a 260nm y 280nm (260/280). En el caso del ADN, una relación próxima a 1.8 es generalmente aceptada como “pura”, y 2 en el caso de ARN. Si el valor de la relación en ambos casos es menor, esto puede indicar la presencia de proteínas, fenol u otros tipos de contaminantes que absorben fuertemente en rangos cercanos a los 280nm.

TRANSFECCIÓN DE VECTORES

La transfección es el proceso por el cual una célula eucariota es alterada genéticamente por la incorporación de un ADN exógeno. En el presente estudio la incorporación de nuestros vectores se realizó por una vía no liposomal usando el reactivo FuGENE 6 (Promega), reactivo usado ampliamente para la inserciones celulares de ADN plasmídico y de ARN en cultivos celulares, con un alto nivel de eficiencia y baja toxicidad. Brevemente, 400µL de la mezcla Opti-MEM (Gibco) y FuGENE en una relación 97:3 e incubada 5 minutos a temperatura ambiente es adicionada a 4µg de ADN plasmídico (valores para un pozo de una placa de 6), después de 20 minutos la mezcla con el ADN es goteada sobre el cultivo celular. A las 24 horas post-transfección es reemplazado el medio de cultivo.

ENSAYO DE RETROTRANSPOSICIÓN

Para la generación de la inserciones del L1 nos basamos en el ensayo transiente de retrotransposición propuesto por Weiet al.²⁰⁸. Brevemente, el sistema usa un L1 (pRAM/LRE3) al que se le ha clonado un gen reportero su extremo 3' UTR³⁸. Este gen consiste en una copia antisentido de un marcador de selección (neo), un promotor heterólogo y una señal de poliadenilación. El gen de *neo* está interrumpido por un intrón transcripcionalmente en el mismo sentido al L1 de tal forma que las células resistentes a neomicina (G418) sólo provendrán cuando el transcrito iniciado a partir del promotor que dirige la expresión del L1 es procesado, reversotranscrito, y reintegrado en una región cromosómica, de esta manera el gen reportero puede ser expresado a partir de su propio promotor (Figura 2-2).

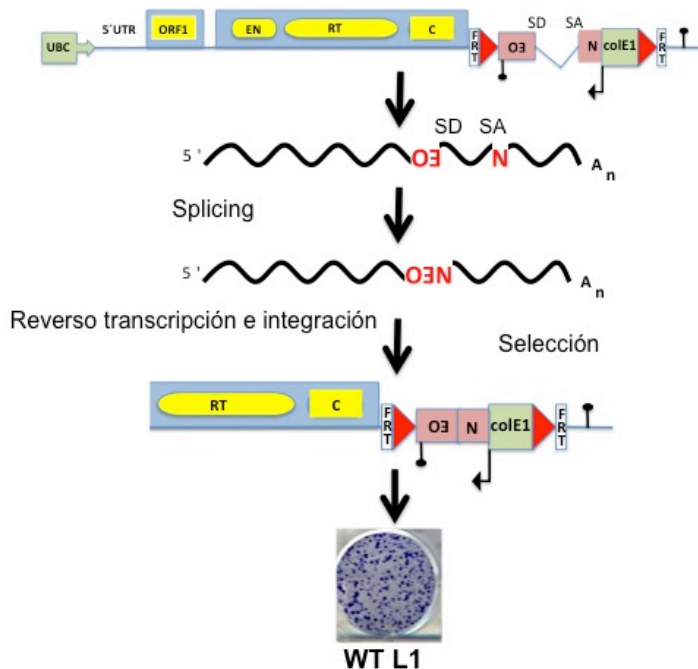


Figura 2-2. Ensayo de Retrotransposición. Un L1 activo marcado con una copia del gen reportero de resistencia a kanamicina (NEO) en sentido inverso al L1, está interrumpido por un

intrón ubicado en sentido transcripcional del L1. Los sitios donador y aceptor de splicing son indicados como DS y AS. Cuando el L1 es transcrito a partir de su promotor, el intrón presente puede ser procesado en la maduración del RNA del L1. Debido a la orientación del gen reportero solo podrá expresarse si este es reverso transcrito e integrado en el ADN genómico, de esta manera solo si se completa un ciclo de retrotransposición las células podrán expresar la proteína que confiere la resistencia al antibiótico (Adaptado de García-Pérez et al; 2007).

Brevemente, se siembran 8×10^4 células (conteo en cámara de Neubauer) en placas de 100 x 20mm y se incuban (en condiciones estándar de cultivo) entre 16 y 18 horas. Posteriormente las células son transfectadas con el L1 modificado, que en nuestro caso es el pRAM/LRE3. Al tercer día post-transfección se inicia la selección con G418 durante un periodo de 12 días, haciendo cambio de medios programados cada dos días. Una vez finalizado el proceso de selección, aquellas células en las cuales haya surgido al menos un evento de retrotransposición se convierten en unidades formadoras de colonias.

Una vez seleccionadas por antibiótico las células, pueden ser fijadas para ser teñidas y ser cuantificadas, pero también pueden ser aisladas y expandidas para la generación de líneas clonales con la presencia del transgen.

Fijación y Tinción Celular

A las placas con las colonias resistentes se les extrae el medio de cultivo y son aclaradas con 1mL de PBS 1X (Gibco). Seguido se adiciona 0,5mL de solución fijadora (Formaldehído 37%, Glutaraldehído 25%, PBS 10X) y se incuban a temperatura ambiente por 30 minutos. Luego las células fijadas son teñidas con cristal violeta (Merk) por un periodo de 15 a 30 minutos (final de la figura 2-2).

Aislamiento clonal (Colony Lifting)

Para el aislamiento de las colonias se realizó mediante "colony lift". Brevemente, a las placas con las células resistentes se les extrae el medio de cultivo, posteriormente las colonias son desprendidas por medio de un pequeño fragmento de papel de filtro (previamente autoclavado) humedecido con tripsina/EDTA al 0,05% (Gibco). Una vez desprendidas las colonias el papel es puesto en un nuevo pozo de una placa de 48/pozos con medio completo, e incubadas en condiciones estándares de cultivo.

Eficiencia de Transfección

Para determinar qué tan eficiente fue el proceso de transfección, es decir la proporción de células que pudieron ser transfectadas con éxito con el plásmido de interés, usando nuestro sistema de transfección, se realizó una co-transfección con pCEP-GFP para identificar por fluorescencia la cantidad de células que han sido transfectadas eficazmente. Brevemente, Al igual que el ensayo de retrotransposición, placas de 100 x 20 mm son plaqueadas con 8×10^4 células, después de las 16 horas las células son co-transfectadas con

2µg de pCEP-GFP y pRAM cada uno y puestas en cultivo. Al tercer día las células son desprendidas con tripsina/EDTA (Gibco) al 0,05% y bloqueada con una solución de PBS 1x suplementado con 10% SFB y 50mM de EDTA. Las células son centrifugadas y resuspendidas nuevamente en 300µL de PBS. Las células fluorescentes fueron cuantificadas y analizadas en el citómetro de flujo FACS Canto II (BectonDickinson, Laser 488nm y 633nm).

Expansión y congelación-descongelación celular.

Para la expansión de las células se usaron las condiciones estándar de cultivo celular. En el caso de colonias resistentes aisladas, dichos focos celulares son sembrados inicialmente en pozos individuales de una placa de 48/pozos. A medida que el cultivo alcanzaba niveles de confluencia superiores al 80%, cada una de las líneas eran subcultivadas para ser sembradas de nuevo en una placa de 24/pozos, esto con el fin de ir aumentando cada vez el área de superficie de adherencia en cada pase, hasta que puedan ser sembradas en un frasco T75.

Con el fin de preservar el material biológico generado, se congelaron cada una las líneas transgénicas expandidas. Brevemente, el proceso comienza con el mismo procedimiento para realizar un pase al cultivo, hasta el momento en el que se bloquea la actividad enzimática de la tripsina con medio suplementado con suero. Una vez las células en suspensión, son centrifugadas a 1000 rpm durante 5 minutos. Una vez formado el pellet celular, se retira el exceso de líquido, y al pellet se resuspende en 1mL de solución de crio-preservación constituida por SBF y DMSO al 10%. Las muestras luego son puestas en crioviales y almacenadas a -80°C con la ayuda de un contenedor de congelación "Mr. Frosty" (ThermoScientific). Después del congelamiento las células se transfieren a nitrógeno líquido para su preservación final.

Todo el material biológico congelado (líneas celulares transgénicas) es susceptible de ser descongelado y cultivado. Brevemente, los viales ya congelados son puestos en hielo mientras son sumergidos en un baño con agua a 37°C durante aproximadamente 3 minutos. Una vez descongelados las células son pasadas a un tubo Falcon de 15mL con 4mL de medio completo (el tipo de medio varía según la línea celular), con el fin de diluir la concentración de DMSO. Posteriormente se centrifugan las células a 1000 rpm por 5 minutos. Una vez peleteadas las células es descartado el sobrenadante y las células resuspendidas en 6mL de medio completo para ser sembradas en un frasco T25.

SISTEMA DE RESCATE DE INSERCIÓN (RECOVERY).

Para la identificación en el genoma de las inserciones del L1 se empleó la técnica de rescate en plásmidoRecovery²⁰⁹. Esta técnica usa el concepto de la PCR inversa aplicada a la técnica de transformación bacteriana. De esta forma se tiene una base genética para poder aislar inserciones de novo del LINE-1 de manera rápida, eficaz y fiable. Brevemente, 8µg de ADN genómico extraído de

las líneas transgénicas generadas con el pRAM, es digerido con enzimas de restricción que no corten el casete de neo, y que digieran fácilmente el genoma mamífero en fragmentos. Las enzimas de restricción que se usaron fueron la HindIII y SspI (figura 2-3), por ser enzimas las cuales sus secuencias de reconocimiento constan de 6 nucleótidos, hace que haya una probabilidad de un evento de corte una vez cada 4096pb. Una vez alcanzada la digestión completa del ADN genómico, se realiza un ligación intramolecular la cual es favorecida en un medio altamente diluido (figura 2-3). El nuevo plásmido formado a partir de ADN genómico es concentrado por centrifugación en filtro Amicon ultra 0,5mL (Merck Millipore) y almacenado a -20°C, en este momento el plásmido está listo para ser transformado en bacterias competentes de tipo XL-1 Gold (figura 2-3). Para la transformación, a 20µL del plásmido concentrado se le adicionan 0,5mL de bacterias competentes y son incubadas en hielo por 30 minutos. Posteriormente se procede a realizar un choque térmico a 42°C durante 38 segundos, y de nuevo regresadas al hielo por 2 minutos. Un mililitro (1mL) de medio LB (37°C) es adicionado a cada muestra y puestos en un tubo de 15mL (Falcon) a medio abrir y se dejan en agitación overnight. Finalmente las muestras son centrifugadas y sembradas tanto en placas de LB agar como LB liquido ambas con kanamicina. Las primeras se ponen a incubar a 37°C overnight, y las últimas en agitación a 37°C a 200 rpm.

La formación de una colonia bacteriana (figura 2-3), indica que una bacteria se ha transformado con el plásmido (ADN genómico) concentrado adquiriendo de esta forma la resistencia a la kanamicina. Para el aislamiento y purificación del plásmido, las colonias bacterianas son picadas y puestas a crecer en 2mL del medio LB con kanamicina overnight a 37°C y 200 rpm durante aproximadamente 12 horas; posteriormente el plásmido es aislado por medio del sistema miniprep (ver aislamiento y purificación de ADN plasmídico).

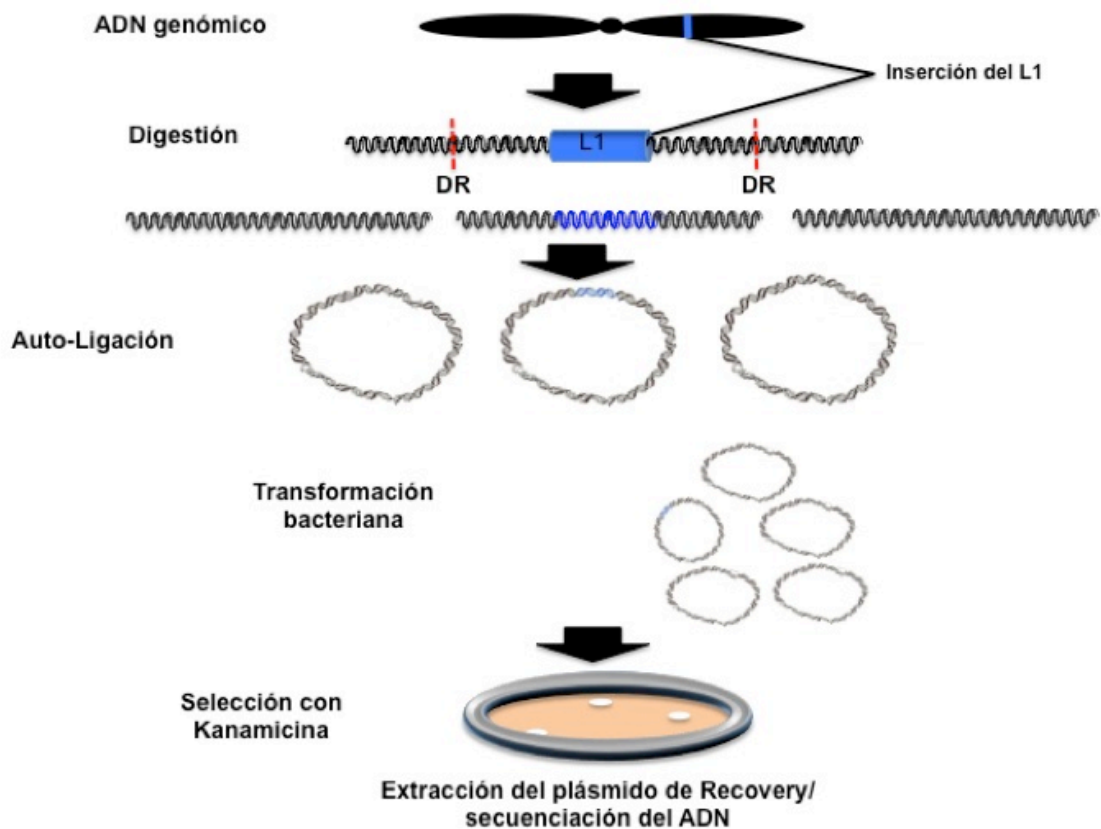


Figura 2-3. Sistema de rescate en plásmido RECOVERY. El ADN genómico es sometido a una digestión enzimática con enzimas de dianas de restricción conocidas (DR). El ADN fraccionado es ligado en medio diluido con el fin de promover la circularización (autoligación) de los fragmentos. El plásmido resultante es transformado en bacterias competentes para su amplificación y aislamiento.

Análisis de las Secuencias de Inserción

Para la determinación de las secuencias flanqueantes del L1 se usaron primers propios del L1/LRE3 (walkingprimers) y del casete de neo (ver en primers) y de esta manera identificar el sitio de pre-integración tomando como referencia el genoma humano de referencia hg38(BLAT, ver software usado).

Recombinación FLP

La recombinación Flp es un procedimiento genético usado para modificación genética celular. Se realiza a partir de una enzima FLP (flipasa) de 43kDa, que es obtenida del plásmido circular de *Saccharomyces cerevisiae*. Dicha enzima reconoce una secuencia de 34pb conocidas como FRT (del inglés *flipase recognition target*). En general esta enzima actúa haciendo un trabajo de “corte y pegado” siempre que se encuentren con estas secuencias específicas. Por este motivo es posible usar esta enzima a modo de tijera molecular con el fin de escindir de una manera controlada una secuencia deseada²¹⁰.

Según como estén presentes las secuencias FRT, es decir su localización y orientación dependerá el tipo de recombinación, por tanto es posible obtener:

Eliminación de un fragmento intermedio, cuando ambas secuencias se encuentran en la misma molécula lineal de ADN y con la misma orientación.

Inversión del segmento intermedio, si los sitios específicos de corte están en la misma molécula pero con orientación invertida.

Translocaciones, cuando las secuencias se encuentran en moléculas de ADN separadas es posible translocar los fragmentos de ADN.

Este sistema ha sido ensayado con éxito en células madre embrionarias (ES), de carcinoma embrionario (EC) y transgénicas de ratón^{211; 212}.

Como se mencionó anteriormente, el plásmido pRAM posee secuencias FRT en los extremos del gen reportero, esto permite que, de manera controlada se pueda escindir la secuencia “artificial” interna de cada inserción del L1 modificado. Brevemente, 3×10^5 células son co-transfectadas con los plásmidos cmv-FLP y el pCEP4-eGFP que nos permitirá sortear las células por fluorescencia (ver transfección de vectores). A las 72h pos-transfección las células son tripsinizadas seguido la inactivación con medio completo. Posteriormente se centrifugan, y el pellet resultante es resuspendido en un buffer de citometría (PBS 1X, SFB 10%, EDTA 50mM y ampicilina/estreptomicina 2%), luego las células son pasadas por un filtro de 70 μ m. Las células verdes son separadas por un separador celular BD FACSAria (BD Biosciences) y cultivadas en un pozo de poco diámetro con el fin de que se recuperen y puedan ser expandidas.

Dilución Clonal

El objetivo de esta técnica es la de diluir una población celular de tal manera que puedan crecer colonias aisladas derivadas a partir de células individuales en el aislamiento de variantes genéticas o fenotípicas²¹³. Brevemente, la población celular la cual presenta la variante genética buscada (células fluorescentes expandidas), son tripsinizadas y resuspendidas en medio completo. Inicialmente se hace un conteo celular seguido de diluciones seriadas hasta obtener una dilución con una concentración celular aproximada de 10×10^4 células/mL (figura 2-4). Posteriormente las células son plaqueadas con una densidad celular de 0,5 células/pozo en placas de 96/pozos; después de una semana son marcados aquellos pozos en los cuales ha crecido solo una célula. A las dos semanas aproximadamente se puede observar la caída del pH a causa del crecimiento celular (figura 2-4). Una vez que el cultivo celular ha alcanzado niveles de confluencia superiores al 80%, los pozos seleccionados son tripsinizados con el fin de proceder a la expansión clonal (ver Expansión y congelación-descongelación celular)

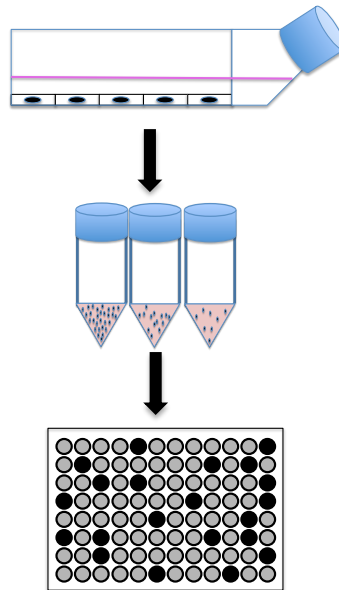


Figura 2-4. Aislamiento de líneas celulares clonales (Dilución clonal). Una vez resuspendidas las células se hacen diluciones seriadas hasta obtener una densidad celular del orden de 10×10^4 células/mL. Las células son sembradas en placas de 96 pozos, sembrando 1 célula por cada 2 pozos. Círculos en gris representan pozos sin crecimiento celular; círculos en negro representa la caída del pH a causa del crecimiento celular.

OLIGONUCLEÓTIDOS

A continuación se hará una relación de los oligonucleótidos sintetizados para llevar a cabo las diferentes técnicas moleculares

Oligonucleótidos ó Primers para Identificación del casete de Neo

Para la identificación inicial de las inserciones de las líneas transgénicas, en la determinación de una inserción verdadera a partir del pRAM, se usaron primers en los extremos del intrón del casete de neo, con el fin de poder analizar por PCR la diferencia de tamaños cuando éste se ha spliceado del ARN mensajero del pRAM en el evento de RTN (ver ensayo de retrotransposición).

Neo3Fw: ATCGACAAGACCGGCTTCCATC
 Neo3Rv: GCTATTCCAGAAGTAGTGAGGAG

Primers para Recovery

Para una precisa identificación de las inserciones del elemento LINE-1 es necesario secuenciar los plásmidos generados a partir de ADN genómico (ver sistema de rescate de inserción), proceso que usualmente conlleva varias rondas de secuenciación (dependiendo de naturaleza del plásmido generado y el tipo de inserción del L1). En general para el mapeo de la inserción se usaron dos parejas de oligonucleótidos según su ubicación.

La primera pareja de oligos, toma como referencia el casete de *neo* (en dirección sentido y antisentido) y la secuencia del L1 como anzuelo para el rescate de la inserción. Los primers orientados en dirección 5' del L1 son llamados *walking primers* debido a que permiten avanzar a través del L1 en busca del fragmento de ADN genómico ubicado en la región 5' del L1:

Reco3: primer orientado en el mismo sentido al L1, está ubicado en el origen de replicación bacteriano y está creado con el fin de identificar el ADN genómico contiguo a la cola de poli A (identificación del extremo 3').

Neo210As: como su nombre lo indica, está ubicado espacialmente en sentido contrario al L1 sobre el gen de resistencia de neomicina. El producto de secuenciación será en dirección 5' del L1.

Walking primers: son primers antisentido diseñados para "caminar" por toda la secuencia del L1 desde el extremo 3' del ORF2 hacia la región 5' URT.

RECO3	CTAAAGTATATATGAGTAACC
NEO210AS	GACCGCTTCCTCGTGCTTTACG
JORF2J_AS:	GGTGCGCTGCACCCACTAATG
JORF2I_AS:	CACATTTTCTTAATCCAGTC
JORF2H_AS:	CTAACTGGTGTGAGATGATATC
JORF2G_AS:	GAATTGATTTTTGTATAAGGTG
JORF2F_AS:	CATTTTCACGATATTGATTCTTCC
JORF2E_AS:	GTCCCATCAATACCTAATTTATTG
JORF2D_AS:	GAATTCGGCTGTGAATCCATCTGG
JORF2C_AS:	TTAATTGTGATGTTAGGGTGTC
ORF2B_AS:	GATTTGGGGTGGAGAGTTCTG
ORF2A_AS:	GCCTTCTTTGTCTCTTTTGAT
BGL2_AS:	CAGAGACTAGGATTGCAACCCCTGC
IG_AS:	CATTAGTTGATGCAGTTTCTTCC
ORF1C_AS:	CTTTCCATGTTTAGCGCTTCC
1598NEW_AS:	CCTGAATCTGAACGTTGGCCTGC
ORF1B_AS:	GGTCTTTTACATAGTCCCAT
ORF1A_AS:	CGTTCCTTTGGAGGAGGAGAGGC
JA8_AS:	GGTTTTATCTACTTTTGG
5'UTR_AS:	GCAGGCAGGCCTCCTTGAGCTG

La segunda pareja de primers están diseñados en ambas direcciones (sentido y antisentido con respecto al L1) siendo específicos de cada inserción. Fueron generados a partir del ADN genómico identificado con la primera pareja de primers. Los primers fueron diseñados manualmente y evaluados en programas informáticos (OligoCalc y Oligoevaluator) más adelante descritos.

Las secuencias del segundo set de primers fueron las siguientes.

PP6-1-C3FW	GCTAGCGCAAGCCAGCTCC
PP12S_C1AS	GATGTCCTCCTTTACAGC
PP12S_C1FW-1	GCTGCTACCTGAACACCCAG
NR2H1_CHR9AS	GTCACACTACTCGTGGGATC
NR2H1_CHR9AS1	CTCACTTCTCCCTGCATTTG
NR2H1_CHR9AS2	CTCTGGCCCTTGGATGTTGC
NR2H1_CHR9AS3	CTCTTAGTTACCTTGGAGGC
NR2H1_CHR9AS4	GCAGCCAAACTGCATAGAAC
NR2H1_CHR9AS5	GCCAGTTCAGTGGTCTTCAG
NR2H1_CHR9FW	GTGAGGTGAAAGGGAATGTGG
NR2H1_CHR9FW-2	GTGAGGTGAAAGGGAATGTG
PA1NR9S1_CHR11AS	GTCTCGCCAGCATCTGTGGT
NR9S1_CHR11FW	GCAGGAGAATCGCTTGAACC
NR9S1_CHR11FW-2	GGAGTTCGAGACCAGCATGG
NR13H4_CHR6AS	GTGATGGATTGGTGGCCGTG
NR13H4_CHR6AS1	GCGGCCGGGAACAGTGGCTC
NR13H4_CHR6AFW	GACCTGCGCCACCATGCC
NR13H4_C6FW-2	CAAAGTGCTGGGATTACAGG
NR25H1_CHR16AS	GGTTCACACCATTCTCCTGC
NR25H1_CHR16AS1	GCTATCAGTGCCTCAAGCATGTG
NR25H1_CHR16FW	CAGGGATTACTCAGATCC
NR25H1_C16FW-2	CACTTCAGCCTCCCTAGTAG
P-RAM 4/CHRM2A	GCTGTTGAGTGTTATGAGTACAG
P-RAM4/CHRM2B	ACTGTGCATACACTGGATG
P-RAM4/CHRM2	GTTTATTGCTGGTATGGGTGG
PRAM6 PCR INSERT 5'	CTGTCCTTCTGGTACTTGAAAGAG
PRAM6PCRINSERT 2	ACAGCCTGGACCATACTGAATGATG
PRAM 8/CHRM1A	AACAAAGAGCCATAAAGTGATC
PRAM8/CHRM1B	CTGTGCTTCTTTGGTGTGTTTGC
PRAM8/CHRM1C	GTGTCTACCTACAAATTAACCAT
P-RAM8/CHRM1D	GCAGTGACTTAAGCCCAGTC
P-RAM8/CHRM1E	CACATTGGTCTGAGTACATTAC
P-RAM8/CHRM1F	GTTCCCTGCTGCTATCTGACTG
P-RAM8/CHRM1G	GACGTTTATAGATGCTGCAC
P-RAM8/CHRM1H	CCTCCCGGCATTTCAATTCTG
PRAM8/CHRM1I	CAGGTTGTGGTTAGGCTTAGTG
PRAM8/CHRM1PCR	CATCTGGGGCTCTCTCTGTTCTC
P-RAM 15H/CHRM143'B	CTCAGCAGCAGTGAATGTTC
PRAM15 PCR INSERT	GTGTCCTGAAAGCCTGAGTGGCAG
P-RAM 33/CHRM 16	GTGTTGGGATTACAGGAGTGAG
P-RAM33/CHRM16B	TTGTGGTTTGGCCTGATTCTC
PRAM33/CHRM165'PCR	CATTTGCTGGCTCTGTTGCTTTC
P-RAM 36/CHRM15	GGCCATTCTAACAGGCATGAG
PRAM36/CHRM15PCR	AGGTGGAGGTTACAGTGAGCTG
PRAM41/CHRM4PCR	GATATTAGCTACATGCTTTCCAGCC
P-RAM 44/CHRM5 A	CTTGATAAGGTAGATTGCAGAAG

P-RAM44/CHRM5B	CATTTGCACAAGAACA
P-RAM44/CHRM5B	CATTTGCACAAGAACA
P-RAM44/CHRM 12	TTAGCTGAGTTTGGTGGTGC
PRAM44/CHRM12B	CTGGAGCACAGCATTGTCTTAG
P-RAM 44/CHRM12 PCR	CTATGAATGGGCCACTGCAC
PRAM44/CHRM12PCRB	GAGACCGGGTATTTGGGCTTAC
PRAM44/ORF2 5'	GTATTTCTGTGGGATCGGTGG
PRAM 45/CHRM9	TGCTCTGATATGAACTTAGG
PRAM45/CHRM9B	GGTATTAGGAGGTGATTTTGC
P-RAM45/CHRM9C	GATCTCCCTAGCCATGCTGAAC
PRAM45/CHRM9D	GCATATCAGAGGTCTTCATG
P-RAM45/CHRM9E	CATGGGAGCCTACTTCTTG
PRAM45/CHRM9F	GAGGAAAGTGAGCTATGCCTAC
P-RAM45/CHRM4	AGTTGGGCAAGAGACATCTTG
P-RAM45/CHRM4PCR	TAGACGGCTCTTTCCTCTGACTG

Primers para Identificación de la recombinación *Flp*

Para la identificación por PCR de la pérdida del casete de neo se usó para todas las muestras un “walking primer” en dirección sentido (JORF2H-S) y como primer antisentido uno ubicado en la región del promotor SV40.

JORF2H_S	GATATCATCTCACACCAGTTAG
SV40_REV2_FLP	TAAAGCAATAGCATCACAAATTCAC

Primers para RT-qPCR

Diseño de primers para la evaluación de la expresión de los genes impactados por el L1. La ubicación de las parejas de primers se hizo con respecto al sitio de inserción, por tanto, se diseñaron primers específicos para los exones que limitaban una inserción (regiones de intersección). Igualmente se diseñaron primers que alineaban en los exones ubicados aguas arriba (*upstream*) y aguas abajo (*downstream*) del sitio de inserción

La nomenclatura de los primers viene determinada de la siguiente forma: en primera posición aparece el nombre de la línea transgénica generada, seguido del nombre del gen evaluado especificando el exón de ubicación y terminando por la dirección del primer, ejemplo:

PP6CPNE4EX1FW1



Nombre	Nombre	Numero	Orientación
línea	Gen	Exón	
celular			

Las parejas de primers fueron diseñados principalmente en dos herramientas informáticas (Universal Probe Library Assay Design Center y el Primer-BLAST, descritas mas adelante) de forma independiente.

PP6CPNE4EX1FW1	AAAGGAGAGCTGGGAGGAGA
PP6CPNE4EX2RV1	GCAGCTCAACTTTGGTCAGG
PP6CPNE4EX1FW2	AAAAGGAGAGCTGGGAGGAG
PP6CPNE4EX2RV2	CTTGAGGATGACACAGGGGT
PP6CPNE4EX1FW2-2	CATGGCTGTAGAACCCAGCA
PP6CPNE4EX2RV2-2	CACGCAGCTCAACTTTGGTC
PP6CPNE4EX15FW1	CCTTCAGGAACTTCAAACACG
PP6CPNE4EX16RV1	ACTTGTTTTGGGACTTCAGC
PP6CPNE4EX15FW2	TCATCGTGGGAGTAGGGAAC
PP6CPNE4EX16RV2	CCACAACCTTGGTTTGGGACT
PP12DDAH1EX1FW1	GCAACTTTAGATGGCGGAGA
PP12DDAH1EX2RV1	CTTTTGGAAAGGCCACAA
PP12DDAH1EX1FW2	AATGCAACTTTAGATGGCGG
PP12DDAH1EX2RV2	CAGCCAAGATTCAGCACCT
NR2TRPM3EX1FW1	TGCTTTTGGGACCATTGAGT
NR2TRPM3EX2RV1	TGTATCAAAAGATACTCGCACATACA
NR2TRPM3EX1FW2	ATGCTTTTGGGACCATTGAG
NR2TRPM3EX2RV2	CTGCTTTGATGAGCCCTTTC
NR2TRPM3EX22FW1	GATACGGGTGACTTCAGAAAGG
NR2TRPM3EX23RV1	TGGAGTGCTCTCTCTCGTTG
NR2TRPM3EX22FW2	TGATGAGAGGATACGGGTGA
NR2TRPM3EX23RV2	GTCCACGGTCTGGAGTGAAG
NR2TRPM3EX24FW1	ACTTCAGAGAAAAGGATGATCGGT
NR2TRPM3EX25RV1	GTCCACGGTCTGGAGTGAAG
NR2TRPM3EX24FW2	GAAAAGGATGATCGGTTCAACTCA
NR2TRPM3EX25RV2	GACTTCTCCAGCCGCATAG
NR9ARHGAP42EX1FW1	ATTGGGGCGTTGAGGAAT
NR9ARHGAP42EX2RV1	AATACATTCAAACCTGGAAATCTTGC
NR9ARHGAP42EX1FW2	ATTGAGCTGGAGCGAACCAA
NR9ARHGAP42EX2RV2	AGACAGATTCCCAACGCC
NR9ARHGAP42EX21FW1	CCCAAAGCTTCACCAAACC
NR9ARHGAP42EX22RV1	GCTTTTGCTGCCACTACTGA
NR9ARHGAP42EX21FW2	GACCTGCCTCCGAAAATGTG
NR9ARHGAP42EX22RV2	TGAGCTTTTGCTGCCACTACT
PPNR10FGGYEX4FW1	GAAGGCAACAGGTGTCACAG
PPNR10FGGYEX5RV1	CAGCCTTTCTCTGCTGAATATG
PPNR10FGGYEX4FW2	TAAGGCGGGACATTTCTTTG
PPNR10FGGYEX5RV2	TTCCAAACCAATCATTTTCCA
PPNR10FGGYEX5FW2	TCCCGGACTTCTTATCGTGGA
PPNR10FGGYEX6RV2	GTCCACTTACACACCAGGGAG
PPNR10FGGYEX13FW1	CAAATGCATGCGGACATTAC
PPNR10FGGYEX14RV1	CAGCAGCACCCACAAGAAC
PPNR10FGGYEX13FW2	TTGGGGACTCGCTTCATTAT

PPNR10FGGYEX14RV2
NR13CASC15EX1FW1
NR13CASC15EX2RV1
NR13CASC15EX1FW2
NR13CASC15EX2RV2
NR13CASC15EX11FW1
NR13CASC15EX12RV1
NR13CASC15EX11FW2
NR13CASC15EX12RV2
PPNR25PMM2EX2FW1
PPNR25PMM2EX2RV1
PPNR25PMM2EX2FW2
PPNR25PMM2EX2RV2
PPNR25PMM2EX7FW1
PPNR25PMM2EX8RV1
PPNR25PMM2EX7FW2
PPNR25PMM2EX8RV2
DR4_OSBPL6EX3FW1
DR4_OSBPL6EX4RV1
DR4_OSBPL6EX3FW2
DR4_OSBPL6EX4RV2
DR4_OSBPL6EX9FW1
DR4_OSBPL6EX10RV1
DR4_OSBPL6EX9FW2
DR4_OSBPL6EX10RV2
DR6_FAM65BEX1FW1
DR6_FAM65BEX2RV1
DR6_FAM65BEX1FW2
DR6_FAM65BEX2RV2
DR6_FAM65BEX1FW2-2
DR6_FAM65BEX2RV2-2
DR6_FAM65BEX9FW1
DR6_FAM65BEX10RV1
DR6_FAM65BEX9FW2
DR6_FAM65BEX10RV2
DR8_LPHN2EX1FW1
DR8_LPHN2EX2RV1
DR8_LPHN2EX1FW2
DR8_LPHN2EX2RV2
DR8_LPHN2EX7FW
DR8_LPHN2EX8RV
DR8_LPHN2EX8FW
DR8_LPHN2EX9RV
DR15_RAD51B_EX1FW1
DR15_RAD51B_EX2RV1
DR15_RAD51B_EX1FW2

AGAACGGACTCCACCTCTTG
TCCCTTAAAAGTGCGACGA
TCATGATGCTTTCTAGATGCTGA
GGGGGAAATGTCCCTTAAAA
GAGAATTCCCAGCGATCAA
ACTCCAGCGGAGAAGTGTGT
TGAGGTTGCTGAAACAGTGG
CCAGAAGGCCAAAATCAAAA
GCAGGTAGGGGACTCCTTTC
GGACTTTGAGAAAGTGCAGGAG
AAGCCATTTTCTGGAAACACAT
CCAAAGAAATGGATGACTTCT
TGCTACCAAGCCATTTTCTG
CGACATGTGGAAAATGACGGT
GAGTAGCCCATGGTTCTGGG
AGGCCAGATCAGCTTTGATG
GGTTCTGGGGTCTGTGAAGA
CCAAAGGGACAGTAGGCAGA
GCTCGGTGCTAGAGGAAGC
GGGCATTTCCCCTGCTCATA
CTCGGTGCTAGAGGAAGCAG
TGCACAAGTGGCCAGAGTAA
AGTTCCACAAGGTTTGACTGG
CAGCCTGGTTACAGGACTCG
GTTCCACAAGGTTTGACTGGC
TGTCCTCTCCAAGATGCAGTT
GGGAGTCTGGTCCGGTAGTCC
TGTCCTCTCCAAGATGCAGT
TGATTTCCGGGAGTCTGGTC
CCTGGTCCGATGAAGAGGATGAT
TAATGATCCCATTGGGCCCTC
GTCTGGCTGGCTTTGCAC
CTGCCGGCCATACTTCAT
ACGCCTCTGTCTGGAGAT
CTTCTCCATCCCAGCTCTGC
AGTGACTGAAAGGACTCAATTTACTG
TTTGTGAAGTTTAGAGTAAGTGGTCTG
CAACTGCTGCCTGGCTTTAC
ACGTAGGCTCCATACTCTTCAG
ATGGTCGGACGGATGACAAG
AGGAAACACATCTGACCCAGT
TGGGTCAGATGTGTTTCTGA
TCCATGGCACTTGTATCCCTG
AAAGGGTGGGGAAACTTGAA
CCCCTCGTTTTAGTTTCTTGC
TCCGCGGGGAAACTGTGTAA

DR15_RAD51B_EX2RV2	GCTCTTGTGATAAACCCCACTCG
DR15_RAD51B_EX9FW1	AGGAAGGCCTGGTTCTTCA
DR15_RAD51B_EX10RV1	TGGTGGCACATCTGTAGTCC
DR15_RAD51B_EX9FW2	TTATTGCCAAGTCCCCTCTG
DR15_RAD51B_EX10RV2	TGGTGGCACATCTGTAGTCC
DR15_RAD51B_EX10FW1	ACACCATCAAGGAGGAAGGC
DR15_RAD51B_EX11RV1	TGGATTGAGAAACCAGATATGCAA
DR33_ANKRD26P1EX8FW1	CAGCTAATGAAGACTTTGATTTTGTAT
DR33_ANKRD26P1EX9RV1	CCATACCATTTTGTCTTTTTTCCA
DR33_ANKRD26P1EX8FW2	CGACTCACGACCTATGTCAGC
DR33_ANKRD26P1EX9RV2	GTGCTCTTGTGGAGCACTCT
DR33_ANKRD26P1EX8FW1-2	ACCATCTATTGACGACTCACGAC
DR33_ANKRD26P1EX9RV1-2	TTTGCTGGCTCTGTTGCTTT
DR33_ANKRD26P1EX12FW1	AGGCCTCAAAGAGGAAATCAG
DR33_ANKRD26P1EX13RV1	TGCATTACATGTAGACCTTCTG
DR33_ANKRD26P1EX12FW2	CCACTAGAGGCTACACCATGT
DR33_ANKRD26P1EX13RV2	GCAGTTTCCACTGCTTCTGTAAG
DR33_ANKRD26P1EX9FW1	GCAATCACATTTGGACAGCAAG
DR33_ANKRD26P1EX10RV1	GCTCGTTGTTTTCACTTTCTTCA
DR36_THSD4EX3FW1	AAGTCACTGGGGACAGAAGG
DR36_THSD4EX4RV1	GGGCCTTACCATAGCCATACT
DR36_THSD4EX3FW2	CCTCGAAGTCACTGGGGACAG
DR36_THSD4EX4RV2	ACCATAGCCATACTTGCCAGG
DR36_THSD15EX3FW1	TGTTGGACCCCTCTGAATGT
DR36_THSD4EX16RV1	CTGGCAGCTCTTGGAACAC
DR36_THSD15EX3FW2	GCCAAATGGTTTAGCACCGA
DR36_THSD4EX16RV2	TGTCATCAGACAGACACCGC
DR41_MAPK10EX1FW1	AACTTCAGAATCAAACTATGACCCTA
DR41_MAPK10EX2RV1	GCTGTCCTAGTGGTTTGAACG
DR41_MAPK10EX1FW2	ACGACCCCCACACTTCTAAC
DR41_MAPK10EX2RV2	GGCAAATGCTGTCCTAGTGGT
DR41_MAPK10EX1FW1-2	TAGGCCAAGTTGAAAATGCAGC
DR41_MAPK10EX2RV1-2	TAATGTGGCAAATGCTGTCCTA
DR41_MAPK10EX5FW1	CGCTACCAGAATCTAAAGCCTATT
DR41_MAPK10EX6RV1	TCAAGGACAGCATCATACGC
DR41_MAPK10EX5FW2	ACAGTTCTCAAGCGCTACCA
DR41_MAPK10EX6RV2	AAGGACAGCATCATACGCGG
DR44_LOC100287944EX1FW1	GCTCGATAGCCCCTGCTT
DR44_LOC100287944EX2RV1	GCACATTCTCAGCCTCAGTTC
DR44_LOC100287944EX1FW2	CCTCAGGACCCGCTCGATAG
DR44_LOC100287944EX2RV2	CAGCCTCAGTTCTTGCTTTTCT
DR44_LOC100287944EX1FW1-2	GCCCGGACGATGTAGAGGG
DR44_LOC100287944EX2RV1-2	TGCATAGTTACCACCACCACG
DR44_LOC100287944EX1FW2-2	CCTCAGGACCCGCTCGATAG
DR44_LOC100287944EX2RV2-2	CAGCCTCAGTTCTTGCTTTTCT
DR44_LOC100287944EX3FW1	TGTGCTACTTAATCCACACAACG

DR44_ LOC100287944EX4RV1	CCTTCCCATACTTCAGTTTCTTCA
DR44_ LOC100287944EX3FW2	GGGAGGAAGAACTGTACCT
DR44_ LOC100287944EX4RV2	GTGGCCTTGGACAAGTTGCT
DR44_ LOC100287944EX2FW1	GTTATGATGAGGGGACGCAGG
DR44_ LOC100287944EX3RV1	AGGTGACAGTTCTTCCTCCCT
DR44_ LOC100287944EX2FW2	ATGTTATGATGAGGGGACGC
DR44_ LOC100287944EX3RV2	AGTTCTTCCTCCCTGGATCG
DR45_ FRAS1EX1FW2	TCCTCAAAGTGTGGCTCGG
DR45_ FRAS1EX2RV2	GCCAACAAGGAATCCTGATAGA
DR45_ FRAS1EX3FW1	TTGCAGAAACCCTCAATGTG
DR45_ FRAS1EX4RV1	CATTGGTTGGCAGCTATTTG
DR45_ FRAS1EX3FW2	GCAGAAACCCTCAATGTGCC
DR45_ FRAS1EX4RV2	ACACTCAGGACAGCATTGGT

REACCIÓN EN CADENA DE LA POLIMERASA (PCR)

Empleamos la técnica de la reacción en cadena de la polimerasa para la amplificación de fragmentos de ADN en la identificación (por presencia o ausencia) y caracterización de secuencias. A continuación se mencionan los diferentes usos aplicados a esta técnica.

PCR NEO

Es usada en la identificación de casete reportero *neo*. Por medio de esta técnica se puede identificar un evento de RTN verdadero por pedida del intrón presente en el casete de resistencia, así la diferencia de tamaños del amplicón determinará la presencia del L1/LRE3 en el genoma de analizado.

Para la generación de los ampliaciones se usó una ADN *Taq* polimerasa recombinante (Invitrogen) junto con los primers Neo3Fw y el Neo3Rv, las cantidades usadas fueron determinadas según indicaciones de la casa comercial. El ciclo de amplificación se llevó a cabo en un termociclador *Mastercycler ep gradient S* (Eppendorf) con los siguientes parámetros:

	Temperatura °C	Tiempo	Ciclos
Inicio	95	2 min	40
Desnaturalización	94	20 seg	
Alineamiento	59	20 seg	
Elongación	72	2 min	
Elongación final	72	10 min	

PCR FLP

Es usada para la identificación de la escisión del casete reportero *neo* por acción de una *Flipasa*. Debido a la diferencia en tamaños de un estado positivo

a un negativo, además de usar la *Taq* polimerasa convencional, se usó el sistema de expansión de molde extenso “*Expand Long Template PCR System*” (Roche). Este último contiene una mezcla de una ADN *Taq* polimerasa termoestable y una ADN polimerasa con actividad correctora de errores termoestable. Mezcla que facilita la obtención de productos de PCR de larga extensión con un alto rendimiento. Las cantidades usadas fueron determinadas según indicaciones de la casa comercial y los primers usados fueron: el JORF2H_S y el SV40_REV2_FLP.

El ciclo de amplificación se llevó a cabo en el termociclador *Mastercycler ep gradient S* (Eppendorf) con los siguientes parámetros:

	Temperatura °C	Tiempo	Ciclos
Inicio	96	1 min	40
Desnaturalización	94	15 seg	
Alineamiento	54	30 seg	
Elongación	68	10 min	
Elongación final	68	10 min	

PCR Para la identificación del extremo 5' del L1.

En la identificación del extremo 5' de la inserción del L1, se usó el sistema de expansión de molde extenso mencionado anteriormente. Cada una de las PCRs es individual para cada una de las líneas ya que cada línea posee una inserción en diferentes sitios del genoma, por lo que cada uno de los primers que van en el mismo sentido de la inserción está diseñado con base en el ADN genómico huésped de cada inserción. Las cantidades usadas fueron las indicadas por la casa comercial.

El ciclo de amplificación se llevó a cabo en el termociclador *Mastercycler ep gradient S* (Eppendorf) con los siguientes parámetros:

	Temperatura °C	Tiempo	Ciclos
Inicio	96	1 min	40
Desnaturalización	94	15 seg	
Alineamiento	Variable	30 seg	
Elongación	68	10 min	
Elongación final	68	10 min	

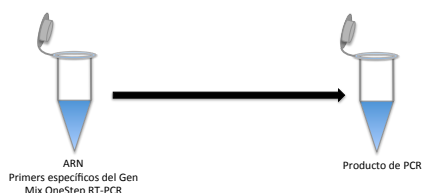
Análisis de la expresión génica RT-qPCR

Para el análisis de la expresión de los genes se empleó el sistema GoTaq 1-step RT-qPCR (Promega). El método consiste en un sistema de reactivos para el análisis cuantitativo de ARN, usando un protocolo de una PCR cuantitativa

en un solo paso de reversotranscripción. Las cantidades aportadas al sistema se hicieron bajo las indicaciones del fabricante. En general, el uso de esta técnica permite:

1. Una representación precisa del número de copias de la secuencia diana.
2. Reducir la posibilidad de errores y de contaminación debido a un menor número de pipeteo.
3. Tener la mejor opción de análisis para exploraciones de alto rendimiento en números reducidos de ensayos.

Los primers usados fueron exclusivos para cada gen evaluado y alineaban en exones consecutivo de cada gen, reduciendo el riesgo de falsos positivos debido a la imposibilidad de generarse un amplicón de larga extensión. Lo anterior es solventado debido a la gran longitud de la mayoría de los intrones presentes en los genes, además del corto tiempo de extensión dado, lo que dificulta la posibilidad en la amplificación largas secuencias.



El ciclo de amplificación se llevó a cabo en el termociclador *StepOnePlus Real-Time PCR System* (AppliedBiosystems) con los siguientes parámetros:

	Temperatura °C	Tiempo	Ciclos
Generación ADNc	37	30min	1
Inactivación de la RT	95	10	
Desnaturalización	95	15 seg	40
Alineamiento/ELONGACION	60	1 min	
Elongación final	72	10 min	

Estadística

Para el análisis de significancia estadística se realizó la prueba *t* de dos colas para muestras relacionadas con varianzas iguales.

HERRAMIENTAS INFORMÁTICAS

Serial Cloner

Este es un software libre que permite el alineamiento de dos secuencias mediante blast. Se usó en el alineamiento de las secuencias “rescatadas” con la secuencia del pRAM/LRE3. Las secuencias que no alineaban se analizaban en otro software para identificar su localización genómica (Blat).

BLAT

Es una aplicación bio-informática desarrollada por la universidad de California en Santa Cruz (UCSC) en su portal en línea “*Genome Bioinformatics*” (<http://genome.ucsc.edu>). Permite determinar la localización genómica de la secuencia introducida en el genoma humano.

GeneCard

Es una base de datos de genes humanos que provee e integra información genómica, proteómica, transcriptómica, genética y funcional de todos los genes conocidos. Fue usada en la caracterización de los genes impactados por el L1.

The Human Protein Atlas

Es una base de datos que provee el conocimiento aportado por el programa *The Human Protein Atlas*, el cual es un programa de investigación que tiene como fin el de explorar todo el proteoma humano basados en la detección por anticuerpos. La información aportada proviene de ensayos en 44 tejidos normales diferentes de humanos y 20 tipos de cáncer diferentes, así como 46 líneas celulares humanas diferentes. Esta base de datos fue usada para la caracterización de las proteínas codificadas por los genes que han sido objeto de un evento de RTN del L1.

Primer-BLAST

Es una herramienta informática de libre acceso del NCBI (National Center for Biotechnology Information) para el diseño y búsqueda de primers específicos. Esta aplicación integra una herramienta de diseño de primer y la búsqueda en BLAST en una sola aplicación. Esta herramienta fue usada en el diseño específico de los primers usados para el análisis de expresión (RT-qPCR).

Universal Probe Library Assay Design Center

Herramienta informática de la casa comercial ROCHE para el diseño de los ensayos de RT-qPCR, incluyendo la sondas y primers específicos de genes. El

software muestra las secuencias específicas de la diana de los cebadores además que es capaz de identificar y separar las regiones intrónicas de las exónicas.

Esta herramienta junto con el primer-BLAT fueron las usadas en el diseño de los primers empleados en las evaluaciones de la expresión génica de las regiones codificantes impactadas por un evento de RTN del L1.

OligoCalc y Oligoevaluator

Son dos aplicaciones informáticas diferentes el cual aportan parámetros físicos necesarios a la hora del diseño de primers, tales como %G-C, temperatura de melting, auto-complementariedad entre otros. Estas herramientas fueron usadas en el diseño manual de primers.

OligoCalc es de acceso gratuito y se puede acceder en la siguiente dirección electrónica: <http://www.basic.northwestern.edu/biotools/oligocalc.html>.

Oligoevaluator es una aplicación proporcionada por SIGMA-ALDRICH, que funciona por suscripción gratuita, y se puede encontrar en la siguiente dirección electrónica: <http://www.oligoevaluator.com/OligoEvaluator/Login.jsp>.

Capítulo 3: RESULTADOS

EVALUACIÓN DEL VECTOR PRAM/LRE3

Ensayo de RTN en células HeLa

Para determinar la eficiencia de RTN del nuestro plásmido, se realizó inicialmente un ensayo de RTN en células HeLa (materiales y métodos) en el que se transfectaron las células con el pRAM-LRE3, a diferentes densidades celulares (2×10^5 , 2×10^4 y 2×10^3 células).

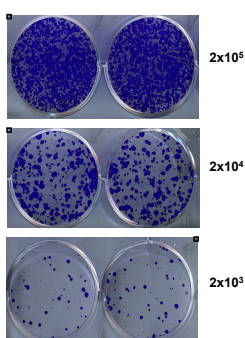


Figura 3-1. Ensayo de retrotransposición en células HeLa implementando el sistema pRAM/LRE3

Una vez seleccionadas las células en kanamicina, las células fueron fijadas y teñidas como se muestra en la figura 3-1. Claramente pudimos observar que el pRAM/LRE3 era capaz de movilizarse eficientemente en este tipo celular independientemente de la densidad celular al que fue transfectado, datos que soportan observaciones previas en donde se evalúa la eficiencia de retrotransposición del pLRE3/mneol/ColE1 en HeLa mostrando igualmente una alta eficiencia de RTN^{1; 2}.

Capacidad generadora de colonias transgénicas

Con el fin identificar la capacidad del pRAM/LRE3 para generar clones celulares resistentes a G418 debido a la realización de un evento de RTN. Nuevamente partimos de un ensayo convencional de RTN, en donde al finalizar la selección con el antibiótico tendremos focos celulares resistentes a G418, indicando de esta manera que cada una de estas colonias provienen de un evento de retrotransposición único sucedido en una célula individual.

Los análisis moleculares llevado a cabo en un *pool* celular indicaron que mediante el uso del sistema pRAM se obtenían inserciones del elemento L1 con eventos de retrotransposición canónicos de la vía endonucleasa (EN) dependientes.

GENERACIÓN DE LÍNEAS CLONALES EN PA1

Debido a la semejanza en el perfil transcripcional de las células del carcinoma embrionario PA-1 con las células madre embrionarias (hESCs)³, se usaron como modelo celular del desarrollo humano temprano. No obstante, estudios previos en donde se evaluó la eficiencia de RTN del LRE3 en este tipo celular se observó que su eficiencia disminuía notablemente con respecto a células HeLa, debido a que aproximadamente el 90% de las inserciones producidas eran silenciadas². Con el fin de asegurarnos de que nuestras células tenían el mismo comportamiento que el descrito anteriormente, decidimos usar los mismos vectores descritos por García-Pérez et al (2010), sobre nuestras células PA-1 para determinar su comportamiento cuando estas son tratadas con TSA (inhibidor de la histona deacetilasa) (figura 3-2).

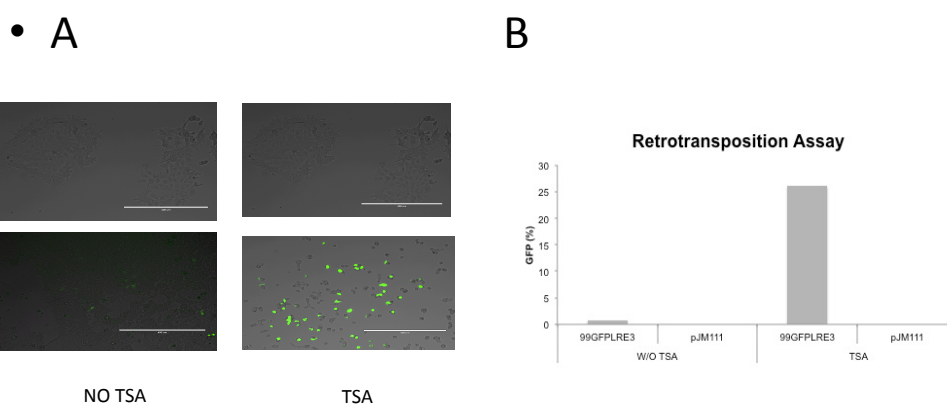


Figura 3-2. Identificación del silenciamiento de las inserciones en células PA-1

Como se observa en la figura 3-2, nuestras células PA-1 se comportan semejante a lo reportado por García-Pérez, confirmando que la mayor proporción de eventos de RTN son silenciados al igual que se ha visto en algunos experimentos realizados con células progenitoras neurales⁴.

En la producción de las líneas transgénicas realizamos los ensayos de RTN de igual forma que en células HeLa. Una vez obtenidas las colonias celulares producto de la selección con G418 se procedió al aislamiento de cada uno de los focos de selección. Teniendo en cuenta que cada colonia proviene de un evento de retrotransposición de una célula individual, nos garantiza que cada una de las inserciones identificadas sean clonales. Las colonias generadas fueron susceptibles de ser aisladas y expandidas generando de esta manera líneas celulares transgénicas con al menos un evento de retrotransposición en su genoma.

En tres experimentos de RTN diferentes se pudieron aislar aproximadamente 30 líneas celulares clonales. El uso del sistema pRAM nos permitió recuperar los eventos de RTN de cada una de las líneas celulares, y de esta manera nos permitió la caracterización molecular a nivel de nucleótido de las inserciones rescatadas. Los nombres de las líneas celulares varían según del experimento del cual proceden, por lo que tendremos líneas celulares pRAMPA1, pRAMPA1-NR y pRAMPA1-D

Identificación de las inserciones PCR-NEO

Inicialmente la determinación de los eventos de RTN se realizó mediante PCR de *neo*, en donde podremos determinar por peso molecular del amplicón la pérdida por *splicing* del intrón presente en el casete reportero (figura 3-3A).

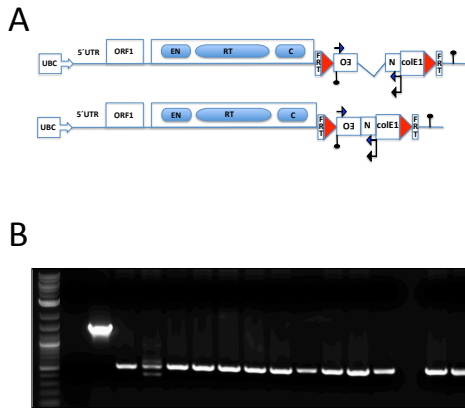


Figura 3-3. Identificación de las inserciones por medio de la pcr de neo (PCR-NEO)

Es importante determinar que en cada una de las líneas analizadas ya no este presente el plásmido como tal, de ahí de que para cada una de las líneas transgénicas generadas solo se observa el amplicón correspondiente a la secuencia del gen reportero activo (figura 3-3B).

Las líneas celulares que inicialmente fueron seleccionadas con G418 y que posteriormente resultaron ser positivas para la PCR de *neo* fueron seleccionadas como material biológico para la realización del presente trabajo.

Mapeo de las inserciones

En la identificación de la ubicación genómica de las inserciones *de novo* del LINE-1 de cada una de las líneas celulares generadas se usó el sistema de rescate en plásmido *Recovery* (ver materiales y métodos).

Los plásmidos recuperados pertenecientes a los subclones bacterianos aislados de cada línea celular clonal transgénica inicialmente son digeridos con la enzima de restricción *EcoRI* con el fin de determinar el tamaño del plásmido y tener un patrón de banda específico para cada plásmido tal como se muestra en la figura 3-4. Una vez identificados los patrones se inicia el análisis molecular de las inserciones. Es importante identificar la presencia de diferentes plásmidos por cada muestra; en los casos en el que los patrones de bandas resultaron diferentes, el análisis e identificación se hace en paralelo para descartar de que sean inserciones diferentes.

Los tamaños de los plásmidos rescatados variaron desde aproximadamente 3 kb hasta 10 kb de longitud (figura 3-4).

En la identificación de extremo 3' del L1 se secuenciaron los plásmidos (obtenidos de ADN genómico) con el oligo denominado Reco3_fw que está ubicado al finalizar el casete de *neo* en la misma dirección transcripcional del L1 (figura 3-5). De esta forma podemos identificar la ubicación exacta de la inserción, teniendo en cuenta que el inicio del TPRT se da a partir del extremo 3' de ARN mensajero del L1. Hay que tener presente que al estar identificando el extremo 3' supone igualmente la identificación de la cola de poli A típica de las inserciones del L1, lo que a la hora de la secuenciación supone un problema, debido a que la secuenciación consecutiva de adeninas hace que la polimerasa "patine" impidiendo de esta manera la obtención completa de la secuencia y por ende de la porción del ADN genómico colindante a la inserción del L1. En aquellos casos en el que la cola de poli A impedía la correcta secuenciación del fragmento, solo quedaba "caminar" por el plásmido en dirección 5'.

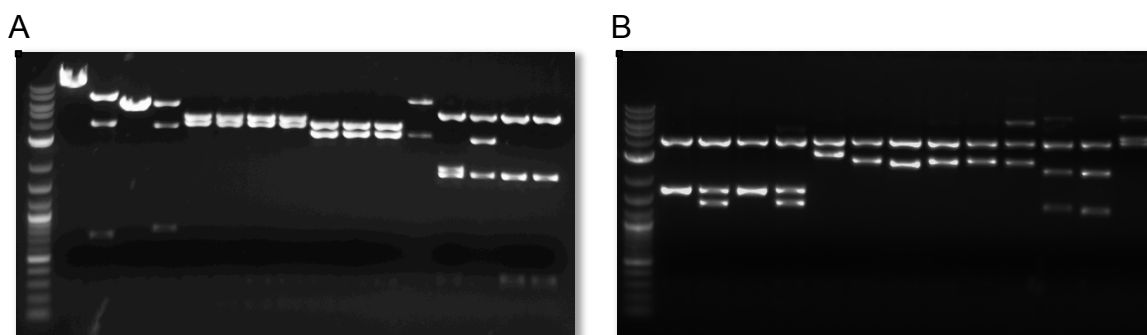


Figura 3-4. Digestión de los plásmidos del recovery con endonucleasas de restricción A) enzima de restricción HindIII. B) Enzima de restricción SspI.

Para la identificación del extremo 5', la secuenciación del plásmido de *recovery* empezó en primera instancia con el oligo Neo210-AS (figura 3-5), el cual está ubicado espacialmente al inicio del casete de *neo* orientado hacia la porción 5' del L1 es decir transcripcionalmente en dirección opuesta al L1, esto nos permitió secuenciar la fracción del L1 en dirección 5', tomando como referencia el casete de resistencia *neo*. Debido a la extensión de LINE-1 (6 Kb), es necesario realizar varias rondas de secuenciamientos con el fin de alcanzar el extremo 5' de la inserción y por tanto la fracción de ADN genómico colindante. Para esto usamos primers antisentido (walking primers) que abarcan espacialmente la totalidad del L1 y que nos sirven para caminar sobre él en cada ronda de secuenciación (figura 3-5).

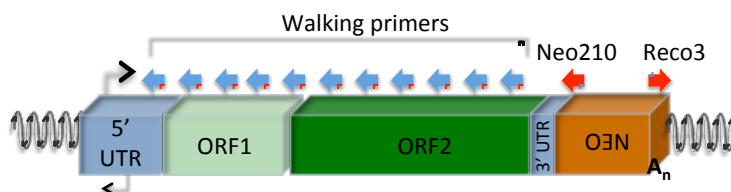


Figura 3-5. Distribución espacial de los oligos de secuenciación en el elemento LINE-1

Tabla 3-1. Lista de las 16 inserciones recuperadas.

Nombre de la línea	Línea Celular	Cromosoma	Posición Cromosómica	Gen	Truncación	TSD	poly A
pRAMPA1_6	PA-1	3	131941027	CPNE4	SI, 5700	SI, 11	46
pRAMPA1_10	PA-1	20	2024578	NO	n/i	n/i	68
pRAMPA1_12	PA-1	1	85496004	DDAH1	SI, 22	SI, 17	77
pRAMPA1NR_1	PA-1	17	22159287	NO	SI, 7	SI, 13	36
pRAMPA1NR_2	PA-1	9	70856401	TRPM3	SI, 10	SI, 14	112
pRAMPA1NR_3	PA-1	4	151310266	NO	n/i	n/i	86
pRAMPA1NR_4	PA-1	6	134895191	NO	n/i	n/i	89
pRAMPA1NR_5	PA-1	7	20144268	NO	SI, 11	16	128
pRAMPA1NR_6	PA-1	16	51716774	NO	n/i	n/i	52
pRAMPA1NR_8	PA-1	4	157237259	NO	n/i	n/i	114
pRAMPA1NR_9	PA-1	11	100909342	ARHGAP42	n/i	n/i	62
pRAMPA1NR_10	PA-1	1	59506338	FGGY	SI, 5530	BLUNT	37
pRAMPA1NR_11	PA-1	1	224963829	NO	SI, 3863	n/i	60
pRAMPA1NR_12	PA-1	12	n/i	NO	n/i	n/i	69
pRAMPA1NR_13	PA-1	6	21699077	CASC15 (ncRNA)	n/i	n/i	67
pRAMPA1NR_14	PA-1	X	n/i	NO	n/i	n/i	57
pRAMPA1NR_17	PA-1	3	181650746	NO	n/i	n/i	32
pRAMPA1NR_25	PA-1	16	8819372	PMM2	n/i	n/i	76
pRAMPA1_D4	PA-1	2	179,182,050	OSPLB6	SI, 5070	SI,13 bp	89
pRAMPA1_D6	PA-1	6	24,930,532	FAM65B	SI, 4231	No	39
pRAMPA1_D8	PA-1	1	81,896,934	LPNH2	SI, 29	SI,14	76
pRAMPA1_D11	PA-1	2	124,238,045	NO	SI, 6027	SI,9	73
pRAMPA1_D15	PA-1	14	68,972,745	RAD51B	SI, 5738	SI,3	39
pRAMPA1_D26	PA-1	5	52,847,392	NO	SI, 3745	No	74
pRAMPA1_D27	PA-1	n/i	13,241,828	NO	SI, 8	SI,12	54
pRAMPA1_D29	PA-1	n/i	136,923,374	NO	SI, 9	SI,14	71
pRAMPA1_D31	PA-1	3	104,792,201	NO	SI, 5818	SI,15	61
pRAMPA1_D33	PA-1	16	46,552,318	ANKRD26P1	SI, 2389	SI,15	45
pRAMPA1_D36	PA-1	15	71,564,736	THSD4	SI, 4879	No	34
pRAMPA1_D37	PA-1	11	75,043,230	ARRB1	SI, 6016	SI,9	44
pRAMPA1_D41	PA-1	4	87,417,392	MAPK10	SI, 9	SI,17	61
pRAMPA1_D44	PA-1	12	107,159,920	LOC100287944	SI, ?	SI,11	33
pRAMPA1_D45	PA-1	4	79,074,813	FRAS1	SI, 29	SI,17	40

Distribución de las inserciones

Teniendo en cuenta los datos de los sitios de inserción generados por los eventos de RTN del L1 (tabla 3-1), se determinó que dichos eventos fueron esparcidos a través del todo el genoma, encontrándose ampliamente distribuidas en el. Aproximadamente un 40% de las inserciones impactaron unidades génicas (conjunto de exones e intrones), sin embargo estas inserciones en su totalidad solo impactaron secuencias intrónicas (tabla 3-1). El 60% restante fueron distribuidas en el aproximadamente 75% de ADN genómico restante⁵.

Dos inserciones no pudieron ser mapeadas por ninguno de los extremos, debido a que fueron inserciones quiméricas en zonas de ADN repetitivo unido a la longitud de la cola de poli A que no permitió su identificación (tabla 3-1).

En adelante el trabajo se centro solo en aquellas líneas celulares que poseían en su genoma inserciones del L1 que estuvieran ubicada estrictamente en unidades génicas (tabla 3-1).

Sitios de corte de la endonucleasa (EN)

La mayoría de las inserciones descritas hasta el momento por medio de la via TPRT cortan en secuencias consenso degeneradas (5'-TTTT/A-3') y variantes de estas secuencias que están dadas en forma general por Pi- Pi- Pi- Pi/ Pu (Pi=base Pirimidínica (T-G), Pu=base púrica (A-T))⁶.

Solo aproximadamente el 40% de los eventos se insertaron en secuencias consenso de reconocimiento de la endonucleasa del L1(identificados en Gilber et al; 2005) (Tabla 3-2). Interesantemente alrededor de 54% presentaban una T (Pi) en la quinta posición (en el sitio de corte) en vez de una base purica (Pu) predictiva de la secuencia consenso (Tabla 3-2). En 5/13 eventos (38%) el corte endonucleolítico ocurrió entre el enlace fosfodiéster C/T, y el 15% se produjo en sitios G/Ty A/T como se muestra en la Tabla 3-2.

En general los sitios de corte identificados en el análisis parecen ser los característicos que son comunes a las inserciones producidas por la EN en el TPRT.

Tabla 3-2. Sitios de corte de la Endonucleasa. Los nucleótidos resaltados no concuerdan con la secuencia degenerada consenso de sitio de corte de la EN del L1 (3'-A/TTTT-5').

MUESTRA	Sitio de Corte
pRAMPA1_6	TTAC/T
pRAMPA1_12	TTTC/A
pRAMPA1NR_2	TATT/G
pRAMPA1NR_9	CTTG/T
pRAMPA1NR_10	TTTC/T
pRAMPA1NR_13	CTTC/T
pRAMPA1NR_25	TCTC/T
pRAMPA1_D4	TTTC/A
pRAMPA1_D6	CTTC/T
pRAMPA1_D8	TCTT/A
pRAMPA1_D15	TCCA/T
pRAMPA1_D41	TCTG/A
pRAMPA1_D45	TTTT/A

TSD y Cola de poli A.

La gran mayoría de nuestras inserciones poseían una gran extensión de la cola de poli A en su extremo 3', hecho que dificultó la caracterización de este extremo por secuenciación *Sanger* con el Reco3. La presencia de esta región homopolímera hace que la polimerasa patine sobre la hebra molde, provocando la formación de un patrón de secuenciación en forma de olas repetitivas que hace imposible su lectura.

Cuando la secuenciación directa del extremo 3' fue de difícil acceso, se identificó esta región por dos métodos diferentes: i) secuenciar el plásmido en dirección 5' partiendo desde el L1 hasta alcanzar la región 3'.ii) con base en la información obtenida del ADN genómico en la región 5' (Genome Browser) de aquellos plásmidos que no fueron cortados por la diana de restricción (HindIII y SspI), se diseñaron oligonucleótidos sintéticos aproximadamente de 200 a 300 nucleótidos aguas abajo del sitio de inserción. Estos primers 3'antisentido fueron luego usados para la secuenciación antisentido del plásmido. La totalidad de los eventos de inserción presentaron longitudes mayores de 30 As por lo que la mayoría de las identificaciones del extremo 3' se realizaron por los métodos indirectos acabados de mencionar.

En el análisis del extremo 5' de los plásmidos que contenían las inserciones del L1, se identificó que la mayoría estaban cortados en las dianas de restricción (correspondientes a las enzimas que fueron usadas en la digestión del ADN) presentes en el L1. En el caso de la enzima HindIII corta en el nucleótido 3662 y en el caso de SspI el nucleótido 2330 del L1. Para estos casos, el extremo 5' del L1 fue identificado mediante PCR (expand long template). Usando la secuencia hallada en el extremo 3' como se mencionó anteriormente, se diseñaron oligonucleótidos sintéticos con base en el genoma humano de referencia (Genome Browser) que alineaban espacialmente en aproximadamente de 200 a 300 nucleótidos aguas arriba del sitio de inserción en dirección transcripcional 5'-3'. Estos primers sintéticos luego fueron usados junto con el primer Neo210As para la amplificación por PCR desde el ADN genómico con posterior secuenciación del amplicón.

Se pudo identificar el extremo 5' a un 70% (9/13) de los eventos, de estos el 88% presentaron TSD y un evento se caracterizó por ser una inserción roma (BLUNT). En la tabla 3-1 se puede observar que 4 de las inserciones estaban truncadas en los primeros 30 nucleótidos lo que representa un tamaño considerable casi llegando a ser un L1-RC.

Definitivamente no se pudo identificar la región 5' del 30% restante debido a la no amplificación de la región buscada con los primers, por tanto fue imposible la determinación de las duplicaciones del sitio diana TSD.

Mapeo intragénico de las inserciones

Los loci de los eventos de inserción que ocurrieron en unidades génicas fueron mapeados a nivel genómico e intragénico. A continuación se detallaran cada

una de inserciones impactadas en las diferentes unidades génicas encontradas en cada una de las líneas celulares transgénicas generadas

pRAMPA1_6-CPNE4

Esta línea celular fue generada con una inserción en el LINE-1 en el gen CPNE4 (Copine IV). Este gen pertenece a la familia copina altamente conservada, codifica para una proteína de unión fosfolipídica a membrana dependiente de calcio, el cual puede regular eventos moleculares en la interface de la membrana celular y el citoplasma, tales como trafico a través de la membrana, mitogenesis y el desarrollo.

Con respecto a detecciones de la proteína por tinción se ha observado alta expresión membranosa luminal en intestino delgado y túbulos renales. Además de actividad citoplasmática de un subconjunto de células linfoides. En tanto que en epitelio escamoso, testículos así como en células del musculo liso se ha observado expresiones moderadas.

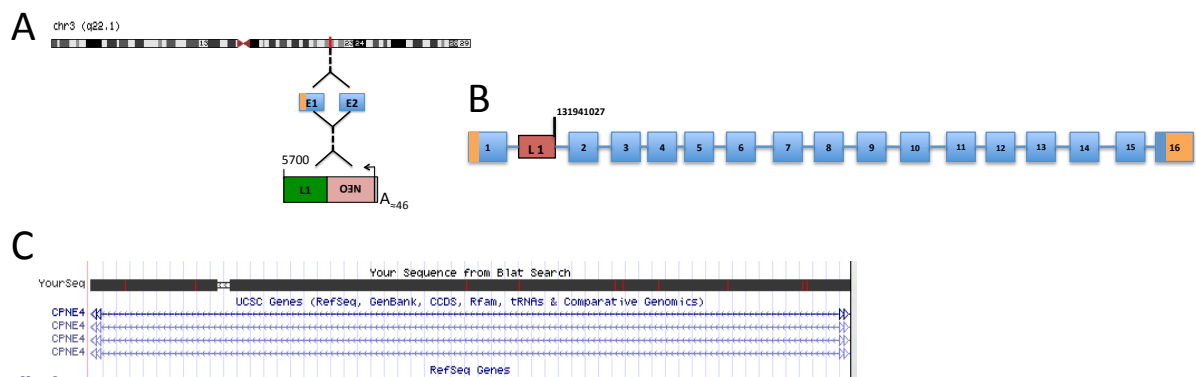


Figura 3-6. Mapeo cromosómico y molecular de la inserción del L1 en el gen CPNE4. A) Ubicación cromosómica de la inserción y su ubicación frente al gen, en donde se muestra el grado de truncación y la extensión de la cola de poli A. B) Esquema general del gen CPNE4 en el que se identifica la inserción del L1 (rojo) con el locus cromosómico en el extremo 3' y su orientación con respecto al gen, los exones (azul) y secuencias UTR (naranja). C) Isoformas identificadas en la base de datos (Genome Browser).

Múltiples variantes de la proteína resultan por splicing alternativo (UCSC Genome Browser) (figura 3-6 C). El gen está ubicado en la cadena antisentido del cromosoma 3 (3q22.1) y está formado principalmente por 16 exones incluyendo las regiones UTRs (figura 3-6 B).

Una vez caracterizado el sitio de inserción del L1 se determinó que este tuvo lugar en entre los exones 1 y 2 (locus 131941027) del gen y en la misma dirección transcripcional (Figura 3-6 A).

pRAMPA1_12-DDAH

Línea celular impactada por el LINE-1 en un gen de la familia de genes dimetil-arginina dimetil-aminohidrolasa (DDAH). Codifica para una enzima que esta involucrada en la generación de oxido nítrico (NO) regulando la concentración celular de metilargininas que a su vez inhiben la actividad de la oxido nítrico sintasa. Se ha observado que en la mayoría de tejidos normales hay de una moderada a fuerte expresión detectada por tinción citoplasmática. En tanto que en epitelio escamoso así como en tejido muscular no hay evidencia de expresión por tinción de tejido. En líneas celulares tumorales se ha observado que esta ampliamente expresada con excepción de líneas celulares de origen linfóide. Entre las líneas con alta expresión a partir de ARN se encuentran: la línea celular de neuroblastoma metastásico (SH-SY5Y), línea celular de carcinoma de célula pequeña de pulmón (SCLC-21H), línea celular de adenocarcinoma de colon (CACO-2), células endoteliales microvasculares con telomerasa inmortalizada (TIME) entre otras.

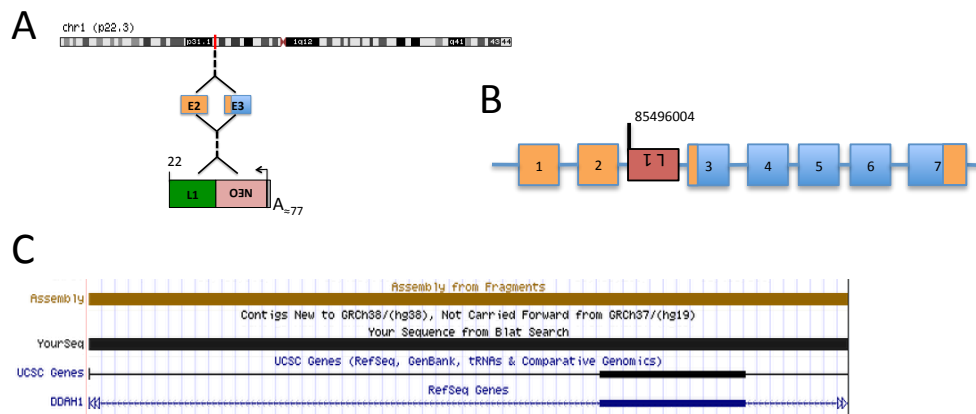


Figura 3-7. Mapeo cromosómico y molecular de la inserción del L1 en el gen DDAH. A) Ubicación cromosómica de la inserción y su ubicación frente al gen, en donde se muestra el grado de truncación y la extensión de la cola de poli A. B) Esquema general del gen DDAH en el que se identifica la inserción del L1 (rojo) con el locus cromosómico en el extremo 3' y su orientación con respecto al gen, los exones (azul) y secuencias UTR (naranja). C) Isoformas identificadas en la base de datos (Genome Browser).

La secuencia codificante esta ubicada en la cadena antisentido del cromosoma 1 (1p22.3) y esta formada principalmente por 7 exones incluyendo las regiones UTRs (figura 3-7 A y B). Solo se reporta una sola isoforma para este gen según la base de datos consultada (UCSC Genome Browser) (figura 3-7 C)

En el análisis de sitio de inserción se determinó que el L1 impactó el gen en el intrón 2 (locus 85496004) justo entre los exones 1 y 2 UTR. La inserción esta ubicada en dirección transcripcional opuesta al gen.

pRAMPA1NR_2-TRPM3

Esta línea celular fue impactada por el LINE-1 en un gen que pertenece a la familia de canales receptores de potencial transitorio (TRP). Los canales TRP son canales de selección catiónica importantes para la ruta de señalización del calcio y la homeostasis. La proteína codificada por el gen TRPM3 media la entrada del calcio y esta entrada es potenciada por el agotamiento de las reservas de calcio. Existen evidencias de su expresión de una manera marcada en tejidos de riñón con respecto a los demás tejidos evaluados que poseen una expresión escasa o nula, comportamiento que es observable igualmente en líneas celulares in vitro.

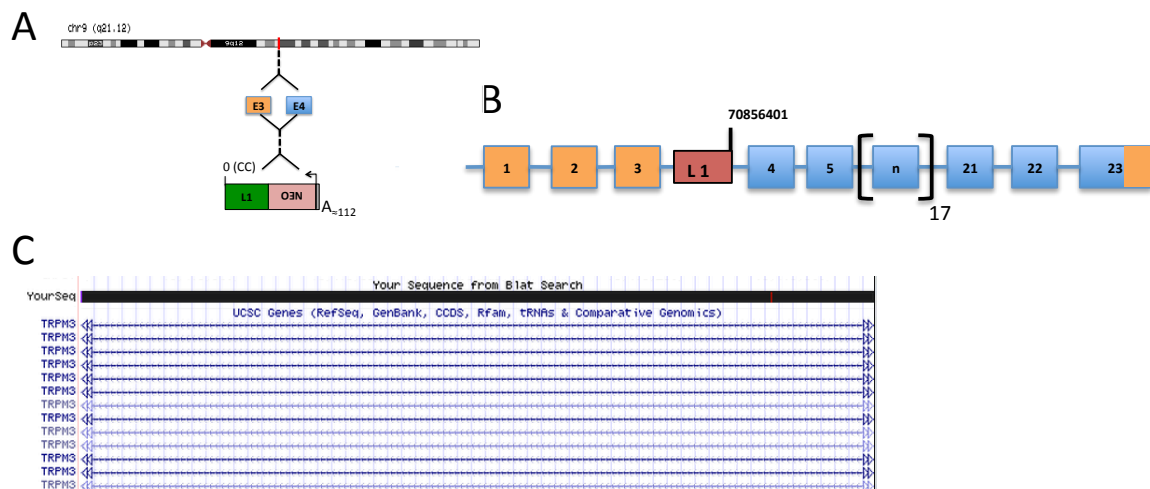


Figura 3-8. Mapeo cromosómico y molecular de la inserción del L1 en el gen TRPM3. A) Ubicación cromosómica de la inserción y su ubicación frente al gen, en donde se muestra el grado de truncación y la extensión de la cola de poli A. B) Esquema general del gen TRPM3 en el que se identifica la inserción del L1 (rojo) con el locus cromosómico en el extremo 3' y su orientación con respecto al gen, los exones (azul) y secuencias UTR (naranja). C) Isoformas identificadas en la base de datos (Genome Browser).

Alternativamente han sido identificado 13 variantes de los transcritos spliceados que codifican para diferentes isoformas (figura 3-8 C). El gen está ubicado en la cadena antisentido del cromosoma 9 (9q21.12), la región codificante del gen la constituyen 26 exones incluyendo las regiones UTRs (figura 3-8 A y B).

Se pudo determinar que la inserción del L1 se realizó específicamente entre los exones 3 y 4 (con respecto al genoma de referencia) del gen (locus 70856401) en la misma dirección transcripcional (figura 3-8 A y B).

pRAMPA1NR_9- ARHGAP42

La inserción presente en esta línea celular tomó lugar dentro de la secuencia del gen ARHGAP42 (del inglés (Rho GTPase Activating Protein 42), que consiste en una proteína que está relacionada con la ruta de señalamiento del citoesqueleto. Anotaciones del GO (Gene Ontology Consortium) con respecto a este gen revelan actividad de unión a fosfolípidos, además de actividad

activadora de GTPasa. Observaciones de su expresión proteica, revelaron que una gran cantidad de tejidos poseen, una débil o moderada actividad citoplasmática, a excepción del tracto gastrointestinal, las glándulas adrenales, los túbulos renales, los ductos seminíferos y el musculo cardiaco que exhibieron una fuerte inmunoreactividad con un patrón granular. Tanto en hígado como en los tejidos linfoides en general no hay evidencias de actividad de la proteína.

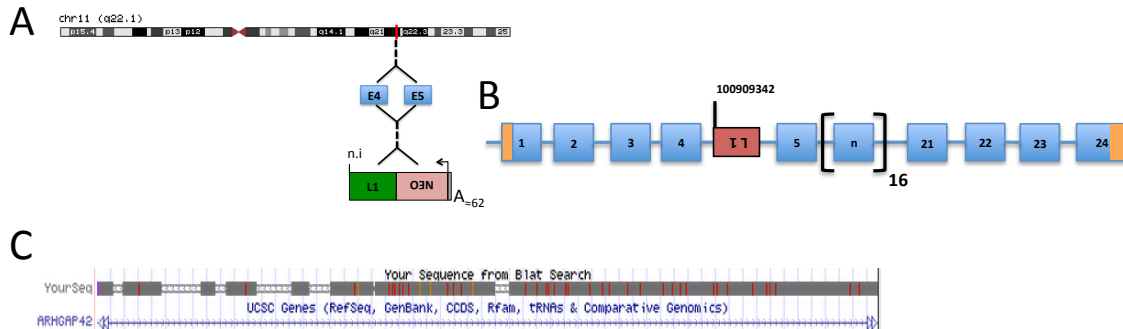


Figura 3-9. Mapeo cromosómico y molecular de la inserción del L1 en el gen ARHGAP42. A) Ubicación cromosómica de la inserción y su ubicación frente al gen, en donde se muestra el grado de truncación y la extensión de la cola de poli A. B) Esquema general del gen ARHGAP42 en el que se identifica la inserción del L1 (rojo) con el locus cromosómico en el extremo 3' y su orientación con respecto al gen, los exones (azul) y secuencias UTR (naranja). C) Isoformas identificadas en la base de datos (Genome Browser).

El gen está ubicado en la cadena con sentido del cromosoma 11 (11q22.1) y está constituido por 24 exones incluyendo las regiones UTRs (figura 3-9 A y B), identificándose una isoforma en la base de datos (figura 3-9 C). Análisis del sitio de inserción determinaron que la inserción del L1 tuvo lugar en el intrón 4 entre los exones 4 y 5 del gen (locus 100909342) en dirección transcripcional opuesta al gen (figura 3-9 B).

pRAMPA1NR_10- FGGY

Esta línea celular ha sido generada con una inserción de un LINE-1 dentro del gen FGGY. Este gen codifica una proteína que fosforila carbohidratos como la ribulosa, ribitol y el L-arabinol. Análisis de GWAS han encontrado en algunas poblaciones una asociación entre polimorfismos en este gen y la esclerosis lateral amiotrofica esporádica, aunque no ha sido posible replicar esta relación en otras poblaciones. Igualmente anotaciones del GO relacionadas con este gen incluyen actividades quinasa y fosfotransferasa así como aceptor de grupos hidroxilo.

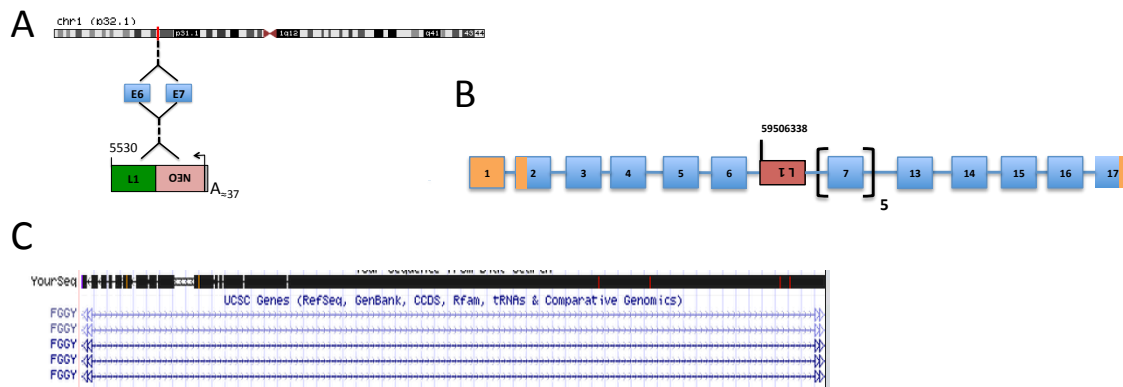


Figura 3-10. Mapeo cromosómico y molecular de la inserción del L1 en el gen FGGY. A) Ubicación cromosómica de la inserción y su ubicación frente al gen, en donde se muestra el grado de truncación y la extensión de la cola de poli A. B) Esquema general del gen FGGY en el que se identifica la inserción del L1 (rojo) con el locus cromosómico en el extremo 3' y su orientación con respecto al gen, los exones (azul) y secuencias UTR (naranja). C) Isoformas identificadas en la base de datos (Genome Browser).

Análisis de expresión en 81 tejidos normales solo en 2 mostraron niveles medios o altos de expresión como el caso del hígado. Datos que concuerdan con las observaciones hechas en líneas celulares, en donde en general se obtienen bajos niveles de expresión de la proteína codificada, a excepción de dos líneas que presentan niveles medios o altos, como es el caso de la línea celular de linfoma de Hodgkin (HDLM-2).

Splicing alternativo de este gen resulta en la formación de múltiples variantes de la transcripción (5 isoformas) (figura 3-10 C). El gen está ubicado en la cadena con sentido del cromosoma 1 (1p32.1) y está constituido por 17 exones incluyendo las regiones UTRs (figura 3-10 A y B).

Análisis de sitio de inserción determinaron que la RTN del L1 en este gen tuvo lugar en el intrón 6 entre los exones 6 y 7 (locus 59506338) del genoma de referencia, ubicándose en dirección transcripcional opuesta al gen (figura 3-10 A y B).

pRAMPA1NR_13- CASC15

El evento de RTN del L1 presente en esta línea celular tuvo lugar en una región que codifica para un gen de ARN. Este ARN largo no codificante lncARN se le ha denominado CASC15 (del inglés cáncer susceptibility candidate 15). El ARN antisentido es transcrito a partir de la cadena con sentido del ADN en vez de la molde, por tanto su secuencia es complementaria al ARN mensajero. Bajos niveles de expresión de la isoforma corta de CASC15 está altamente asociada con neuroblastomas avanzados y poca supervivencia⁷. Igualmente se ha visto que está relacionada con la progresión del melanoma⁸.

Este gen de ARN está ubicado en la cadena con sentido del cromosoma 6 (6p22.3) y está constituido por 12 exones (figura 3-11 A y B).

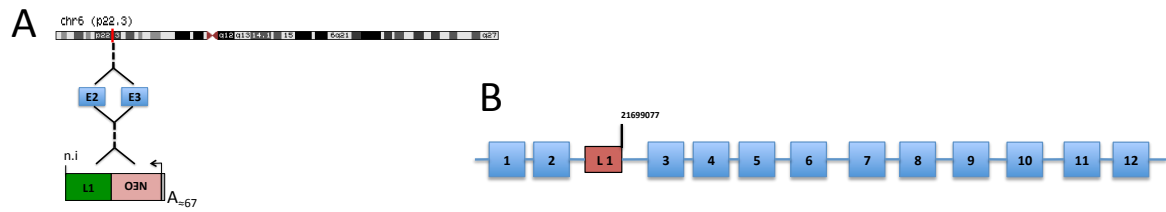


Figura 3-11. Mapeo cromosómico y molecular de la inserción del L1 en el gen de ARN CASC15. A) Ubicación cromosómica de la inserción y su ubicación frente al gen, en donde se muestra el grado de truncación y la extensión de la cola de poli A. B) Esquema general del gen CASC15 en el que se identifica la inserción del L1 (rojo) con el locus cromosómico en el extremo 3' y su orientación con respecto al gen, además el conjunto de exones (azul).

Se pudo identificar que el sitio de inserción del L1 en este gen ocurrió en el intrón 2 entre los exones 2 y 3 del gen (locus 21699077). Tanto el gen como la inserción del LINE-1 están ubicadas en la cadena consentido del cromosoma 6 (figura 3-11 A y B)..

p RAMPA1NR_25-PMM2

Línea celular a partir de la inserción del elemento LINE-1 en el gen fosfomanomutasa 2 (PMM2). La proteína codificada por este gen cataliza la isomerización de la manos 6-fostato a manos 1-fostato, el cual es un precursor de la manosa GDP necesaria para la síntesis de oligosacáridos dolicolfofostato. Mutaciones en este gen se ha visto que causan defectos en la biosíntesis de la glicoproteínas manifestándose como síndrome de glicoproteínas carbohidrato deficientes del tipo I. Dentro de las enfermedades asociadas a este gen se incluyen el desorden congénito de glicosilación y ruptura miocárdial. Análisis en 80 tejidos normales, 39 de ellos presentaron una expresión de la proteína de moderada a alta. En la mayoría de las células de tejido como en líneas celulares se ha identificado actividad citoplasmática aunque en algunos casos se pudo identificar en el núcleo. Igualmente el nivel de expresión en líneas celulares fue semejante al observado en los tejidos evaluados.

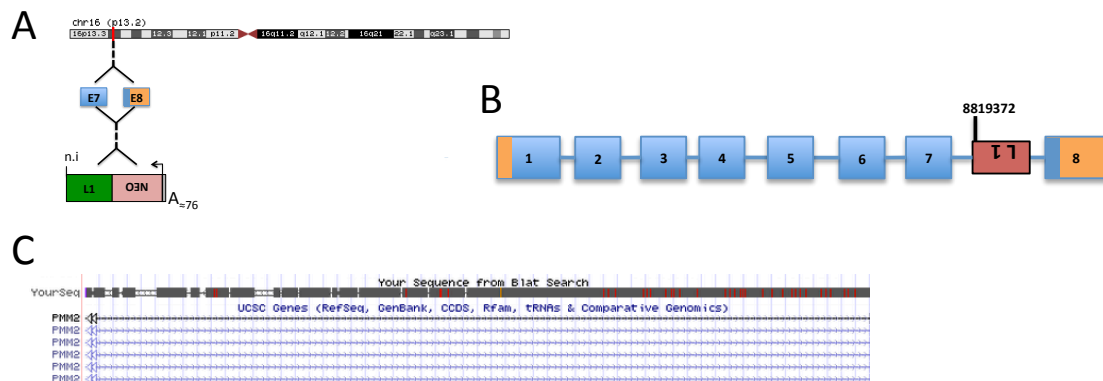


Figura 3-12. Mapeo cromosómico y molecular de la inserción del L1 en el gen PMM2. A) Ubicación cromosómica de la inserción y su ubicación frente al gen, en donde se muestra el

grado de truncación y la extensión de la cola de poli A. B) Esquema general del gen PMM2 en el que se identifica la inserción del L1 (rojo) con el locus cromosómico en el extremo 3' y su orientación con respecto al gen, los exones (azul) y secuencias UTR (naranja). C) Isoformas identificadas en la base de datos (Genome Browser).

El gen está ubicado en la cadena consentido del cromosoma 16 (16p13.2) y esta constituido por 8 exones, incluyendo las regiones UTRs (figura 3-12 A y B). Se identificaron 6 isoformas para este gen segun la base de datos consultada (UCSC Genome Browser) (figura 3-12 C).

La RTN del L1 ocurrida en este gen fue identificada entre los exones 7 y 8 en el intrón 7 (locus 8819372) y tuvo lugar en dirección transcripcionalmente opuesta al gen (figura 3-12 A y B)..

pRAMPA1D_4- OSBPL6

Este gen codifica para un miembro de la familia de proteínas de unión al oxysterol (OSBP), que hacen parte del grupo de receptores lipídicos intracelulares. La mayoría de sus miembros contienen un dominio PH N-terminal, y un dominio altamente conservado C-terminal OSBP semejante a unión de esteroides. Inmunotinciones de la proteína expresada en musculo esquelético mostró una fuerte reactividad, mientras que en musculo liso las células fueron ligeramente teñidas. En evaluaciones hechas en otros tipos de tejidos normales no hubo evidencia de expresión.

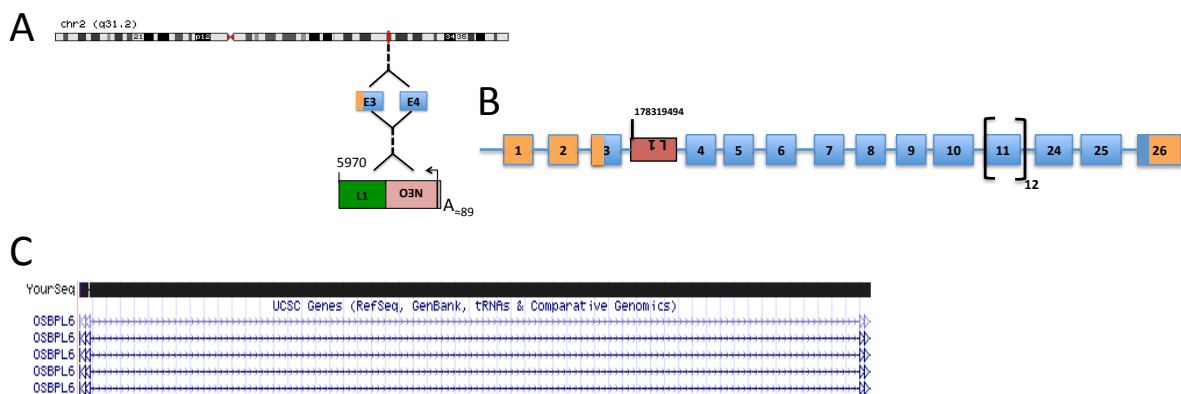


Figura 3-13. Mapeo cromosómico y molecular de la inserción del L1 en el gen OSBPL6. A) Ubicación cromosómica de la inserción y su ubicación frente al gen, en donde se muestra el grado de truncación y la extensión de la cola de poli A. B) Esquema general del gen OSBPL6 en el que se identifica la inserción del L1 (rojo) con el locus cromosómico en el extremo 3' y su orientación con respecto al gen, los exones (azul) y secuencias UTR (naranja). C) Isoformas identificadas en la base de datos (Genome Browser).

Se han identificados variantes de transcripción que codifican 5 diferentes isoformas (figura 3-13 C). El gen esta ubicado en la cadena con sentido del cromosoma 2 (2q31.2) y está constituido en total por 26 exones incluyendo las regiones UTRs (figura 3-13 A y B). El sitio de inserción de L1 fue identificado en el locus 178319494 que representa el intrón 3 entre los exones 3 y 4 del gen (figura 3-13 A y B). La dirección de la secuencia de inserción es transcripcionalmente opuesta a la del gen (figura 3-13 B).

pRAMPA1D_6-FAM65B

Este gen codifica para una proteína que estimula la formación de un sincitio multinucleado no mitótico de citotrofoblastos proliferativos durante la diferenciación del trofoblasto. Tejidos como los encontrados en la vesícula biliar, túbulos renales, conductos pancreáticos y biliares se ha encontrado una fuerte presencia de esta proteína. Con respecto a la expresión vista en líneas celulares, se identifico que en aquellas con origen linfoide poseían una mayor expresión con respecto a las demás líneas analizadas.

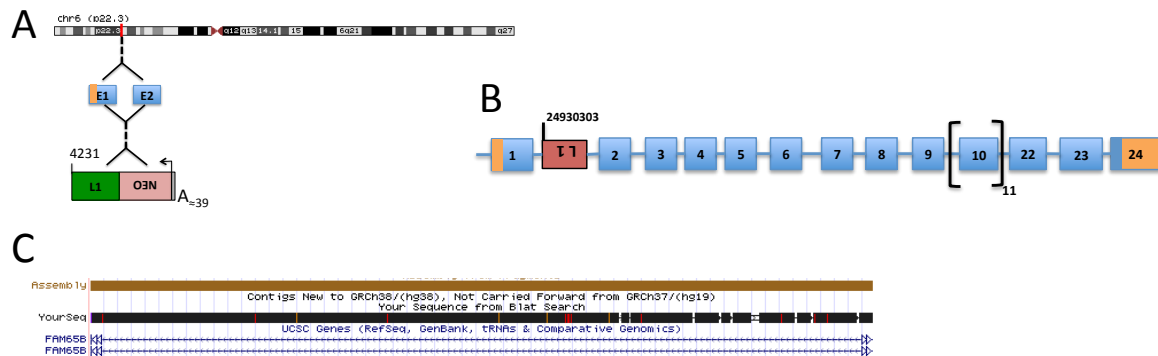


Figura 3-14. Mapeo cromosómico y molecular de la inserción del L1 en el gen FAM65B. A) Ubicación cromosómica de la inserción y su ubicación frente al gen, en donde se muestra el grado de truncación y la extensión de la cola de poli A. B) Esquema general del gen FAM65B en el que se identifica la inserción del L1 (rojo) con el locus cromosómico en el extremo 3' y su orientación con respecto al gen, los exones (azul) y secuencias UTR (naranja). C) Isoformas identificadas en la base de datos (Genome Browser).

Splicing alternativo de este gen da lugar a 2 variantes de transcripción (figura 3-14 C). El gen está ubicado en la cadena antisentido del cromosoma 6 (6p22.3), y está formado por 24 exones incluyendo las regiones UTR (figura 3-14 A y B). La inserción del L1 se identificó en el locus 24930303 que pertenece al intrón 1 del gen, entre los exones 1 y 2, y está ubicada en dirección transcripcional opuesta a la dirección del gen (figura 3-14 A y B).

pRAMPA1D_8- LPHN2

La proteína codificada por este gen es miembro de la subfamilia de las latrofilinas que son receptores acoplados a la proteína G (GPCR del inglés G-protein-coupled receptors). La proteína resultante participa en la regulación de la exocitosis. Se piensa que quiebres proteolíticos endógenos en un dominio rico en cisteína GPS (del inglés G-protein-coupled receptor proteolysis site) resultan en la formación de dos subunidades (una subunidad extracelular grande N-terminal de adhesión y otra subunidad de gran similitud a la familia de las secretina/calcitonina de las GPCRs) las cuales no son unidas covalentemente en la membrana celular. Inmunoanálisis hechos en tejidos sanos mostraron una mediana a una fuerte expresión en 17% de los tejidos analizados (el mismo resultado fue observado en cáncer). Dentro de estos se observó una positividad diferencial entre las membranas lumbinales del intestino y la vesícula

biliar. Además de una fuerte presencia citoplasmática en músculo estriado y liso.

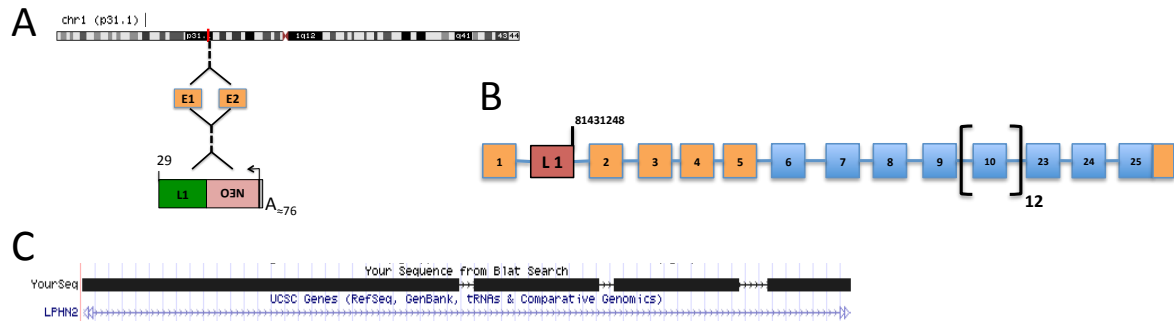


Figura 3-15. Mapeo cromosómico y molecular de la inserción del L1 en el gen LPHN2. A) Ubicación cromosómica de la inserción y su ubicación frente al gen, en donde se muestra el grado de truncación y la extensión de la cola de poli A. B) Esquema general del gen LPHN2 en el que se identifica la inserción del L1 (rojo) con el locus cromosómico en el extremo 3' y su orientación con respecto al gen, los exones (azul) y secuencias UTR (naranja). C) Isoformas identificadas en la base de datos (Genome Browser).

Solo se observó una variante de transcripción en la base de datos consultada (figura 3-15 C). El gen está ubicado en la cadena sentido del cromosoma 1 (1p31.1), y está compuesto por 25 exones incluyendo las regiones UTR (figura 3-15 A y B). El evento de RTN se identificó en el intrón 1 entre los exones 1 y 2 (locus 81431248) (figura 3-15 B). La inserción se encuentra en la misma dirección transcripcional del gen (figura 3-15 A y B).

pRAMPA1D_15- RAD51B

La proteína codificada por este gen es un miembro de la familia de proteínas RAD51. Los miembros de esta familia se conocen por estar involucrados en la recombinación homóloga y la reparación del ADN. Se ha visto que esta proteína forma un heterodímero estable con miembros de la familia RAD51C, el cual además interactúa con otros miembros de la familia como RAD51, XRCC2, Y XRCC3. La sobre-expresión de este gen se ha visto que causa retraso en el ciclo celular específicamente en G1 además de apoptosis celular, lo que sugiere el rol de esta proteína como censora del daño al ADN. Ensayos por anticuerpos en tejidos sanos mostraron que el 97% de las muestras presentaban una fuerte inmunoreactividad nuclear, en el caso de muestras de cáncer el porcentaje aumentó al 100%.

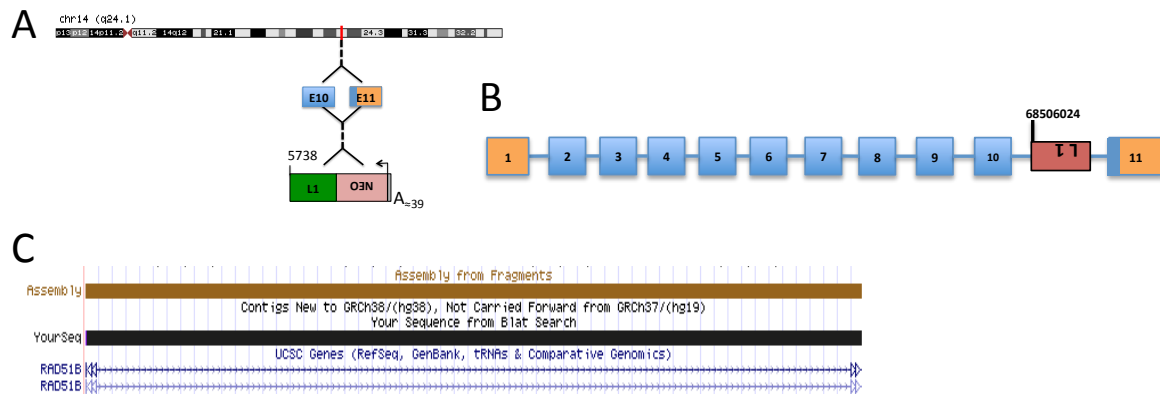


Figura 3-16. Mapeo cromosómico y molecular de la inserción del L1 en el gen RAD51B. A) Ubicación cromosómica de la inserción y su ubicación frente al gen, en donde se muestra el grado de truncación y la extensión de la cola de poli A. B) Esquema general del gen RAD51B en el que se identifica la inserción del L1 (rojo) con el locus cromosómico en el extremo 3' y su orientación con respecto al gen, los exones (azul) y secuencias UTR (naranja). C) Isoformas identificadas en la base de datos (Genome Browser).

Se ha observado que al menos 2 variantes de transcripción formados por splicing alternativo codifican para ambas isoformas (figura 3-16 C). Rearreglos entre este locus y el grupo AT-hook 2 de alta movilidad (HMGA2) han sido observados en leiomioma uterino. El gen está ubicado en la cadena con sentido del cromosoma 14 (14q24.1), y está formado por 12 exones incluyendo las regiones UTR (figura 3-16 A y B). La inserción del L1 fue identificada en el locus 68506024 en dirección transcripcional contraria al gen, en el intrón 10 entre los exones 10 y 11 (figura 3-16 A y B).

pRAMPA1D41_ MAPK10

La proteína codificada por este gen es un miembro de la familia de las MAP quinases. Esta familia actúa como punto integrador de múltiples señales bioquímicas, y están involucradas en una amplia variedad de procesos celulares, como la proliferación, diferenciación, regulación y desarrollo de la transcripción. Esta proteína es una forma neuronal específica de las quinases c-Jun N-terminal (JNKs). A través de su fosforilación y localización nuclear, esta quinasa desempeña funciones reguladoras en las vías de señalización durante la apoptosis neuronal. La quinasa dependiente de ciclina 5 puede fosforilar e inhibir la actividad de esta quinasa el cual puede ser importante en el control de la apoptosis neuronal.

Análisis de expresión proteica por inmunotinciones, reportaron que solo el 9% de las muestras evaluadas poseían un grado de tinción medio o alto, tal es el caso de los tejidos del sistema nervioso central. Tinciones diferenciales citoplasmáticas fueron vistas en neuronas y neuropilos, además también fueron positivas muestras de nervios periféricos e islotes pancreáticos. En líneas celulares en general la expresión a partir de ARN es baja, en muestras de cáncer solo el 7% fue positivo ante la tinción por anticuerpo.

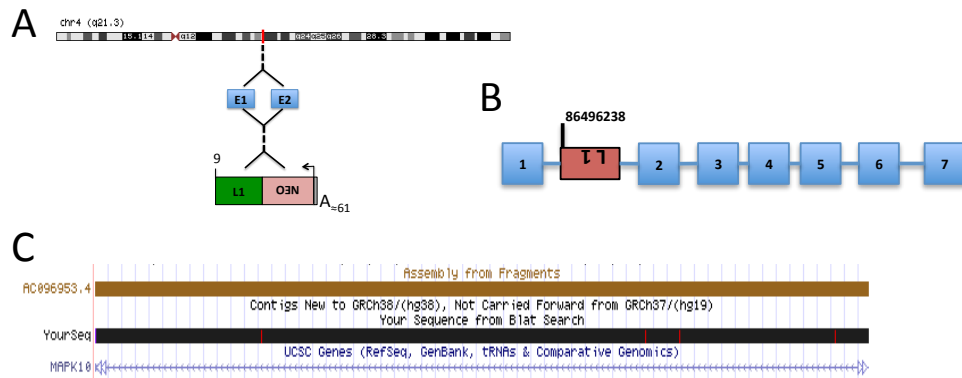


Figura 3-17. Mapeo cromosómico y molecular de la inserción del L1 en el gen MAPK10. A) Ubicación cromosómica de la inserción y su ubicación frente al gen, en donde se muestra el grado de truncación y la extensión de la cola de poli A. B) Esquema general del gen MAPK10 en el que se identifica la inserción del L1 (rojo) con el locus cromosómico en el extremo 3' y su orientación con respecto al gen, los exones (azul). C) Isoformas identificadas en la base de datos (Genome Browser).

Se reportó una sola variante de transcripción para el gen MAPK10 (figura 3-17 C). El gen está ubicado en la cadena antisentido del cromosoma 4 (4q21.3), además de estar formado por 7 exones (figura 3-17 A y B). La inserción del L1 se identificó en la cadena con sentido transcripcionalmente en sentido opuesto al gen en el locus 87417391 en el intrón 1 entre los exones 1 y 2 (figura 3-17 A y B).

pRAMPA1D44_LOC100287944

La inserción del L1 en este caso impactó un gen de ARN denominado LOC100287944 (el cual permanece sin caracterizar) que pertenece a la clase de los denominados ARN no codificantes (ncARN). El gen está ubicado en la cadena sin sentido del cromosoma 12 (12q23.3), y está conformado por 4 exones (figura 3-18 A y B). El evento de RTN se determinó en el intrón 1 entre los exones 1 y 2, en dirección transcripcional opuesta al gen (figura 3-18 A y B).

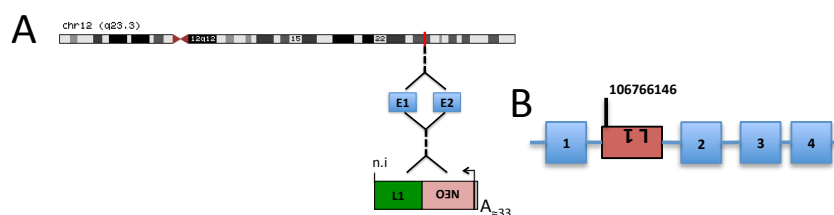


Figura 3-18. Mapeo cromosómico y molecular de la inserción del L1 en el gen LOC100287944. A) Ubicación cromosómica de la inserción y su ubicación frente al gen, en donde se muestra el grado de truncación y la extensión de la cola de poli A. B) Esquema general del gen LOC100287944 en el que se identifica la inserción del L1 (rojo) con el locus cromosómico en el extremo 3' y su orientación con respecto al gen, además del conjunto de exones (azul).

pRAMPA1D45_ FRAS1

Este gen codifica para una proteína de la matriz extracelular que parece que cumple una función en la regulación de la adhesión de la membrana basal epidermal y la organogénesis durante el desarrollo. Mutaciones en este gen causa el síndrome de Fraser, que consiste en una malformación multisistémica que puede incluir anomalías en los sistemas cráneo-facial, urogenital y respiratorio. Ensayos de la expresión proteica por inmunotinción mostró que aproximadamente el 68% del total de las muestras analizadas mostraron una de media a alta. Igualmente la mayoría de los tejidos mostraron una reactividad moderada citoplasmática y/o membranosa. Dentro de los tejidos sanos con mayor reactividad encontramos los de pulmón, la vesícula biliar, Langerhans de la piel, células granulares de mama y células hematopoyéticas.

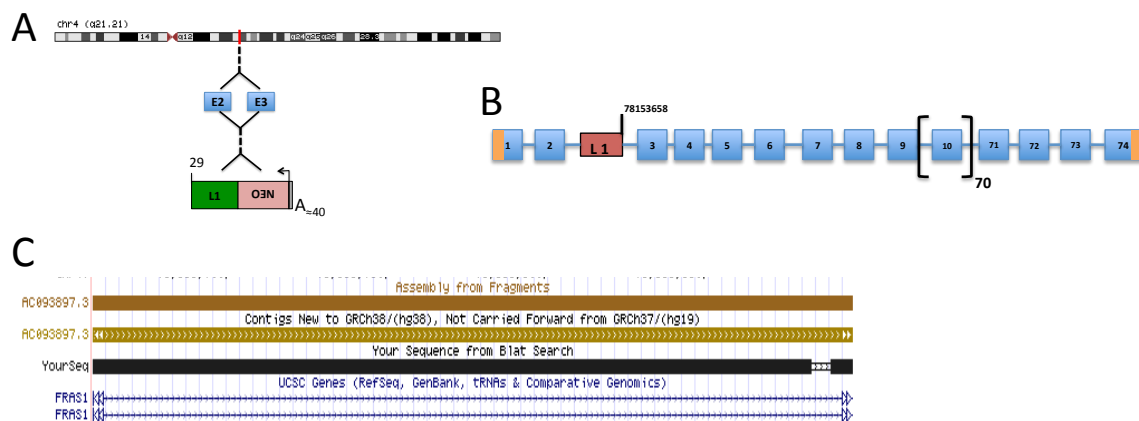


Figura 3-19. Mapeo cromosómico y molecular de la inserción del L1 en el gen FRAS1. A) Ubicación cromosómica de la inserción y su ubicación frente al gen, en donde se muestra el grado de truncación y la extensión de la cola de poli A. B) Esquema general del gen FRAS1 en el que se identifica la inserción del L1 (rojo) con el locus cromosómico en el extremo 3' y su orientación con respecto al gen, los exones (azul) y secuencias UTR (naranja). C) Isoformas identificadas en la base de datos (Genome Browser).

Splicing alternativo del gen resulta en la formación de 2 variantes de transcripción (figura 3-19 C). El gen está ubicado en la cadena con sentido del cromosoma 4 (4q21.21) y está constituido por 74 exones incluyendo las regiones UTRs (figura 3-19 A y B). Este gen fue impactado por el L1 en el intrón 2 (locus 78153658) entre los exones 2 y 3. Tanto el gen como el L1 están ubicados en la cadena con sentido del cromosoma 4 (figura 3-19 A y B).

RECOMBINACIÓN FLP: ANÁLISIS DE LA ESCISIÓN DEL CASETE DE NEO

Con el fin de obtener un tipo de inserción *in vitro* que imite la constitución molecular del un L1 endógeno (L1 naturalizado), mediante el uso de la FLP recombinasa, se eliminó del L1 la secuencia del casete reportero (*neo*) presente en nuestro evento de inserción (L1 transfectado y modificado genéticamente) presente en las líneas celulares generadas. Una vez eliminada esta secuencia, las células nuevamente vuelven a adquirir el estado de sensibilidad al antibiótico G418 (figura 3-20)

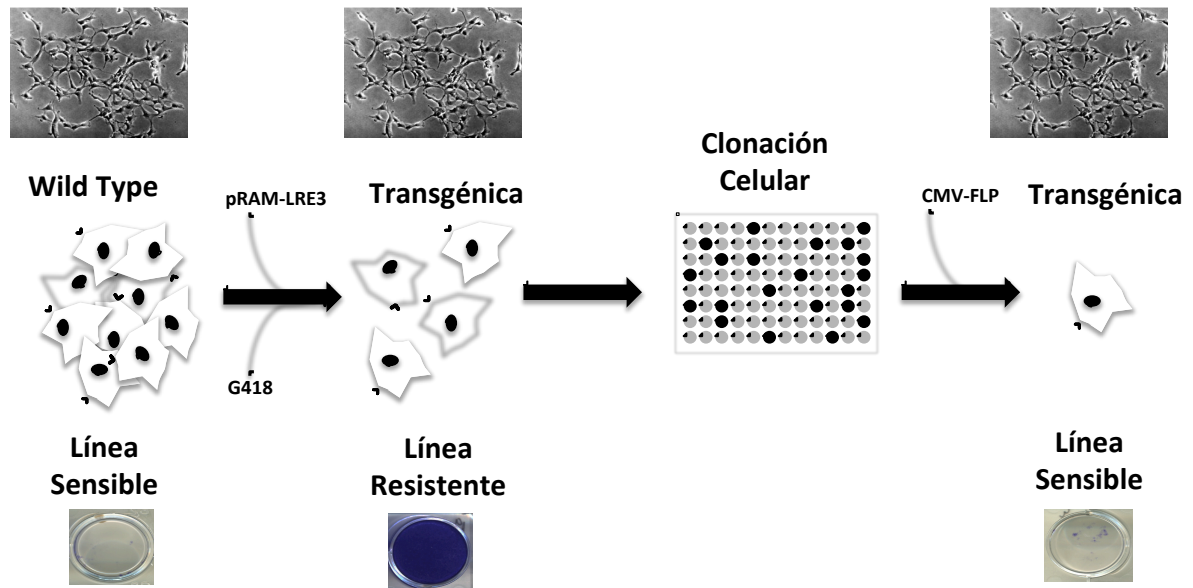


Figura 3-20. Representación grafica de la recuperación de la sensibilidad al antibiótico G418.

Determinación Molecular por PCR-NEO

Para la generación de las sublíneas clonales, las líneas celulares anteriormente descritas fueron transfectadas con el plásmido *cmv-FLP* junto con el *pCEP-eGFP* (ver materiales y métodos). Las células GFP positivas fueron sorteadas, separadas y puestas en cultivo nuevamente para su recuperación. La generación de los subclones se llevó a cabo realizando el procedimiento de dilución clonal con el fin de obtener la mayor cantidad de clones posibles para ser analizados (figura 2-4).

Inicialmente el comportamiento de todas las células después de pasar por el sorter fue heterogéneo, debido a esto el número de clones obtenidos para los análisis fue variable para todas las líneas celulares. Aquellas líneas celulares que permitieron el aislamiento y expansión de una gran cantidad de clones, en primera instancia se les realizó un ensayo preliminar de presencia del casete de *neo* (PCR-NEO), por medio del buffer de extracción rápida QuickExtract (figura 2-1). Una vez obtenidos los clones candidatos (figura 3-21 B), dicha población era expandida para ser igualmente analizada por la misma PCR, con el fin de disminuir los sesgos causados por el sistema de extracción rápida del ADN genómico (Figura 3-21 C).

La baja recuperación de algunas líneas celulares luego del sorter, nos permitió expandir directamente estos subclones celulares para ser analizados por PCR NEO convencional (Figura 3-21 C).

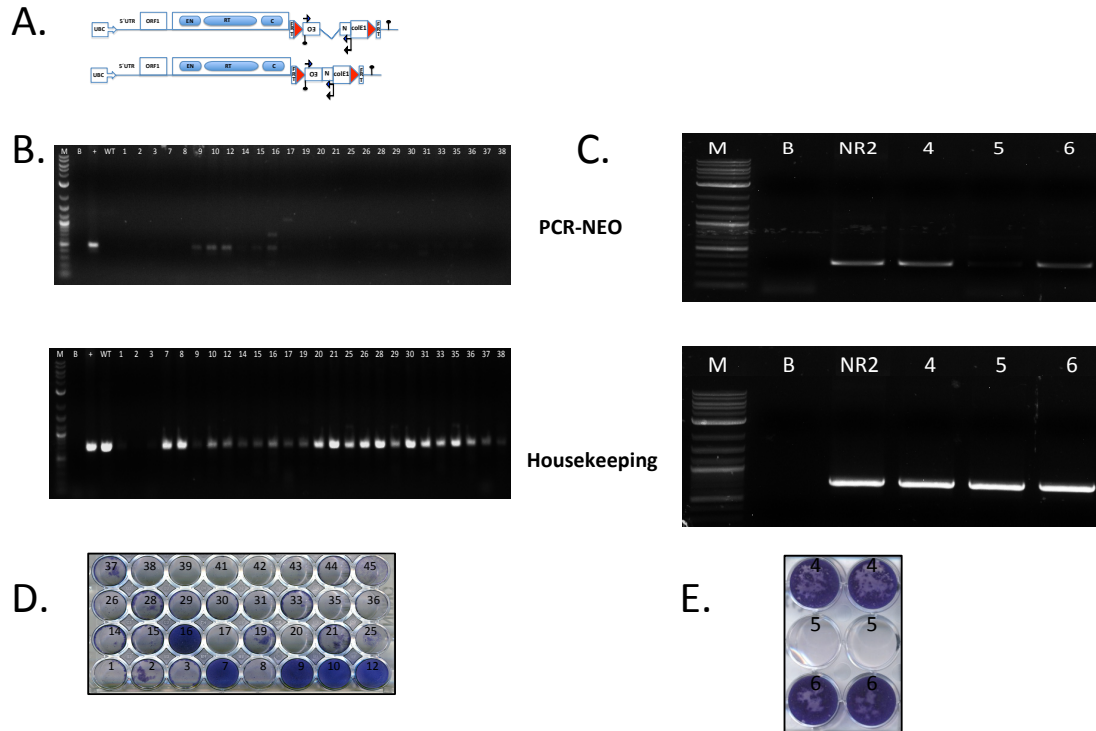


Figura 3-21. Identificación de la presencia del casete de neo por PCR-NEO. A) Representación gráfica de las longitudes del pRAM/LRE3 con y en ausencia del intrón. B) PCR-NEO a partir del ADN genómico generado por el buffer QuickExtraction. C) PCR-NEO a partir de ADN genómico extraído por medios convencionales. D y E) Ensayo de resistencia a G418 para las células sorteadas.

De las 13 líneas generadas con inserciones génicas solo se pudo eliminar el casete reportero en 9 de ellas, debido a que fue imposible obtener una estabilidad celular una vez fueron sorteadas las células (tabla 3-2)

Tabla 3-2: Lista de líneas celulares transgénicas impactadas en unidades génicas y subclones seleccionados luego de la eliminación del casete reportero *neo*.

Nombre de la línea	Gen	Subclon (<i>Flp</i>)
pRAMPA1_6	CPNE4	34
pRAMPA1_12	DDAH1	-
pRAMPA1NR_2	TRPM3	5
pRAMPA1NR_9	ARHGAP42	8
pRAMPA1NR_10	FGGY	4
pRAMPA1NR_13	CASC15 (ncRNA)	-
pRAMPA1NR_25	PMM2	49
pRAMPA1_D4	OSPLB6	14
pRAMPA1_D6	FAM65B	3
pRAMPA1_D8	LPNH2	3
pRAMPA1_D15	RAD51B	4
pRAMPA1_D41	MAPK10	-
pRAMPA1_D45	FRAS1	-

Subclones celulares sensibles

Los subclones celulares que previamente habían demostrado por PCR-NEO que no presentaban el casete de resistencia (ausencia de banda), fueron sometidos a un ensayo de selección con G418 (ver materiales y métodos) con el fin de confirmar desde el punto de vista celular los resultados obtenidos a nivel molecular (figura 3-21 E y D). En dichos ensayos, en los casos que se tenían clones negativos, igualmente se expandían con el fin de tener un control negativo dentro de los subclones.

En la figura 3-21 se puede observar ejemplos de dos líneas celulares diferentes evaluadas, en donde se observa que se pudo recobrar la sensibilidad al G418. Los clones seleccionados fueron aquellos que presentaron una total sensibilidad antes el antibiótico usado y que coincidían con los resultados moleculares.

PCR-FLP

Hasta el momento los resultados desde el punto de vista molecular se habían obtenido con base en la ausencia del casete de *neo*, es decir, la ausencia de banda indicaría la acción positiva de la flipasa. Con el fin de tener resultados con base en presencia de banda del fragmento escindido, realizamos una *pcr* específica que nos permitiera analizar por presencia de banda la delección del casete *neo*.

Para esto se sintetizó un oligoque alineaba en la región del promotor SV40 como cebador antisentido (SV40_REV2_FLP) y el “walkingprimer” JORF2H-S del L1 ubicado en la región del ORF2 como cebador sentido (ver materiales y métodos). Un evento positivo daría como resultado un producto de amplificación de aproximadamente 780 pb, mientras que un evento negativo daría un amplicón de aproximadamente 3,1 Kb. (Figura 3-22 A).

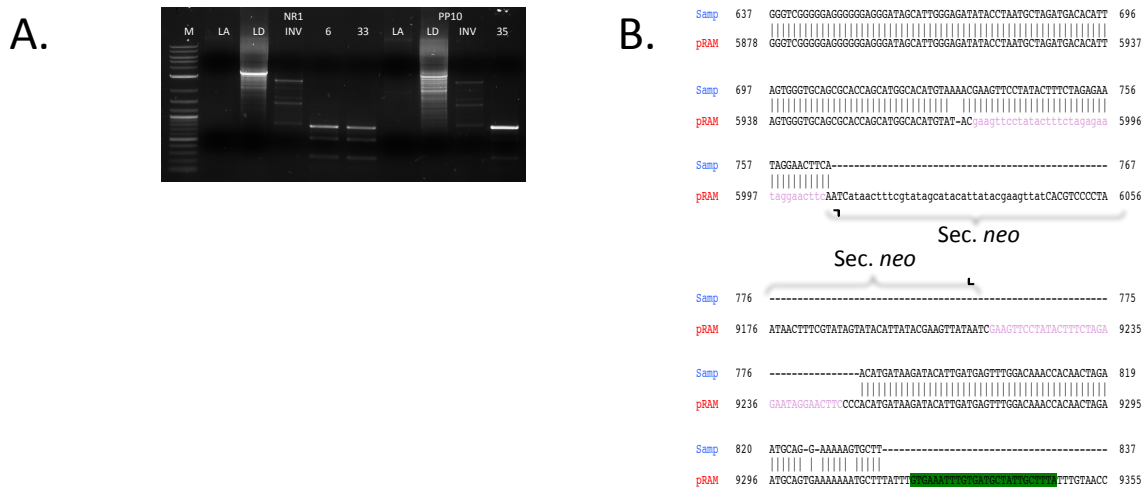


Figura 3-22. Determinación de la recombinación mediada por Flp. A) PCR-Flp. LA=Agua; LD=Muestras parentales con casete neo. B) Alineamiento del amplicón con la secuencias del

pRAM/LRE3. En rosa se muestra la secuencias FRT, y en verde se representa la ubicación de primer antisentido usado.

El análisis de las secuencias de los fragmentos generados evidenció el real efecto de la flipasa en la generación de inserciones del L1 con características moleculares semejantes a los L1s activos endógenos (Figura 3-22 B). Las longitudes de las inserciones resultantes se mostradas en la tabla 3-3.

Tabla 3-3. Relación de la longitud de las inserciones antes (Long trans) y después de ablación del casete reportero *neo* (Long Nat) con la orientación de las inserciones del L1.

Nombre de la línea	Gen	Orientación	Long trans (nt)	long Nat (nt)
pRAMPA1_6	CPNE4	Sentido	3734	492
pRAMPA1_12	DDAH1	Antisentido	9443	-
pRAMPA1NR_2	TRPM3	Sentido	9490	6248
pRAMPA1NR_9	ARHGAP42	Antisentido	n/i	n/i
pRAMPA1NR_10	FGGY	Antisentido	3895	653
pRAMPA1NR_25	PMM2	Antisentido	n/i	-
pRAMPA1_D4	OSPLB6	Antisentido	4407	1165
pRAMPA1_D6	FAM65B	Antisentido	5196	1954
pRAMPA1_D8	LPNH2	Sentido	9435	6193
pRAMPA1_D15	RAD51B	Antisentido	3689	-
pRAMPA1_D41	MAPK10	Antisentido	9440	-
pRAMPA1_D45	FRAS1	Sentido	9399	-

ANALISIS DE LA EXPRESION GENICA

Debido a las características de nuestras muestras, el análisis de la expresión génica se realizó por medio del sistema OneStep RT-qPCR, dada las ventajas que representa (ver materiales y métodos). La expresión de la mayoría de los genes fue analizada con respecto a dos parejas de primers que evaluaban la expresión del gen desde dos posiciones diferentes tomando como referencia la ubicación del L1.

Inicialmente se realizó una evaluación de la eficiencia de los primers en PA-1 realizando curvas estándar de cada una de las parejas y comparándolas con el nivel del expresión de un *housekeeping* que en este caso fue el GAPDH. Una vez realizado en análisis de primers pudimos optimizar la RT-qPCR para 12 de las 13 líneas celulares, igualmente en algunos casos fue imposible la estandarización de las dos parejas de primers, por lo que en estos casos solo se tiene un patrón de referencia dentro del gen.

Variación de la Expresión Génica en los Subclones Celulares Flipeados en Estado Pluripotente

Inicialmente se pretendió evaluar la expresión de los genes impactados por el L1 cuando éste esta en un estado “artificial” (presencia del casete *neo*), y

cuando se ha “*naturalizado*” (sin presencia del casete de resistencia). La evaluación se realizó con base en los resultados de expresión obtenidos en tres “replicas biológicas” y cada una a su vez poseían tres replicas técnicas. Como control en la evaluación del cambio en los niveles de expresión se tomo tres muestras de ARN de células PA-1 extraídas independientemente en donde igualmente se evaluaban por triplicado.

Análisis de la expresión del gen CPNE4

Teniendo en cuenta que la ubicación de la inserción del L1 ocurrió entre los exones 1 y 2 del gen (figura 3-23), la primera pareja de primers evalúa la expresión del gen CPNE4 en la intersección de los exones 1 y 2 como lo muestra la figura 3-23 B. Aguas abajo de la inserción se diseñaron los primers en una región aleatoria que comprendió los exones 15 y 16 (figura 3-23 B).

En el primer caso, se puede observar que cuando el L1 es *naturalizado* la expresión del gen disminuye significativamente con respecto a los niveles de su línea parental (es decir, aquella que presenta el casete de *neo* en su genoma). En general estas variaciones no representan un cambio relevante en la expresión, cuando ambos valores son comparados con los niveles de expresión basales del la línea celular silvestre (figura 3-23 C).

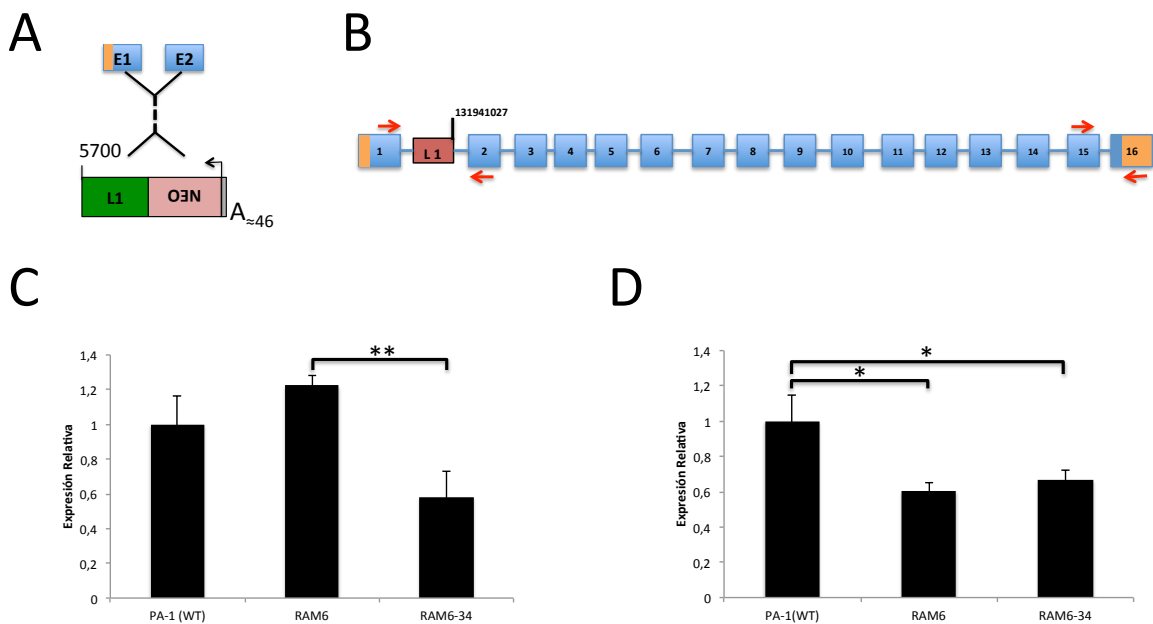


Figura 3-23. Análisis de la expresión del gen CPNE4 en la línea celular transgénica pRAM-PA1_6. A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C y D) Niveles de expresión del gen por RT-qPCR.

Al comparar los niveles de expresión aguas abajo del gen, se pudo observar una disminución notable en los niveles de expresión, tanto para la línea parental transgénica como el subclon *naturalizado*, evidenciando un patrón de expresión diferente al mostrado con la evaluación previa (figura 3-23 D). Los resultados de expresión para este gen, nos mostró que los niveles de

expresión del CPNE4 se vio influenciado por la inserción del L1 en donde en términos generales disminuyó su expresión teniendo en cuenta el tipo de inserción y las condiciones evaluadas.

El tipo de orientación (sentido) y el tamaño adquirido por la línea transgénica parental también son características que pueden determinar la disminución en la expresión del gen por fallos en la elongación del transcrito¹³⁰. Teniendo en cuenta que la inserción se encuentra ubicada en el extremo 5', problemas en la elongación serían mas evidentes en los exones alejados del la inserción en dirección 3'. Lo anterior podría explicar la disminución en los niveles de expresión del gen cuando el transcrito es evaluado en los exones 15 y 16.

A pesar de observar una clara influencia de la inserción del L1 en la expresión del gen CPNE4, datos semejantes en el comportamiento de la expresión del gen en ambas líneas celulares permiten excluir un efecto directo de la secuencia reportera en los resultados obtenidos.

Análisis de la expresión del gen TRPM3

La inserción del L1 en el gen TRPM3 se encuentra entre la región 5' UTR (formada por 3 exones) y el inicio de la región codificante. Aguas arriba de la inserción los primers alineaban en los dos primeros exones, mientras que aguas abajo no se consiguió obtener ningún amplicón del gen (figura 3-24 A y B).

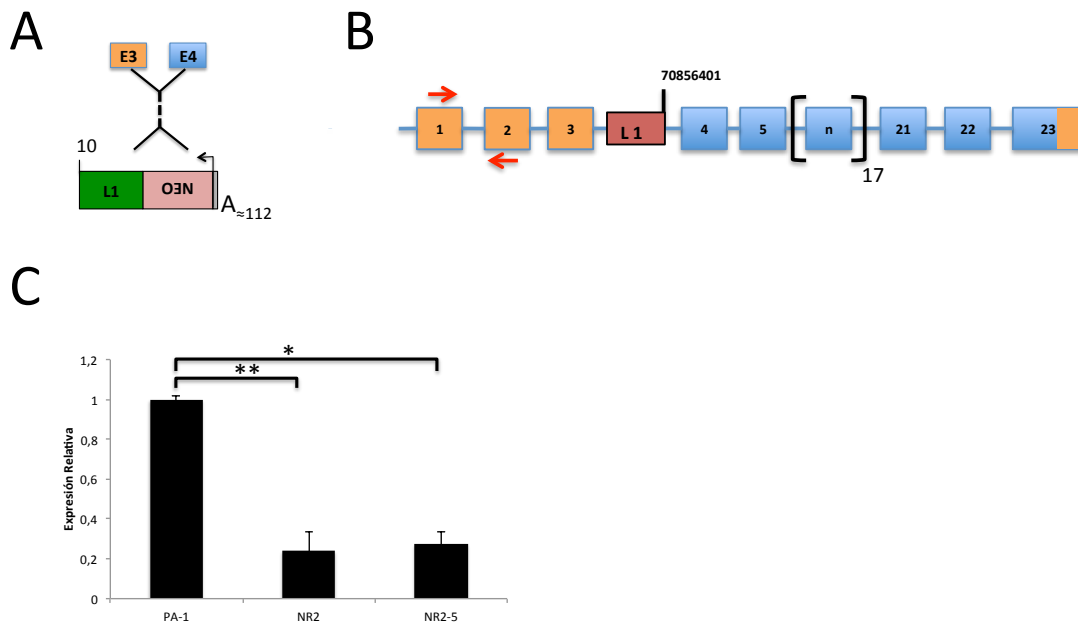


Figura 3-24. Análisis de la expresión del gen TRPM3 en la línea celular transgénica pRAM-PA1NR_2. A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C) Niveles de expresión del gen por RT-qPCR.

Cuando se analiza la expresión del gen con la primera pareja de primers aguas arriba de la inserción, se observa claramente una drástica disminución de la expresión del gen tanto para la línea parental transgénica como para el subclon *naturalizado* (figura 3-24 C). Los presentes datos aportan evidencias claras de reducción de la expresión del gen TRPM3 a raíz de la inserción del L1, y dicha reducción es independiente de la presencia del casete reportero presente en la línea transgénica parental (figura 3-24 C).

La disminución en la expresión de este gen esta en sintonía con lo reportado en estudios previos, en donde inserciones de L1s intrónicas cercanas o de copia completa en dirección sentido desencadenan una reducción en los niveles de expresión del gen hospedero^{130;133}.

Análisis de la expresión del gen ARHGAP42

Teniendo en cuenta que la inserción del L1 yace entre los exones 4 y 5 del gen (figura 3-25 A), la pareja de primers usada aguas arriba de la inserción esta ubicada en los exones 1 y 2 (figura 3-25B). Así mismo los exones usados para identificar la expresión del gen aguas abajo de la inserción fueron ubicados en los exones 21 y el 22 (figura 3-25 B).

Al analizar los niveles de expresión del gen ARHGAP42 aguas arriba y aguas debajo de la inserción del L1, en ambos casos, los niveles de expresión no fueron alterados por el impacto del L1 (figura 3-25C y D); esto tanto para la línea transgénica parental como para el subclon *naturalizado*.

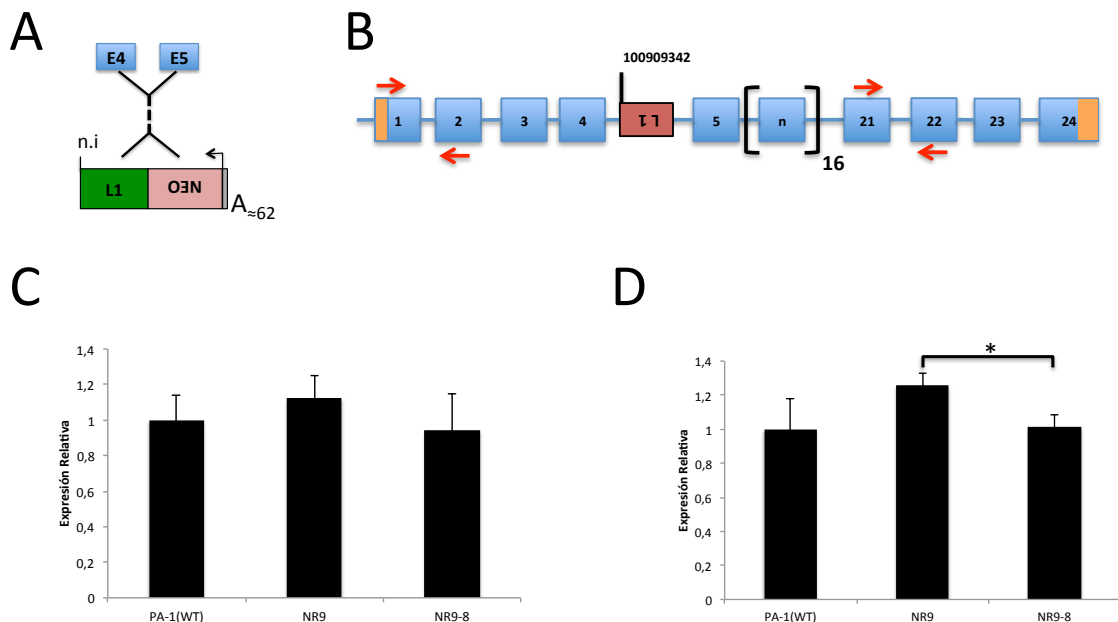


Figura 3-25. Análisis de la expresión del gen ARHGAP42 en la línea celular transgénica pRAM-PA1NR_9. A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo

intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C y D) Niveles de expresión del gen por RT-qPCR.

A pesar que, los niveles de expresión hallados en la evaluación con los primers aguas debajo de la inserción fueron semejantes a los mostrados por la línea PA-1 WT, se pudo observar una diferencia significativa entre la línea parental transgénica y su clon *naturalizado* (figura 3-25 D).

Estos resultados no pudieron ser correlacionados con la longitud del L1, aunque actualmente se conoce que las inserciones que se encuentran orientadas antisentido, poseen un menor efecto hablando especialmente de los niveles de expresión de los genes portadores de dichas inserciones¹⁴².

Análisis de la expresión del gen FGGY

La inserción en este gen tuvo lugar entre los exones 6 y 7 (figura 3-26 A). El análisis de la expresión del gen aguas arriba de la inserción del LINE se hizo usando una pareja de primers ubicados en los exones 4 y 5. A su vez, la ubicación de los primers aguas debajo del sitio de inserción abarcaron los exones 13 y 14 (figura 3-26B).

En el análisis de la expresión del gen agua arriba de la inserción, se observó claramente una reducción de los niveles de expresión de las líneas impactadas con el L1 cuando son comparadas con los niveles basales de las PA-1 silvestres, tanto para la línea que presenta la inserción con el casete reportero como para su subclon *naturalizado* (figura 3-26C). Por otro lado, al observar el patrón de expresión mostrado al usarse los primers aguas bajo de la inserción, no se observan diferencias cambios en la expresión del gen en ninguna de las formas evaluadas (figura 3-26D).

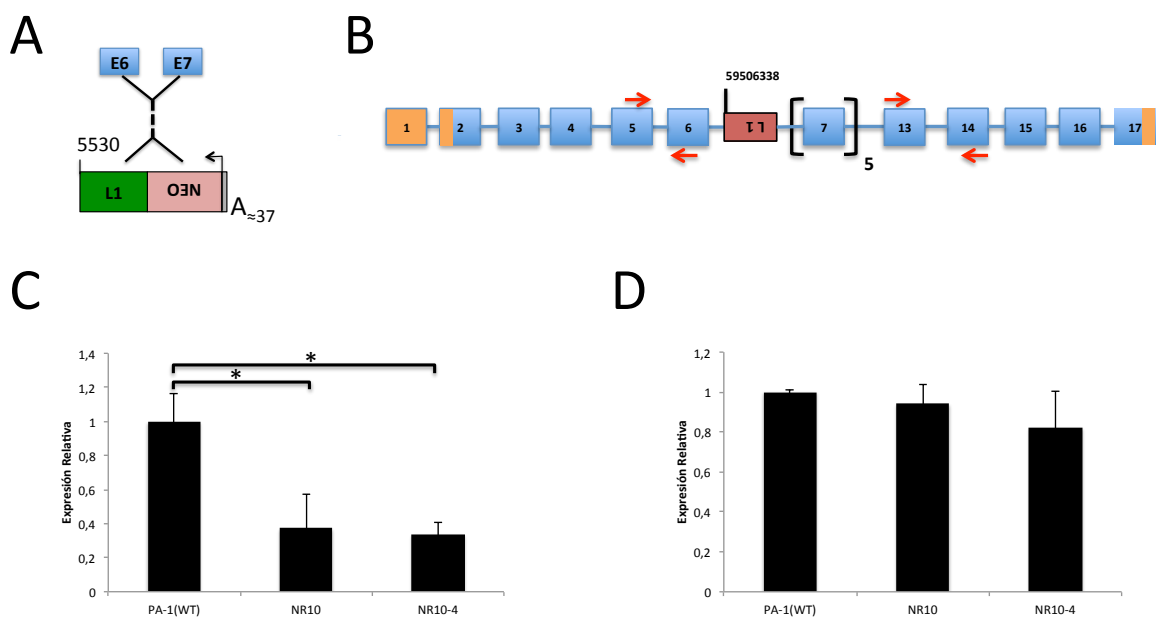


Figura 3-26. Análisis de la expresión del gen FGGY en la línea celular transgénica pRAM-PA1NR_10. A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C y D) Niveles de expresión del gen por RT-qPCR.

Teniendo en cuenta los anteriores resultados, y a pesar de que tanto el tamaño del L1 como su orientación inicialmente no se esperaba un efecto notorio en la expresión del gen. Una posible explicación a la disminución en el nivel de expresión observado podría ser que el impacto del L1 este afectado el splicing alternativo del transcrito, fomentando la formación de nuevos transcritos del gen en el que la isoforma minoritaria resultante es a aquella que conserva los exones 5 y 6.

Análisis de la expresión del gen PMM2

Teniendo en cuenta que la inserción del L1 ocurrió entre los exones 7 y 8 (figura 3-27 A), el análisis de la expresión del gen aguas arriba de la inserción se hizo usando una pareja de primers ubicados en los exones 2 y 3 (figura 3-27B) . Además, se evaluó la región de intersección de la inserción con los exones que limitan con el L1 (exones 7 y 8) (figura 3-27B)

Semejante a lo observado con el gen CPNE4, la expresión del gen PMM2 impactado por el L1, en la línea celular parental transgénica aumenta sus niveles de expresión, mientras que los niveles del expresión parecen normalizarse en el momento que es eliminado el casete reportero de la secuencia del gen (figura 3-27C).

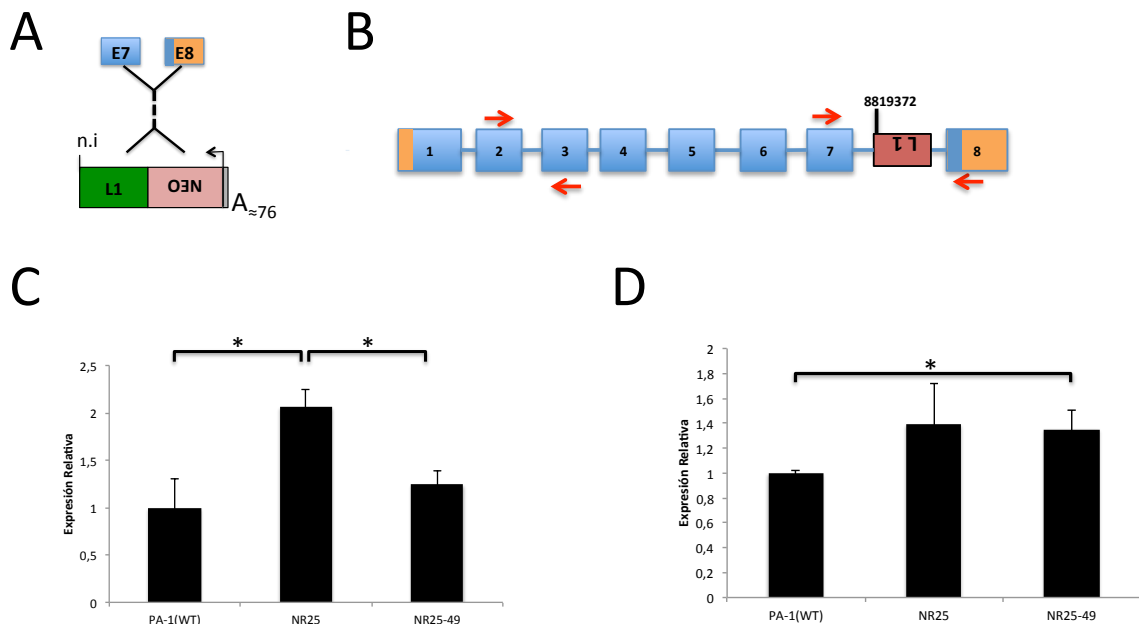


Figura 3-27. Análisis de la expresión del gen PMM2 en la línea celular transgénica pRAM-PA1NR_25. A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C y D) Niveles de expresión del gen por RT-qPCR.

de los primers usados en la evaluación de la expresión del gen. C y D) Niveles de expresión del gen por RT-qPCR.

En la figura 3-27D, se puede observar que la expresión del gen PMM2 aumenta cuando el L1 es *naturalizado* al igual que su línea parental transgénica el cual posee la misma tendencia cuando la expresión del gen es evaluada con la pareja de primers que amplifican la región de intersección que rodea la inserción del L1.

A pesar que no se pudo identificar el tamaño real de la inserción, es claro el efecto de la inserción en la sobre expresión, hecho hasta ahora no reportado, debido a la tendencia a la disminución de la expresión de los transcritos por inserciones intrónicas y más aún cuando estas están en dirección antisentido. Este efecto no es directamente atribuible a la presencia del casete de *neo* en el sistema, dado que este mismo incremento fue observado en la línea naturalizada en la evaluación de los exones 7 y 8 (figura 3-27D)

Análisis de la expresión del gen OSBPL6

La inserción del L1 fue identificada entre los exones 3 y 4 (figura 3-28 A), en tanto que la primera pareja de primers usadas en la evaluación de la expresión del gen OSBPL6 esta ubicada en el sitio de intersección con el elemento L1, que corresponder a los exones 3 y 4 (figura 3-28B). Así mismo los exones usados para identificar la expresión del gen aguas abajo de la inserción fueron el 9 y el 10 (figura 3-28B)

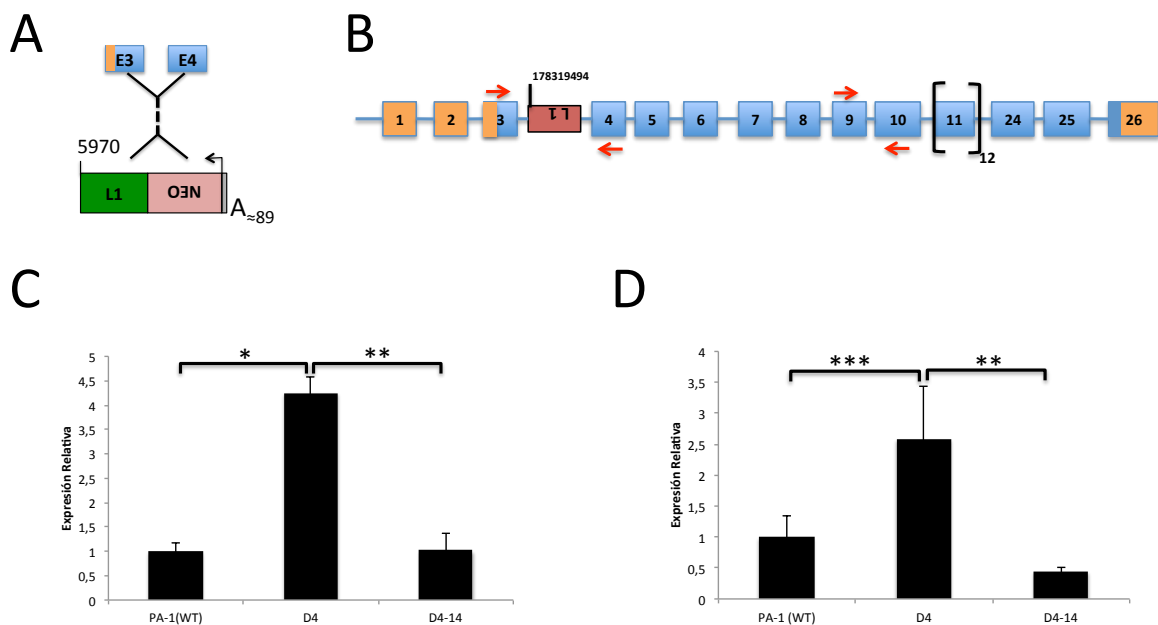


Figura 3-28. Análisis de la expresión del gen OSBPL6 en la línea celular transgénica pRAM-PA1D_4. A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C y D) Niveles de expresión del gen por RT-qPCR.

Los niveles de expresión del gen cuando fue evaluado tanto en el sitio de intersección de la inserción del L1 como aguas abajo de la misma, se vieron aumentados posiblemente influenciados por la presencia del casete *neo* en la estructura genómica. Igualmente cuando era retirado dicha secuencia reportera, los niveles de expresión disminuían significativamente alcanzando niveles semejantes a los observados en células silvestres PA-1 (figura 3-28B y C).

Al igual que los resultados observados en el gen *PMM2*, una variación no ortodoxa fue identificada, encontrando niveles superiores del gen a pesar de que la inserción se encuentra en orientación antisentido y sobrepasa la cantidad de secuencia de L1 encontrada por ejemplo en los genes de alta expresión en humanos (en el caso de inserciones antisentido con una media de 1760 nt) (HAN 2004). Todo esto y lo anterior hacen creer que efecto sobre regulador ejercido por la inserción del L1 este influenciado por la presencia de la secuencia reportera mas que el hecho que representa la inserción en si, y las posibles variaciones que puede esta desencadenar.

Análisis de la expresión del gen FAM65B

La inserción del L1 en este gen tuvo lugar en el intrón 1 (figura 3-29 A), por lo que la primera pareja de primers usada en la evaluación del nivel de expresión esta ubicada con respecto al sitio de intersección de la inserción entre los exones 1 y 2, mientras que la segunda pareja aguas abajo se diseño para los exones 9 y 10 (figura 3-29B).

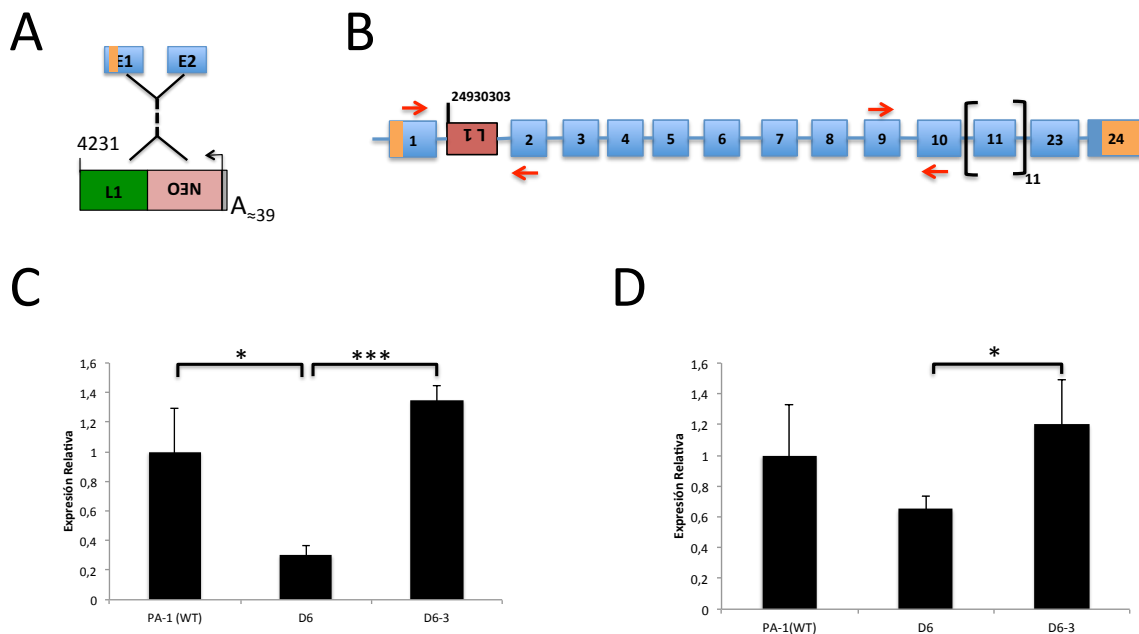


Figura 3-29. Análisis de la expresión del gen FAM65B en la línea celular transgénica pRAM-PA1D_6. A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C y D) Niveles de expresión del gen por RT-qPCR.

En la figura 3-29C podemos observar que la expresión del gen FAM65B se ve reducida notablemente cuando en su estructura se encuentra la inserción del L1 modificado genéticamente con el casete reportero *neo*, mientras que cuando se evalúa la expresión del su derivada naturalizada, la expresión recuperaba nuevamente los valores observados en células silvestres. Este hecho es igualmente observable al evaluar la expresión de los exones aguas debajo de la inserción, en donde el nivel de la línea parental transgénica *neo* tiende a reducir, y aumentándola notablemente una vez el casete reportero es eliminado (figura 3-29C).

Teniendo en cuenta que el tipo de inserción presente en el gen esta en dirección antisentido y que su los niveles de expresión no se ven afectados en la línea con el L1 *naturalizado*, posiblemente podríamos atribuir el efecto causado por la presencia del casete reportero causando un efecto contrario al visto en los resultados de los genes PMM2 y OSBPL6. Lo anterior esta en concordancia con experimentos previos donde han demostrado que los efectos sobre la atenuación de la expresión génica causado por inserciones intrónicas del L1 en muchos casos son dependientes de la longitud de la inserción (HAN 2005), de esta forma al disminuir la longitud del L1 en el proceso de naturalización podríamos estar restaurando la naturaleza no variante típica de las inserciones antisentido de corta longitud.

Análisis de la expresión del gen LPHN2

La inserción del L1 fue identificada en el intrón 1 del gen en una región UTR que esta conformada por 5 exones, por lo que fue difícil lograr una amplificación en la zona de intersección (figura 3-30 A y B). Es por esto que el nivel de expresión de este gen se evaluó solo con una pareja de primers ubicada aguas debajo de la inserción, y que amplifica una región entre los exones 7 y 8 (figura 3-30B).

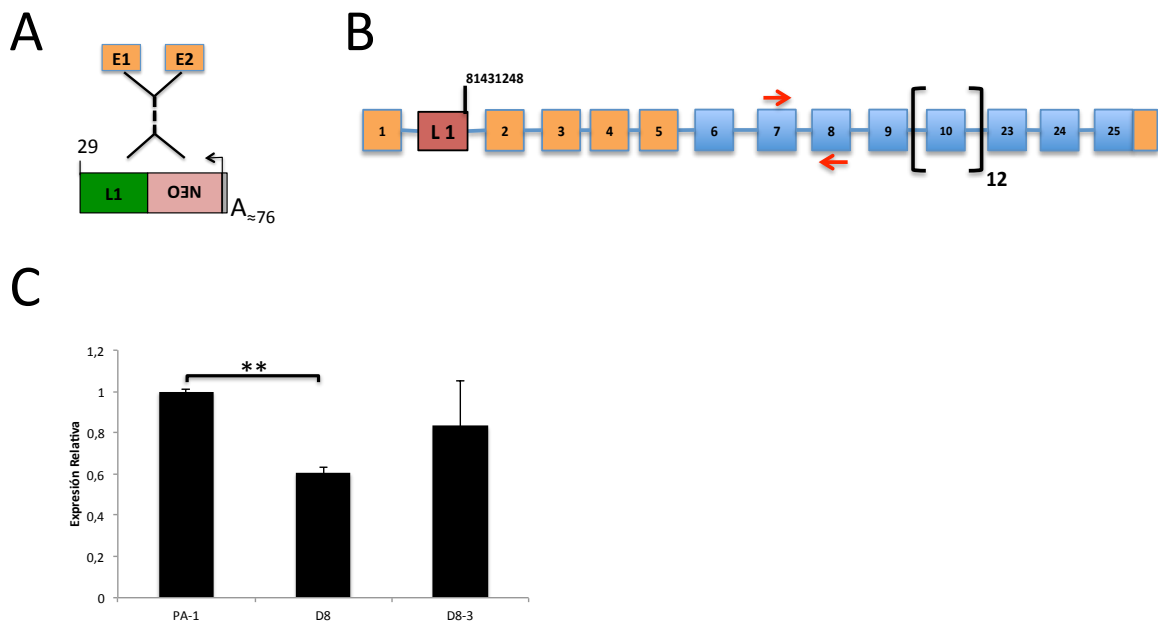


Figura 3-30. Análisis de la expresión del gen LPHN2 en la línea celular transgénica pRAM-PA1D_8. A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C) Niveles de expresión del gen por RT-qPCR.

Los resultados de expresión para este gen, mostraron que sus niveles de expresión fueron reducidos significativamente para la línea transgénica parental *neo*, mientras que la línea con el L1 *naturalizado* se observa que ha recuperado su expresión mostrando niveles semejantes a los observados en las células silvestres (figura 3-30C).

Se cree que las inserciones del L1 de copia completa o cercanas antisentido pueden generar transcritos truncados del gen diana por medio de señales de poliadenilación críptica, además estas inhibiciones han sido relacionadas con la longitud de las inserciones¹⁴², lo que en principio podría explicar la atenuación observada en el caso de la línea transgénica. Teniendo en cuenta que la línea derivada naturalizada recobra los niveles basales, la disminución en la expresión del gen podría deberse a efectos netamente a causados por la presencia de la secuencia del case reportero o simplemente debido a la longitud de la inserción.

Análisis de la expresión del gen RAD51B

Teniendo en cuenta que la inserción del L1 uno se encontró ubicada entre los exones 10 y 11, el nivel de expresión para este gen fue evaluado con una pareja de primers ubicada aguas arriba de la inserción, amplificando una región entre los exones 1 y 2 (figura 3-31A y B).

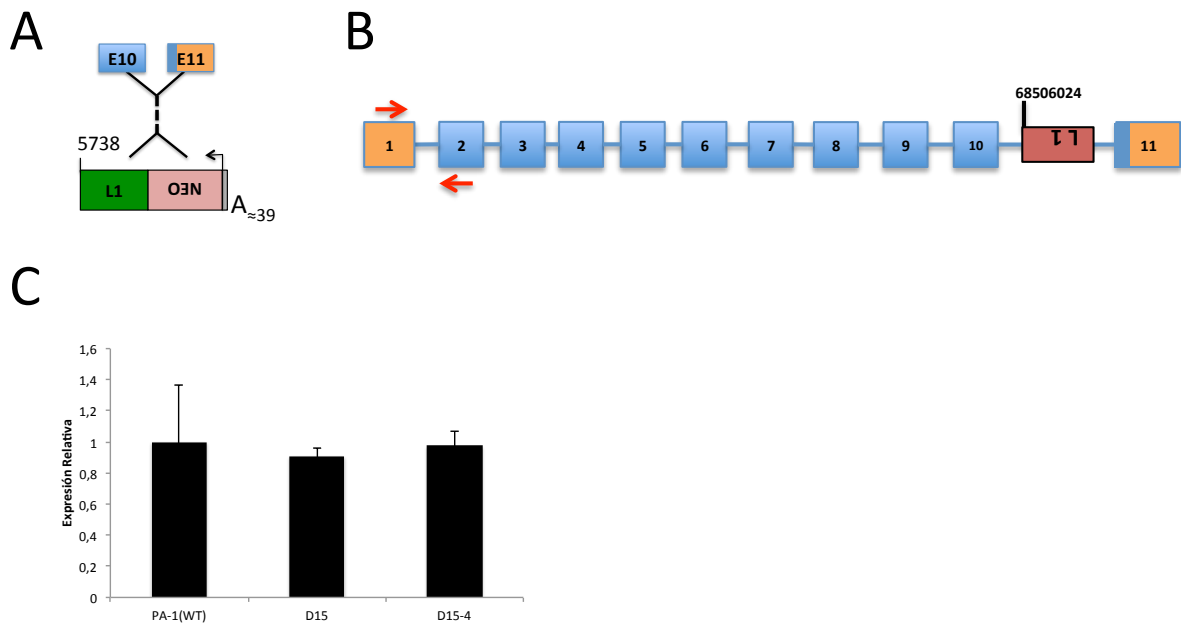


Figura 3-31. Análisis de la expresión del gen RAD51B en la línea celular transgénica pRAM-PA1D_15. A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1

L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C) Niveles de expresión del gen por RT-qPCR.

En la figura 3-31 C se puede observar que los niveles de expresión del gen RAD51B no se vieron alterados aparentemente por la inserción del L1 dentro de su estructura molecular, mostrando valores similares en ambos casos a los mostrados en las células silvestres. Los resultados obtenidos son consistentes con la idea hasta ahora aceptada de que las inserciones antisentido son menos deletéreas para las células comparadas con las que ocurren en dirección sentido¹³³. Igualmente la distancia entre los exones evaluados y el sitio de inserción pueda hacer que tengamos un falso positivo en el caso de que las señales de poliadenilación presentes en el L1 orientado en dirección antisentido y causantes de disminuciones en la expresión de los transcritos¹³⁰, esté causando posibles truncaciones en exones mas cercanos a la inserción.

Análisis de la expresión del gen DDAH

La inserción del L1 en el gen DDAH fue identificada en el intrón 2 entre los exones 3 y 3 en una región UTR (figura 3-32 A). La expresión para este gen solo se pudo obtener mediante el uso de los primers diseñados para la región 5' UTR del gen como se indica en la figura 3-32B.

Al comparar la expresión del gen en la línea transgénica que contiene del casete reportero *neo* con sus similares silvestres, se observó que el nivel de expresión en general no varió, indicando que ni el evento de RTN en si, ni la presencia del casete reportero en su estructura molecular influyo en la determinación de los niveles de expresión del gen (figura 3-32C)

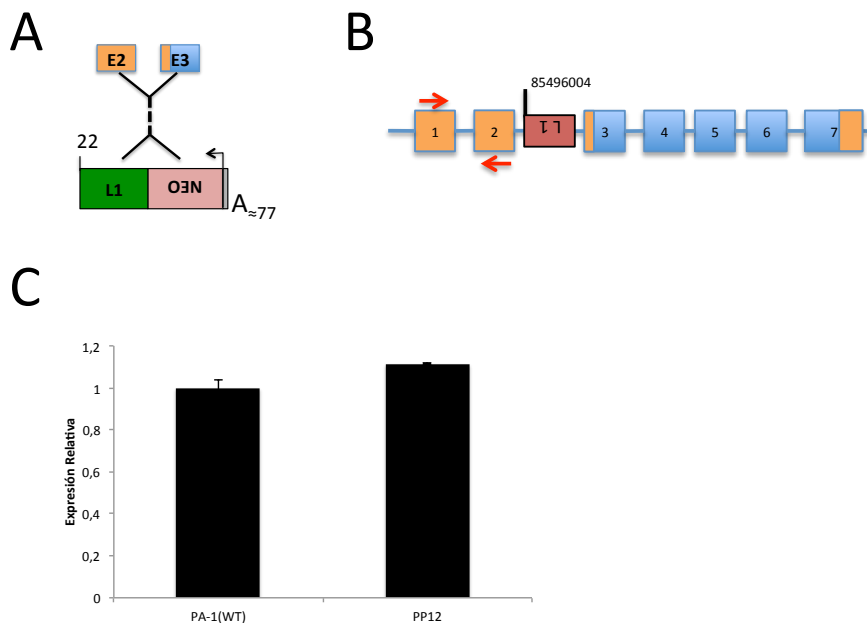


Figura 3-32. Análisis de la expresión del gen DDAH en la línea celular transgénica pRAM-PA1_12. A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del

L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C) Niveles de expresión del gen por RT-qPCR.

Teniendo en cuenta la longitud de la inserción y su orientación podría esperarse que hubiese una disminución de los niveles del gen causada por posibles truncaciones de los transcritos debido al uso de señales de poliadenilación alternativas presentes en las secuencias del L1¹³⁰. Los oligos usados para la evaluación, en este caso no permitieron descartar dicha opción, al no tener un patrón de evaluación aguas debajo de la inserción.

Análisis de la expresión del gen MAPK10

Teniendo en cuenta que la inserción del L1 en el gen MAPK10 tuvo lugar entre los exones 1 y 2 (figura 3-33 A), el análisis de la expresión del gen se realizó inicialmente con una pareja de primers que amplificaban la región de intersección de la inserción entre los exones 1 y 2 (figura 3-33B). Para la identificación de la expresión aguas abajo de la inserción se diseñaron los primers de tal forma que alineaban en los límites de los exones 5 y 6 del gen (figura 3-33B).

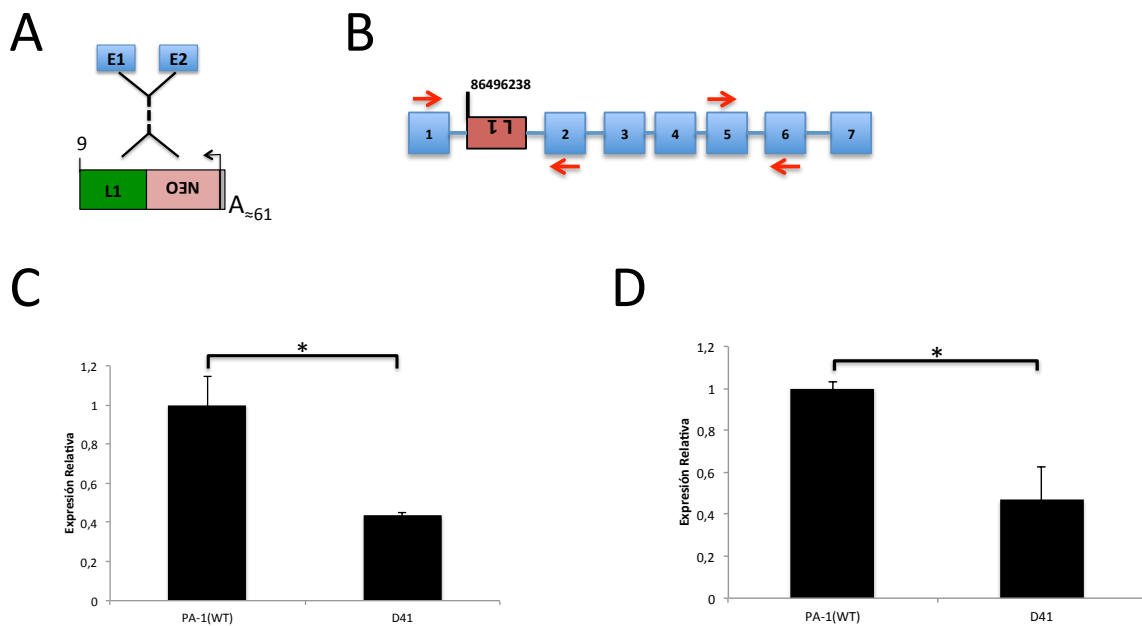


Figura 3-33. Análisis de la expresión del gen MAPK10 en la línea celular transgénica pRAM-PA1D_41. A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C y D) Niveles de expresión del gen por RT-qPCR.

En este caso se realizó la evaluación del gen MAPK10 solo en la línea transgénica que presentaba la secuencia del casete *neo* en su interior. Teniendo en cuenta los dos parámetros de evaluación, se pudo determinar que la expresión de este gen se ve notablemente reducida por la presencia del L1 manipulado genéticamente (figura 3-33C y D).

Como se muestra en la figura 3-33B la inserción del L1 se encuentra en dirección antisentido y con una longitud casi cercana a la copia completa, por lo que los resultados podría atribuirse al un posible efecto de interferencia causado por el uso de señales alternativas de poliadenilación presentes en el L1, a falta del análisis de una línea derivada con un L1 naturalizado no se puede afirmar que el efecto en la disminución sea debido al simple uso del casete reportero *neo*, al uso de señales de poliadenilación o una combinación de ambos factores.

Análisis de la expresión del gen FRAS1

La inserción del L1 se identificó entre los exones 2 y 3 (figura 3-34 A). En análisis de expresión del gen FRAS1 se hizo tomando como referencia en primera instancia los exones 1 y 2 predecesores a la inserción (figura 3-34B). La segunda pareja de primers alinea en la región consecutiva a la inserción, entre los exones 3 y 4 respectivamente (figura 3-34B).

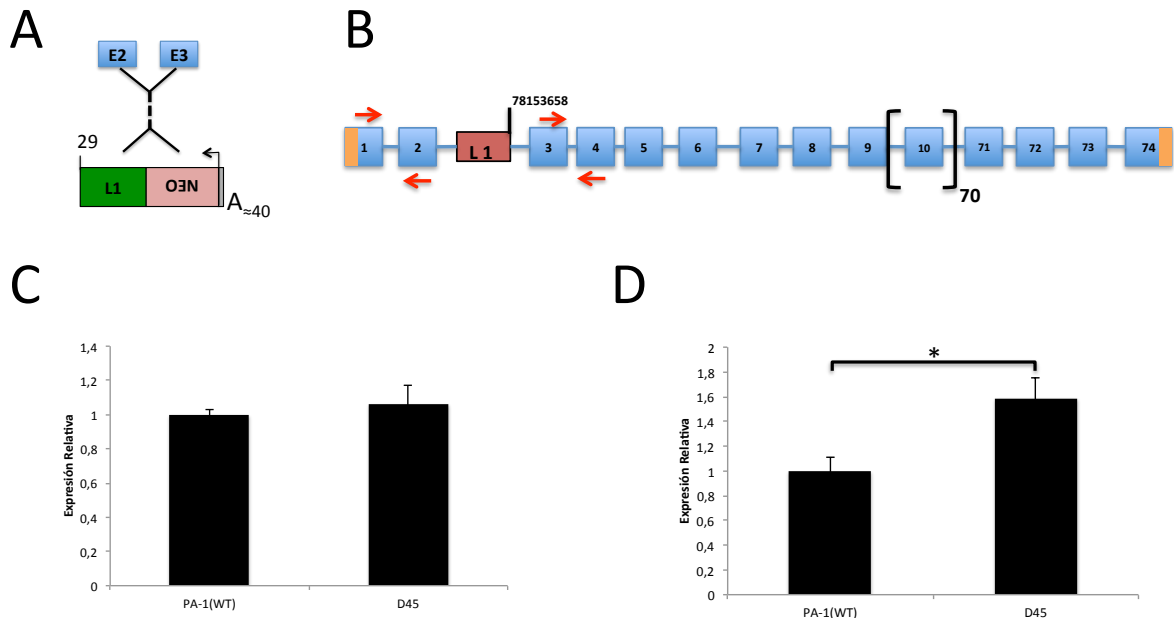


Figura 3-34. Análisis de la expresión del gen FRAS1 en la línea celular transgénica pRAM-PA1D_45. A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C y D) Niveles de expresión del gen por RT-qPCR.

Usando el primer parámetro de evaluación, se pudo observar que no hubo variación en los niveles de expresión del gen para la línea transgénica *neo* (figura 3-34C), aunque al mirar los niveles de expresión de la misma línea con el otro parámetro de evaluación (primer aguas abajo) se observa en este caso que si hubo un incremento notable de los niveles de expresión del gen FRAS1 (figura 3-34D).

Comparando la longitud del L1 y la orientación de la inserción con los resultados obtenidos, es comprensible a la luz de las observaciones previas, en donde los efectos en la variación de los niveles de expresión de inserciones *sentido* son atribuidos a defectos en la elongación, de ser así, los resultados obtenidos al evaluar el transcrito en los exones 1 y 2 no tendría porque sufrir en principio variaciones en los niveles de expresión, como efectivamente es observado (figura 3-34C). Por otro lado, el aumento en la expresión del gen diana (observado en otros eventos) en este caso puede estar influenciado por la presencia del casete reportero *neo*, en tal caso podríamos pensar que dicha secuencia reportera esta interactuando de alguna manera en el ambiente del gen provocando su sobreexpresión .

Variación de la Expresión Génica Posterior al Proceso de Diferenciación Celular.

Para determinar el efecto de las inserciones *de novo* presentes en cada uno de los genes en un proceso de diferenciación, líneas celulares transgénicas PA-1 *naturalizadas* fueron diferenciadas (ver materiales y métodos) con el fin de comparar los niveles de expresión de los genes ya evaluados cuando estos se encuentran en un estado pluripotente, y cuando su genoma posee características de células diferenciadas. Para esto se tomaron los mismos patrones de evaluación mostrados anteriormente en la identificación de variación de los niveles de expresión tanto para las líneas transgénicas naturalizadas parentales como para sus derivadas diferenciadas.

Análisis de la expresión del gen CPNE4

Al evaluar el nivel de expresión del gen en la zona de intersección de la inserción del L1 (figura 3-35 A y B), se observó que a pesar de haber un aumento en la expresión del gen en la línea silvestre diferenciada, los niveles de mostrados por la línea transgénica antes y después de la diferenciación permanecieron constantes (figura 3-35C).

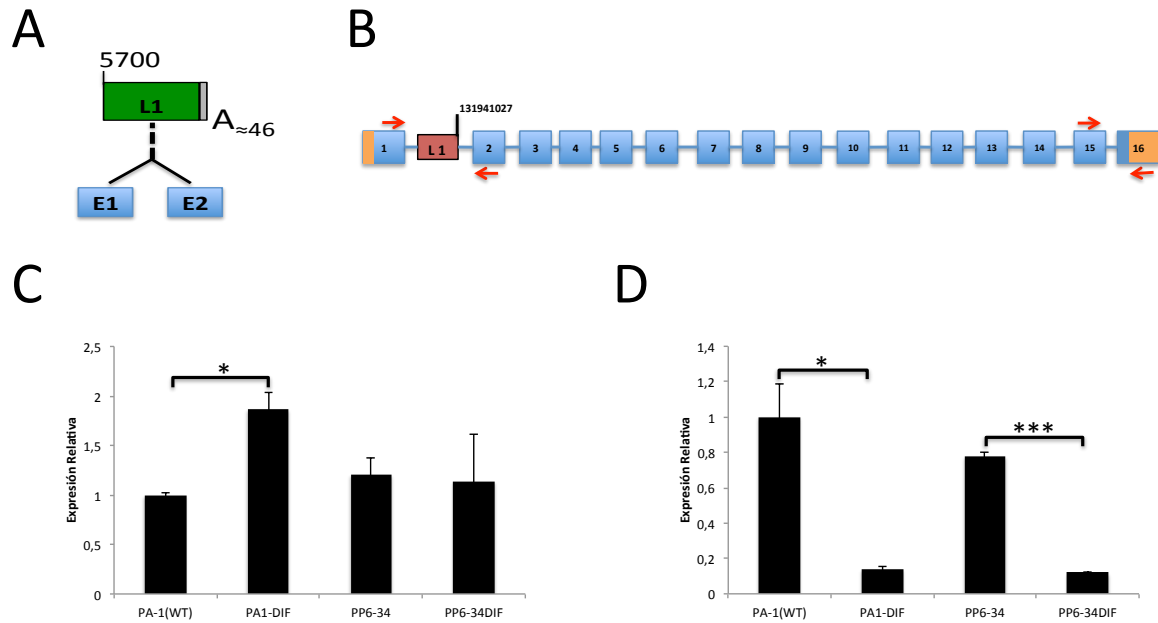


Figura 3-35. Análisis de la expresión del gen CPNE4 en el proceso de diferenciación de la línea celular transgénica pRAM-PA1_6 que contiene un L1 *naturalizado* A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C y D) Niveles de expresión del gen por RT-qPCR.

Curiosamente, cuando se realizó la evaluación de la expresión usando los exones aguas abajo de la inserción, pudimos detectar una gran disminución en la expresión una vez se ha ocurrido el proceso de diferenciación (figura 3-35D). Esta disminución en los niveles de está relacionada más con el proceso de diferenciación que al efecto que podría estar causando la inserción del L1 dentro del gen.

En general y teniendo en cuenta la posible intervención del elemento L1 en la formación de nuevos transcritos, hemos podido observar que en el proceso de diferenciación, al analizar los exones distantes a la inserción no se observa efecto del L1 sobre la expresión, aunque al analizar los exones en la zona de intersección con la inserción, hemos podido observar que el impacto del L1 ha reducido la expresión del transcrito contrario a lo esperado naturalmente (figura 3-35C y D)

Análisis de la expresión del gen TRPM3

Igual que en el anterior análisis, se uso como medida de evaluación los primeros exones UTR del gen aguas arriba de la inserción (figura 3-36A y B).

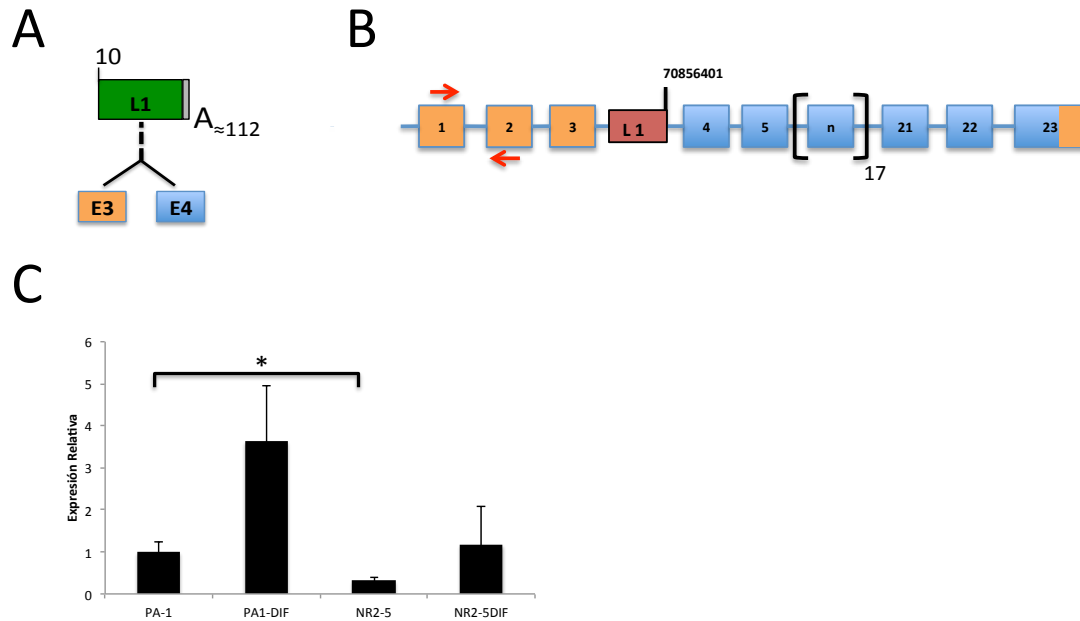


Figura 3-36. Análisis de la expresión del gen TRPM3 en el proceso de diferenciación de la línea celular transgénica pRAM-PA1NR_2 que contiene un L1 *naturalizado* A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C) Niveles de expresión del gen por RT-qPCR.

A pesar que los resultados no adquieren significancia estadística, se observa una tendencia al aumento de los niveles de expresión del gen TRPM3 una vez las células sufren un proceso de diferenciación (figura 3-36 C). Dicha tendencia igualmente no se puede atribuir a la presencia del L1 en la estructura del gen, debido a que este mismo patrón es observable a las líneas silvestres que han sufrido el mismo proceso de diferenciación (figura 3-36C).

Análisis de la expresión del gen ARHGAP42

La expresión del gen ARHGAP42 nuevamente se evaluó con respecto a los parámetros usados anteriormente, usando primers aguas arriba y aguas debajo de la inserción del L1 naturalizado (figura 3-37 A y B).

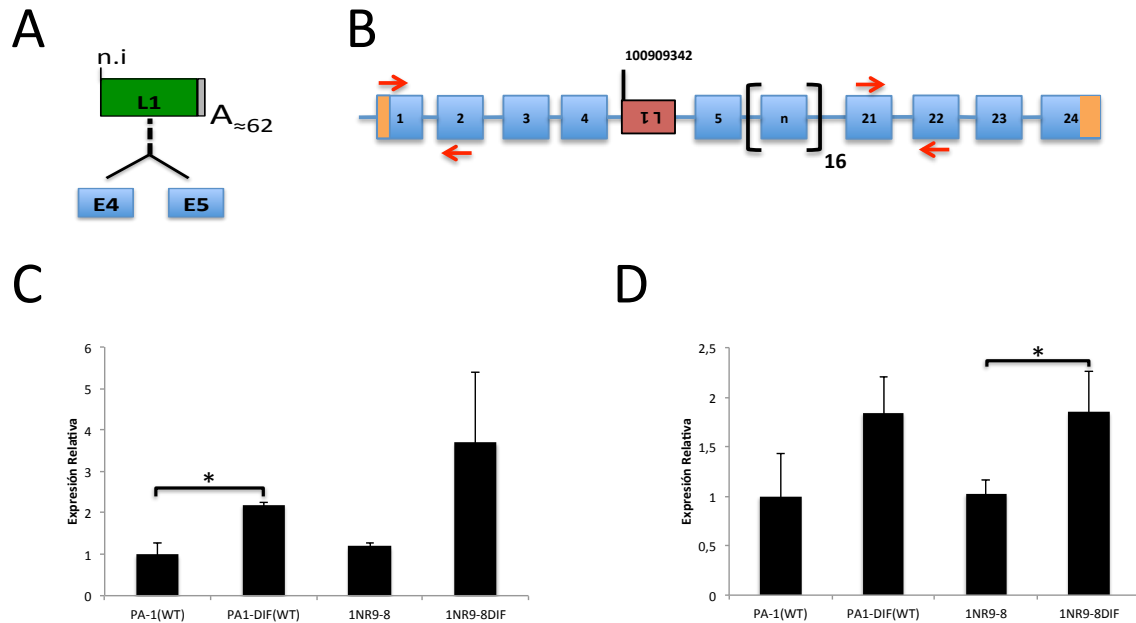


Figura 3-37. Análisis de la expresión del gen ARHGAP42 en el proceso de diferenciación de la línea celular transgénica pRAM-PA1NR_9 que contiene un L1 *naturalizado* A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C y D) Niveles de expresión del gen por RT-qPCR.

En la grafica de la figura 3-37C y D se puede observar como el gen ARHGAP42 tiende a aumentar su expresión una vez las células han sido diferenciadas. Igual que en los casos anteriores, dicha tendencia en la variación de la expresión del gen es dada mas por el proceso de diferenciación, y no puede ser atribuida a la presencia de la inserción del L1 en el gen, dado que la expresión del gen es aumentada igualmente en células diferenciadas silvestres.

Análisis de la expresión del gen FGGY

El análisis de la expresión del gen se realizó teniendo en cuenta los mismos parámetros anteriores, en donde se evaluó la expresión del gen FGGY teniendo en cuenta exones ubicados en ambos lados de la inserción del L1 naturalizado (figura 3-38A y B).

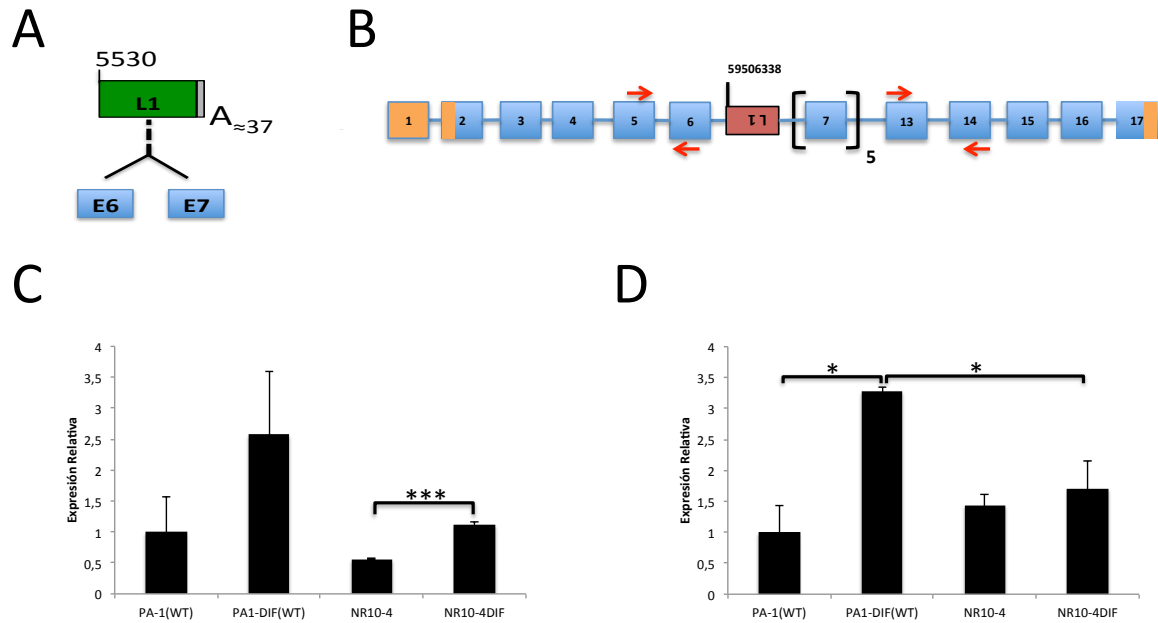


Figura 3-38. Análisis de la expresión del gen FGYY en el proceso de diferenciación de la línea celular transgénica pRAM-PA1NR_10 que contiene un L1 *naturalizado* A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C y D) Niveles de expresión del gen por RT-qPCR.

Los análisis de expresión pudieron determinar que una vez la línea con el L1 *naturalizado* dentro del gen FGYY es diferenciada, la expresión del gen se ve notablemente aumentada luego de este proceso, acorde con la tendencia observada en las células silvestres (figura 3-38 C). En este caso podemos observar el rol en el control de la expresión del gen FGYY por el L1 en el proceso de desarrollo, de tal forma que disminuye su tasa de expresión aún luego de la diferenciación cuando el gen es sobre expresado naturalmente. Si bien la diferenciación aumenta los niveles de expresión de FGYY, la presencia del L1 hace que su máximo valor sea reducido.

En la figura 3-38D podemos observar el mismo comportamiento modulador del L1 impidiendo el aumento de la expresión del gen FGYY una vez el genoma se ha diferenciado.

Análisis de la expresión del gen OSBPL6

En la figura 3-39 A y B se esquematiza la ubicación del L1 *naturalizado* y los patrones de evaluación previamente mencionados. Con el primer patrón de evaluación no se encontraron diferencias significativas en la expresión del gen en ambos estados de diferenciación (figura 3-39C). Igualmente al evaluar los exones aguas debajo de la inserción encontramos una expresión diferencial del gen acorde con la tendencia natural, por lo que en este caso no se puede atribuir a la inserción del L1 aparición la causa de la disminución de la expresión del gen (figura 3-39D).

que los valores de la expresión tendían a disminuir pero en cualquiera de los casos no se puede atribuir a la inserción del L1 aparición de esta tendencia (figura 3-39D).

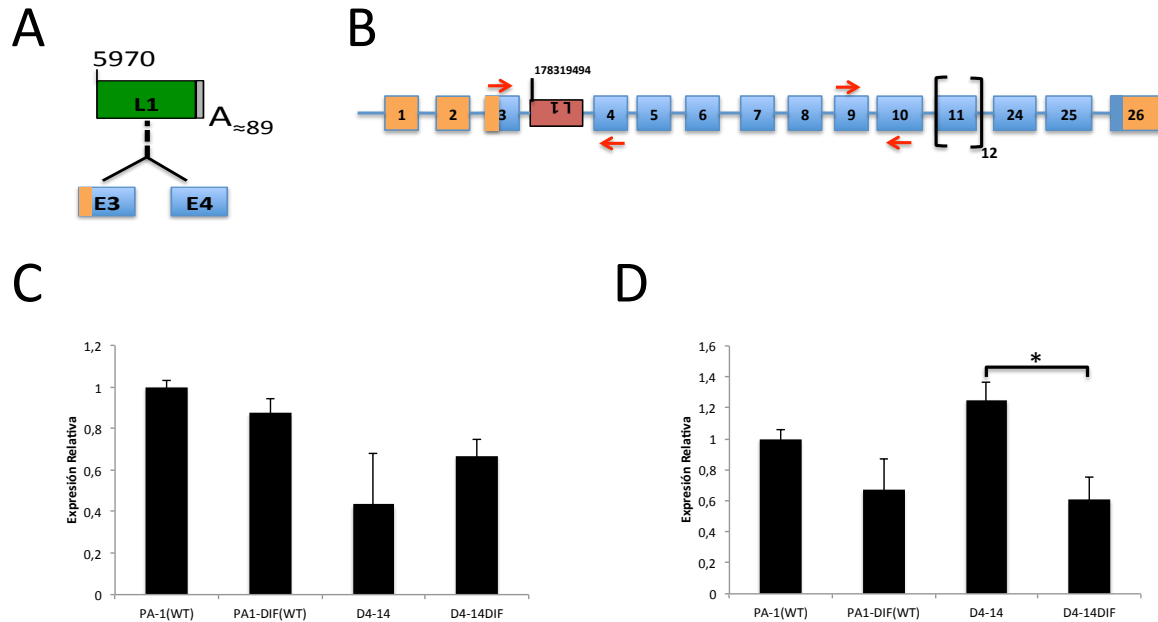


Figura 3-39. Análisis de la expresión del gen OSBPL6 en el proceso de diferenciación de la línea celular transgénica pRAM-PA1D_4 que contiene un L1 *naturalizado* A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C y D) Niveles de expresión del gen por RT-qPCR.

En el proceso de diferenciación la expresión natural del gen tiende a mantenerse estable, en este caso en particular podemos observar que la inserción del L1 no ejerce un efecto claro en el control de la expresión del gen mostrando valores de expresión que están dentro del rango de lo que se podría considerar normal en el proceso natural de diferenciación.

Análisis de la expresión del gen FAM65B

La evaluación de la expresión del gen se realizó según los parámetros previos (figura 3-40 A y B). En primera instancia, se observa que la expresión del gen FAM65B luego de la diferenciación disminuye su nivel de expresión naturalmente, (figura 3-40C). A pesar de esto, un efecto contrario fue visto en la línea transgénica que posee la inserción del L1 en la región intragenica. Los niveles de expresión del gen fueron drásticamente aumentados una vez las células fueron diferenciadas (figura 3-40C). Este efecto diferencial podría estar influenciado por la presencia del L1 de tal forma, que el L1 pueda estar ejerciendo algún tipo de presión en la desregulación de la expresión génica cuando las el genoma de las células han obtenido sus marcas genéticas y epigenéticas luego del proceso de diferenciación.

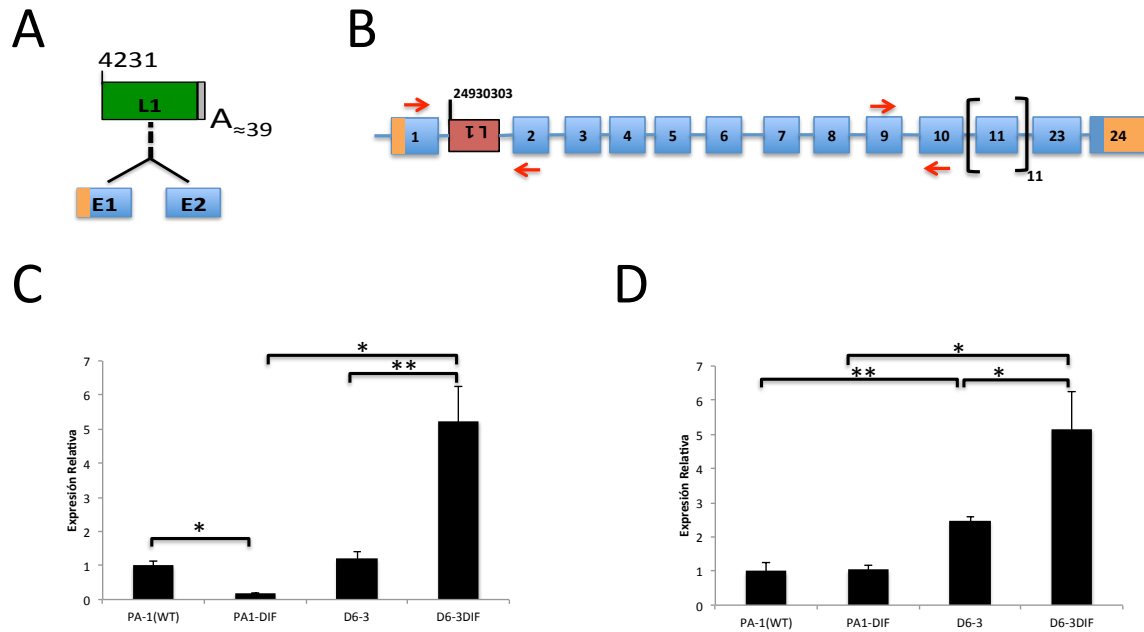


Figura 3-40. Análisis de la expresión del gen FAM65B en el proceso de diferenciación de la línea celular transgénica pRAM-PA1D_6 que contiene un L1 *naturalizado* A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C y D) Niveles de expresión del gen por RT-qPCR.

Al evaluar la expresión del gen con respecto a exones ubicados aguas abajo de la inserción, se observó un comportamiento semejante a pesar de que en las células silvestres no mostraron ningún tipo de variación (figura 3-40D). En general es evidente que la inserción del L1 esta ejerciendo un papel activador de la transcripción del gen FAM65B en el proceso de diferenciación celular.

Análisis de la expresión del gen LPHN2

El gen fue evaluado de igual forma a como se hizo en la evaluaciones previas (figura 3-41 A y B). Los resultados mostraron que los niveles de expresión del gen LPHN2 no varían dependiendo el estado de pluripotencia de sus células (figura 3-41C). De igual forma la presencia de la inserción del L1 dentro del gen, aparentemente no ejerce ningún efecto represor ó activador de la expresión del gen LPHN2 dado que no se observaron cambios en los niveles de expresión luego de la diferenciación de las líneas celulares transgénicas pluripotentes (figura 3-41C)

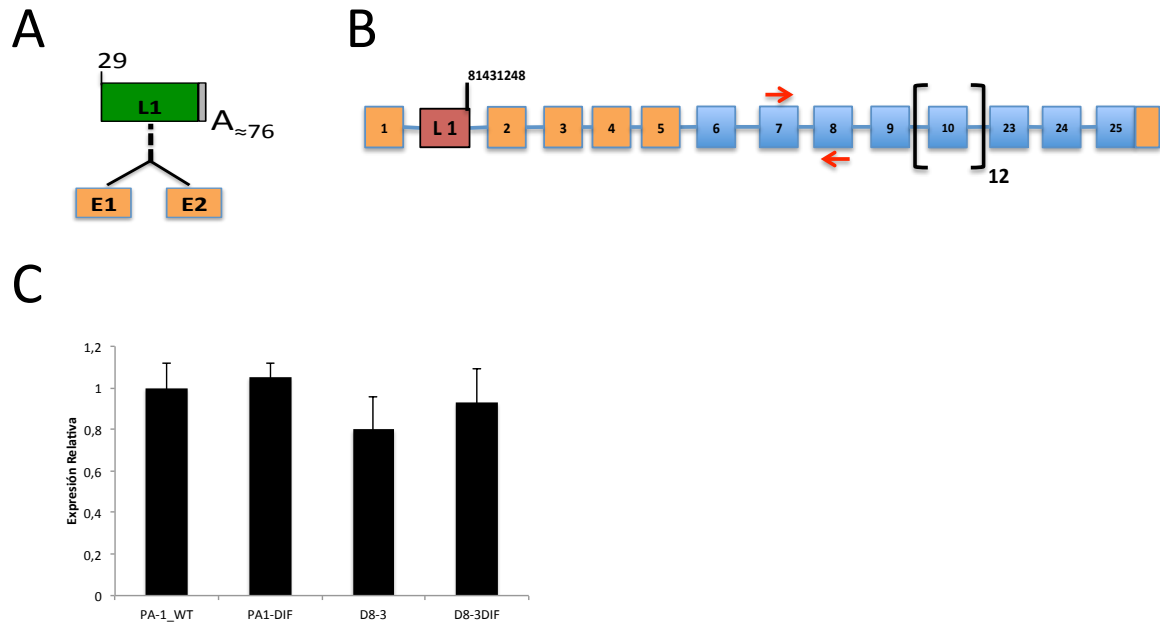


Figura 3-41. Análisis de la expresión del gen LPHN2 en el proceso de diferenciación de la línea celular transgénica pRAM-PA1D_8 que contiene un L1 *naturalizado* A) Esquema del L1 modificado genéticamente con relación a su ubicación en el gen, indicándose el grado de truncación y el tamaño de la cola de poli A. B) Mapeo intragénico del L1 donde se identifica su ubicación y su orientación frente al gen, además de la ubicación de los primers usados en la evaluación de la expresión del gen. C y D) Niveles de expresión del gen por RT-qPCR.

Capítulo 4: DISCUSIÓN

ACTIVIDAD DEL VECTOR PRAM EN CÉLULAS TUMORALES

Los resultados obtenidos demuestran que las células tumorales en este caso las células HeLa fueron capaces de soportar altas tasas de RTN cuando estas fueron transfectadas aplicando un nuevo sistema de RTN basado en el uso del vector pRAM. Dichos datos correlacionan con ensayos de RTN similares realizados en HeLa³⁸.

Estos resultados pudieron ser obtenido debido a que la células adquirieron el gen de resistencia (*neo*) una vez ocurrido un evento de retrotransposición, así que después del tiempo de selección, aquellas células que hayan podido splicear el intrón que interrumpe la codificación del gen de resistencia debido a la transcripción del L1 pudieron sobrevivir³⁸.

USO DEL PRAM COMO SISTEMA GENERADOS DE LÍNEAS CLONALES

Las células resistentes a la selección fueron susceptibles de ser aisladas y expandidas con el fin de tener líneas celulares transgénicas clonales, lo cual representaron eventos de RTN diferentes.

El uso de este nuevo sistema nos permitió rescatar las inserciones por medio del sistema de rescate en plásmido *Recovery* (Pulgarín et al; 2015 *sometido*)²⁰⁹ y hacer la caracterización de los eventos de inserción a nivel de nucleótido. De esta manera se pudo determinar que la vía mayoritaria de RTN, cuando se usa el sistema pRAM, es la vía dependiente de endonucleasa (EN), semejante a las observaciones hechas con los sistemas convencionales anteriormente usados para la realización de ensayos de RTN en células HeLa^{38; 64; 214; 215}.

Igualmente se pudo determinar que todas modificaciones genéticas realizadas en el pRAM seguían presentes, permitiendo de esta manera la manipulación genética de las inserciones seleccionadas en busca de la obtención de inserciones del L1 “naturalizadas”.

ACTIVIDAD DEL VECTOR PRAM EN CÉLULAS PA-1

Ensayos de movilidad del L1 previos realizados en PA-1 habían mostrado ya que aproximadamente el 90% de las inserciones soportadas eran silenciadas por mecanismos epigenéticos¹¹⁰. Con el 10% restante, nosotros fuimos capaces de generar una batería de líneas celulares resistentes transgénicas, es decir, líneas celulares con presencia de una inserción de un L1 modificado genéticamente el cual eran resistentes al antibiótico G418.

Con respecto a la total eficiencia de retrotransposición de nuestro plásmido aun no está del todo claro debido a que no se pudo determinar si la proporción de eventos silenciados era semejante por lo reportado previamente.

. Para determinar el nivel de silenciamiento se trataron de generar líneas celulares transgénicas por medio del sistema pRAM sin selección de antibiótico con el fin de determinar el porcentaje de líneas generadas con eventos de RTN silenciados con respecto a los activos. Dichas líneas celulares fueron definitivamente imposibles de conseguir, posiblemente debido a que los porcentajes obtenidos en las eficiencias de transfección con el pRAM nunca superaron el 30%, lo que aumenta la probabilidad de que una célula no pueda ser transfectada, además se sabe que el porcentaje de RTN representa una pequeña fracción de las células transfectadas³⁸, lo que teóricamente disminuye drásticamente la probabilidad de aislar células individuales con eventos de retrotransposición. De las alrededor de 600 líneas generadas, solo se obtuvieron aproximadamente 7 inserciones las cuales no se pudieron identificar claramente si eran eventos silenciados o no (Datos no mostrados).

Una vez generada la batería de líneas celulares transgénicas, histológicamente su comportamiento y dinámica celular era similar a las células silvestres predecesoras (PA-1 WT). Por lo que los protocolos de mantenimiento y cultivo celular fueron exactamente los mismos que los usados con las PA-1 WT.

Amplificaciones del casete de *neo* por PCR demostraron inicialmente que al igual que en HeLa, las líneas celulares clonales aisladas presentaban la secuencia del casete reportero con ausencia del intrón spliceado en el proceso de RTN, hecho que les confería resistencia al antibiótico G418, y de esta manera generar líneas celulares transgénicas clonales con inserciones de un L1 (LRE3) en un genoma pluripotente.

SITIO DE CORTE NO CARACTERÍSTICO DE LA EN DEL L1

A pesar de que en algunas líneas no se pudo determinar la su región 5', el extremo 3' fue identificada en la totalidad de las líneas celulares generadas, lo que nos permitió identificar el locus exacto si se tiene en cuenta que la reversotranscripción inicia precisamente a partir de esta región.

Con respecto al sitio de reconocimiento de la endonucleasa en las líneas celulares transgénicas, se pudo determinar que las secuencias consenso degeneradas usadas por la EN del L1 se asemejaban en gran medida a las propuestas por Feng et al (1996) y Gilbert et al (2005) a pesar de que la mayoría variaban en el cambio de una purina por una pirimidina (y viceversa) en algunas de las posiciones del sitio de reconocimiento, y teniendo en cuenta que los datos reportados son con base en otras líneas celulares^{214; 216}. Posibles explicaciones se han dado con el fin aclarar estos hechos pero aun sin resultados concluyentes.

LAS INSERCIONES PRODUCIDAS POR EL PRAM TIENEN UNA ALTA INCIDENCIA EN REGIONES CODIFICANTES

La identificación de los *loci* de las inserciones obtenidas para cada una de las líneas generadas, nos permitió observar que los eventos de RTN estuvieron distribuidos a lo largo de todo el genoma. Con respecto al tipo de región donde se llevaron a cabo los eventos de RTN se pudo observar que las nuevas inserciones tuvieron una alta incidencia en secuencias constitutivas de unidades génicas.

Si se tiene en cuenta que el porcentaje de ADN codificante (exones) en el genoma humano es de aproximadamente del 2%^{2; 4}, y que estos dependen en gran medida de regiones no codificantes pero que intervienen en el procesamiento de su información, como es el caso de los intrones que en su conjunto constituyen aproximadamente el 26% del genoma humano^{2; 4} existe una alta probabilidad de que un evento de RTN pueda realizarse dentro de unidades génicas, y dentro de estas los intrones son los más expuestos para sufrir este tipo de inserciones genéticas.

En los resultados obtenidos pudimos identificar una gran tendencia de impacto del L1 en secuencias de unidades codificantes encontrando que alrededor del 50% de las inserciones habían impactado dichas secuencias, y la totalidad de ellas se encontraban en regiones intrónicas, lo que representa una gran preferencia si se tiene en cuenta el porcentaje que representan los intrones en el genoma. Estos datos son comparables con los reportados previamente por Sela y colaboradores en el 2007 en donde habían encontrado que alrededor del 60% de los EMs están localizados en regiones intrónicas en el genoma¹⁴⁰. Igualmente, se ha determinado que este tipo de inserciones intrónicas han contribuido al alargamiento de la extensión de los intrones dentro de los genomas mamíferos²

L1 “NATURALIZADOS”

Las inserciones *de novo* del L1 genéticamente modificado distribuidas por todo el genoma de las células PA-1 fueron susceptibles de ser manipuladas modificando su estructura obteniendo una conformación semejante a la encontrada en la naturaleza.

Claramente esta es una gran herramienta en el estudio de la biología del elemento LINE-1 y su interacción con la célula huésped, debido a que los resultados generados a partir de estos nuevos L1 *naturalizados* disminuyeron los sesgos presentes en los análisis *in vitro* del comportamiento del L1 debido a la presencia de una gran cantidad de artefactos genéticos que en determinadas ocasiones podría estar influenciando los resultados y de esta manera enmascarando el efecto real de las nuevas inserciones.

El uso del vector pRAM/LRE3 como generador de nuevas inserciones naturalizadas nos permitió evaluar el comportamiento de la expresión de genes impactados por un evento de RTN. Además pudimos hacer la comparación

entre resultados obtenidos en presencia y en ausencia de los artefactos genéticos usados para la identificación de las inserciones y de esta manera poder extrapolar con una mayor certeza los efectos de las nuevas inserciones del L1 en unidades génicas en un ambiente natural.

EFFECTO DE LAS INSERCIONES GÉNICAS EN LA EXPRESIÓN DEL GEN.

Los elementos móviles intragénicos son comúnmente encontrados en eucariotas superiores especialmente en aquellos con genomas de gran tamaño^{130; 133; 217}. Cuando un elemento L1 se inserta dentro de un intrón, esta nueva secuencia inevitablemente empieza a hacer parte de la estructura del gen impactado y de sus transcritos. Estudios previos han podido identificar inestabilidades genéticas cuando eventos de RTN han tenido lugar en intrones de genes²¹⁸⁻²²⁰. Igualmente se ha caracterizado inserciones intrónicas sin resultados deletéreos²²¹.

Las inserciones del L1 dentro de unidades génicas interrumpen la estructura de los genes, y en ciertos casos causando silenciamiento de éstos en respuesta del hospedero ante un efecto mutagenico¹¹⁰. Estos silenciamientos pueden afectar negativamente la expresión de los genes cercanos²²². Quizás debido a esto es la razón por la que las inserciones de los EMs son generalmente aisladas de las regiones ricas en genes durante la evolución²²². Aun así, muchos de han sido capaces de escapar de este control y se encuentran dispersos dentro de los cuerpos de los genes, indicando que aquellas inserciones intragénicas han perdido o adquirido algunas propiedades que les permitieron a ellos permanecer dentro de la unidad transcripcional del gen huésped²¹⁷.

En general las inserciones generadas a partir de vector pRAM/LRE3 tuvieron una tendencia a incorporarse en orientación antisentido de las unidades génicas y aunque nuestros resultados no poseen una significancia estadística, estos datos correlacionan con la estadística realizada por Han *et al* en el 2004, donde indica un bajo contenido de secuencias de L1 en dirección sentido en genes expresados. Igualmente se planteó que las inserciones antisentido traían consigo menos efecto deletéreos para el gen, en donde se aportaron algunas evidencias que soportaban esta hipótesis.

La mayoría de las inserciones del L1 tuvieron una gran preferencia hacia los extremos 5' de los genes tanto de las inserciones ocurridas en dirección sentido como las orientadas antisentido del gen. 3 de las 4 inserciones en dirección sentido sobrepasaron los 9kb de longitud inicialmente cuando eran poseedoras del casete de resistencia *neo*. En el caso de las inserciones antisentido 2 de 8 inserciones alcanzaron longitudes semejantes. De acuerdo a lo anterior se podría pensar en la hipótesis acerca de la tendencia de las inserciones *de novo* del L1 en el genoma pluripotente de las células PA-1 impactadas en regiones intragénicas, pueden estar relacionadas con complejos de la ARN polimerasa mas estables y por tanto mas procesivos comparados con los encargados de la elongación en células¹³⁰.

Todas las inserciones intragénicas en dirección sentido analizadas produjeron un efecto en los niveles de los transcritos evaluados, de estas el 75% mostraron disminución en los niveles de los transcritos, mientras que el 25% mostro un amento de la expresión del gen, lo que hasta el momento no se había reportado para inserciones del elemento L1 en secuencias codificantes. La depleción en la expresión de los trascritos en su mayoría podría ser causada, bien sea por problemas en la elongación de los transcritos^{142; 220}, ó debido a la generación de nuevos transcritos de tal forma que en el momento de evaluar las diferentes expresiones pueden mostrar niveles normales para algunos exones y para otros no.

Tres inserciones antisentido de ocho aparentemente no evidenciaron un efecto directo con los niveles de expresión del transcrito. Se tiene conocimiento en la actualidad de que las inserciones antisentido en general son menos deletéreas que las que están ubicadas en dirección sentido^{130; 133; 142}. La primera razón y la mas sencilla de explicar esto, es simplemente que las inserciones impactaron regiones intrónicas en donde no tuvieran relación con el procesamiento del transcrito y que posteriormente fuesen eliminadas como uno proceso normal en la maduración de los ARNm. Igualmente Wheelan *et al.* en el 2005 describió una nueva forma de estructuración génica a partir de la retrotransposición antisentido de elementos L1, en el que mediante el uso de los promotores sentido y antisentido ubicados en la región 5' son capaces de generar transcritos individuales de las proteínas mediante la formación de nuevos ARNm e isoformas de las proteínas generadas, mecanismo al cual se le ha denominado *Gene-Breaking* ó ruptura de gen^{142; 146}.

Recientemente se ha podido observar en plantas un tipo de control epigenético de la retrotransposición, en donde se asocia la metilación CHG de los EM intrónicos con las transcripción propia del gen hospedero, sugiriendo un rol critico de la heterocromatina intragénica en la tolerancia del hospedador hacia las inserciones del los EMs deletéreas en unidades génicas transcripcionales²¹⁷. De esta manera se ha postulado en plantas que mecanismos epigenéticos pueden entrar a enmascarar el efecto de un elemento móvil dentro de un cuerpo génico, a pesar de que no se ha validado su correlación en mamíferos, se ha identificado grandes semejanzas en los procesos epigenéticos llevados a cabo por ambas células.

Aproximadamente el 40% de la inserciones mostraron un efecto reductor de la expresión de los genes asociados, aunque los presentes resultados no se pueden relacionar con la longitud de las inserciones debido a que solo una de ellas presentaba una longitud cercana a la copia completa. Los efectos en la disrupciones de los genes orientados antisentido con el gen, se les ha atribuido al uso de señales de poliadenilaciones alternativas presentes en las secuencias del L1, que el hospedador reconoce y de esta manera causar la truncación prematura de los transcritos¹³⁰.

Dentro de la inserciones antisentido se pudo identificar dos casos de sobre-regulación de la expresión de los transcritos semejante a lo observado en el caso de las inserciones en dirección sentido. Tanto los casos vistos

independiente de la orientación del elemento L1, son resultados que hasta el momento no han sido observado para las inserciones intragénicas del L1, ya que el concepto que se tiene es que sea cual sea orientación del L1 las observaciones siempre radican en una disminución de la producción de pre-ARN. Es bien conocido que los EMs pueden influenciar a expresión génica, ya sea alterando el estado epigenético de los alelos que los contienen^{143; 223}, proporcionando secuencias promotoras y sitios de unión a factores de transcripción, posiblemente alguno de estos factores o una combinación de estos sean los responsables del aumento inesperado de los transcritos evaluados.

La metilación de los promotores internos de un gen es específica de tejidos y así el control epigenético es un mecanismo específico de tejidos¹²⁷, de este modo se especula que la metilación de los promotores del L1 puedan estar sujetos igualmente a factores como el tipo celular donde se este dando la retrotransposición. Por otro lado, se ha observado que las poliadenilaciones alternativas dentro de los genes incrementan la densidad del transcriptoma generando múltiples isoformas de transcripción para un solo gen¹²⁷. Lo anterior se hace de especial importancia si se tiene en cuenta que la secuencia completa del L1 posee múltiples señales de poli A a través de toda su extensión¹³⁰. Igualmente se ha pensado que este proceso puede estar sujeto a regulaciones epigenéticas, aunque pocos ejemplos específicos han sido reportados¹²⁷. Igualmente se ha observado que el splicing alternativo de los exones además esta asociado con el enriquecimiento diferencial de las modificaciones específicas de la histona H3 entre tejidos ya que modulando los niveles de estas modificaciones de histonas resulta en la pérdida del splicing específico de tejido¹²⁷.

En general es fácilmente observable la relevancia de la presencia del elemento L1 dentro de unidades génicas como agente modulador de la expresión génica, ya sea por factores directos o indirectos independientemente de la orientación en que se encuentre. Posiblemente mucho de estos resultados estén relacionados con la influencia de diferentes mecanismos epigenéticos ya sea en la regulación del procesamiento alternativo de los genes o simplemente por sutiles obstrucciones del complejo de la polimerasa en regiones de heterocromáticas y de esta manera reducir los resultados de los transcritos

EFEECTO DEL CASETE REPORTERO EN LA EXPRESIÓN DE LOS GENES.

En el análisis de las expresiones de los diferentes genes pudimos determinar que tres de las inserciones ejercían un efecto diferencial únicamente cuando tenían presente dentro de su estructura las secuencias codificantes del gen de resistencia *neo*. Todas las inserciones estaban ubicadas hacia el extremo 5' de los genes asociado y tenían longitudes en un rango de 4400 hasta 9400pb. Los genes identificaron que eran modulados por el casete reportero son el LPHN2, OSBPL6 y el FAM65B, el primero con una inserción del L1 en dirección sentido y los restantes en dirección antisentido. Así que el efecto modulador

relacionado con la presencia “*artificial*” fue independiente de la orientación de las inserciones.

Curiosamente los efectos moduladores vistos no fueron los mismos para todas las muestras, ya que se observaron efectos que tendían tanto a la regulación positiva como a la negativa de los transcritos. Igualmente el tipo de efecto no se puede relacionar con la orientación de cada una de las inserciones presentes en el cuerpo del gen. Igualmente no podemos descartar que estos resultados estén relacionados con variaciones en el ambiente epigenético de tal forma que las secuencias del casete reportero estuvieran relacionados ya sea con cambios en la conformación del estado de la cromatina o del estado de metilación del ADN y de esta forma favorecer depleción o potenciadores de los niveles de expresión de los transcritos formados con una secuencia de *neo* en su interior. Con respecto a la epigenética de los elementos móviles, hasta el momento se sabe que el estado heterocromatínico de las inserciones intrónicas es muchas ocasiones determinantes para la transcripción propia de los genes asociados²¹⁷, aunque la regulación epigenética de los elementos móviles intragénicos y su impacto sobre las actividades génicas permanecen por se elucidadas²¹⁷.

PAPEL DEL ELEMENTO L1 EN LA DIFERENCIACIÓN CELULAR

Tres de los eventos de retrotransposición fueron identificados como agentes moduladores de la expresión génica en las células diferenciadas. Los genes afectados son CPNE4, FGGY y FAM65B en los que las inserciones tomaron lugar en todos los casos en el extremo 5' de ellos. Dos de las inserciones estaban presentes en dirección antisentido con respecto al gen, y una en dirección sentido. Los efectos observados inicialmente son descartados que sean proporcionales a la longitud de las inserciones debido a que el rango de los fragmentos insertados esta desde los 650pb hasta 1950pb en el caso de las inserciones antisentido y de 490pb en la inserción sentido aproximadamente.

Al igual que habíamos visto en las células pluripotentes, pudimos observar que las inserciones del L1 podían modular la expresión de los transcritos con efectos opuestos, es decir logrando reprimir la expresión del transcrito o bien influenciar el aumento de la expresión del producto de transcripción. Igualmente este tipo de efectos es independiente de la orientación de los transcritos ya que las dos inserciones encontradas en dirección antisentido mostraron efectos opuestos. Sin embargo el numero de muestras no son estadísticamente significativas para validar estas observaciones.

Interesantemente pudimos notar el aproximadamente el 40% de las inserciones que sufrieron algún tipo de efecto a raíz de las inserciones del L1 se sus regiones intrónicas en estado de pluripotencia, una vez diferenciadas no mostraron variaciones en la expresión de los transcritos. A pesar de estar basándonos en un numero de inserciones reducido, estos resultados estarían reforzando la hipótesis del control epigenético que puede ejercer las células hospedadoras pluripotentes para controlar los efectos mutagenicos que pueden

provocar las inserciones de los elementos móviles en el proceso de diferenciación.

Dos de las inserciones que presentaron efectos en la regulación negativa de los transcritos (CPNE4 y FGGY), luego de la diferenciación continuaron teniendo efectos represores de la expresión del gen aunque no de una manera totalitaria, por lo que no se tiene certeza con los presentes resultados si la magnitud de la represión llega a ser deletérea para la célula o no. Por otro lado, sorprendentemente pudimos observar que en una de las inserciones evaluadas anteriormente en las células pluripotentes el cual influía en la regulación negativa del transcrito evaluado (FAM65B), luego de la diferenciación adquirió un papel opuesto en dicha regulación influenciado la regulación positiva del transcrito. De esta manera queda en evidencia la influencia del elemento L1 en la regulación de la expresión génica en las células pluripotentes (EC), convirtiéndose así mismo en un agente moderador de los niveles de expresión en células diferenciadas, y de esta manera contribuir en el desarrollo de variabilidad somática en el que en términos evolutivos podría desencadenar en ciertos casos procesos evolutivos.

En general, los datos obtenidos por las investigaciones evidencian un nuevo papel de los ARN derivados de EMs en la biología de la células madre, que arrojan un nuevo panorama sobre la funcionalidad de lo que hasta hace poco se le denominaba "ADN basura" en nuestro genoma. De hecho, el uso de promotores de EMs por la célula para crear nuevas transcripciones, la adaptación de promotores de EMs como enhancers, la generación de nuevos sitios de unión a factores de transcripción dispersadas a través del genoma, su asociación con modificadores de la cromatina, y la clara relación con el mantenimiento de la pluripotencia en el embrión mamífero, debe desafiar la visión convencional de estos ARN derivados del genoma "basura". De hecho, gracias a todas las evidencias hasta ahora halladas, ponen en evidencia el posible rol de los EMs, brindando a nosotros, y otras especies, la capacidad de evolucionar gracias a su generación inherente de plasticidad genómica

Aunque la comprensión de la relevancia funcional de las inserciones de los L1 intrónicos requiere una evaluación mas profunda, estos datos sugieren un posible papel de los L1 intragénicos en la modulación de la expresión génica ya sea por procesos directamente transcripcionales alterando el splicing alternativo o por la activación o represión de genes a través de regulaciones epigenéticas. Igualmente hay que tener en cuenta que los estados epigenéticos de los elementos móviles intragénicos podrían modular la expresión de genes en respuesta a señales ambientales y/o estrés biótico.

CAPITULO 5: CONCLUSIONES

- Por medio del uso del sistema de rescate en plásmido *Recovery* pudimos generar una batería de líneas celulares transgénicas con inserciones de L1s modificados genéticamente dispersados alrededor de todo el genoma pluripotente de las células de carcinoma embrionario humano PA1
- Las inserciones producidas a partir del vector pRAM son susceptibles de caracterizar el sitio de inserción a nivel de nucleótido cuando se usa la metodología del *Recovery*.
- Se consiguieron generar y caracterizar cerca de 33 líneas celulares clonales a partir de células PA-1, de las cuales aproximadamente el 40% de las inserciones del L1 fueron inidentificadas en unidades codificantes.
- Mediante el uso del vector pRAM/LRE3 usado en la generación de las inserciones fue posible escindir todo el ADN “*artificial*” convencionalmente presente en los ensayos de retrotransposición *in vitro*, logrando de esta forma obtener inserciones del L1 en un estado “*naturalizado*”, imitando de esta forma y de manera controlada un ambiente natural “silvestre” en el que se encuentra normalmente una nueva inserción del L1.
- El casete de resistencia presente en las inserciones llevadas a cabo por el L1 del pRAM pudo ser removido satisfactoriamente después de la retrotransposición generando un estado “*naturalizado*” del L1. El cual se pudo demostrar con varias estrategias moleculares y celulares.
- Las variaciones en la expresión de los genes impactados por el L1 fueron diversos proponiendo panoramas individuales e intervenciones del ADN huésped que quedan por aclarar Los niveles de expresión génica en forma general eran regulados positivamente cuando las inserciones del L1 tenían presente su casete reportero.
- Se pudo determinar que la presencia de un L1 manipulado genéticamente influía de una manera diferencial en algunos resultados de la expresión génica en células pluripotentes, validando de esta forma el sesgo experimental al que están sometidos los ensayos de retrotransposición *in vitro* al usarse ADN exógeno como agente reportero.
- Se identificaron varios casos atípicos de sobre-regulación de la expresión génica tanto en inserciones sentido como antisentido con respecto al gen, en oposición a los siempre referenciados efectos

negativos sobre la expresión de los transcritos por efectos de las inserciones del L1.

- Se logró identificar varios eventos de retrotransposición como agentes moduladores de la expresión génica luego de sufrir un proceso de diferenciación, apreciándose efectos desreguladores y sobrerreguladores que igualmente fueron observados en los estados pluripotentes
- En general los datos aportados por el presente estudio dejan en evidencia la influencia del elemento L1 en la regulación de la expresión génica en las células pluripotentes (EC), convirtiéndose así mismo en un agente moderador de los niveles de expresión en células diferenciadas, y de esta forma contribuir en el desarrollo de variabilidad somática en el que en términos evolutivos podría desencadenar en ciertos casos procesos de adaptación.

PARTE B: ARTICULO ANEXO

Reprogramming triggers endogenous L1, *Alu* and SVA retrotransposition in human induced pluripotent stem cells

Sabine Klawitter^{1,†,#}, Nina V. Fuchs^{1,2,†}, Kyle R. Upton^{3,†}, Martin Muñoz-Lopez^{4,†}, Ruchi Shukla³, Jichang Wang², Marta Garcia-Canadas⁴, Cesar Lopez-Ruiz⁴, Daniel J. Gerhardt³, Attila Sebe¹, Ivana Grabundzija², Sylvia Merkert⁵, Patricia Gerdes³, J. Andres Pulgarin⁴, Anja Bock¹, Ulrike Held¹, Anett Witthuhn⁵, Alexandra Haase⁵, Balázs Sarkadi⁶, Johannes Löwer¹, Ernst J. Wolvetang⁷, Ulrich Martin⁵, Zoltán Ivics¹, Zsuzsanna Izsvák², Jose L. Garcia-Perez^{4,*}, Geoffrey J. Faulkner^{3,8,*} and Gerald G. Schumann^{1,*}

¹Division of Medical Biotechnology, Paul Ehrlich Institute, Langen, Germany.

²Max Delbrück Center for Molecular Medicine, Berlin, Germany.

³Mater Research Institute – University of Queensland, TRI Building, Woolloongabba QLD 4102, Australia.

⁴Department of Human DNA Variability, Pfizer/University of Granada and Andalusian Regional Government Center for Genomics and Oncology (GENYO), PTS Granada, 18016 Granada, Spain.

⁵Leibniz Research Laboratories for Biotechnology and Artificial Organs (LEBAO), Department of Cardiac, Thoracic, Transplantation, and Vascular Surgery; REBIRTH, Cluster of Excellence, Hannover Medical School, 30625 Hannover, Germany.

⁶Institute of Molecular Pharmacology, RSNC, Molecular Biophysics Research Group, Semmelweis University, Budapest, Hungary.

⁷Australian Institute for Bioengineering and Nanotechnology, University of Queensland, St Lucia, Queensland, Australia.

⁸Queensland Brain Institute, University of Queensland, Brisbane QLD 4072, Australia.

#Current address: Division of Inborn Metabolic Diseases, University Children's Hospital, Heidelberg, Germany

†These authors contributed equally to this work.

*Correspondence to:

Gerald G. Schumann (Gerald.Schumann@pei.de)

Geoffrey J. Faulkner (faulknergj@gmail.com)

Jose L. Garcia-Perez (jlgp@genyo.es)

Abstract

Human induced pluripotent stem cells (hiPSCs) are capable of unlimited proliferation and can differentiate *in vitro* to generate derivatives of the three primary germ layers. Genetic and epigenetic abnormalities have been reported to

occur during hiPSC derivation, including mobilization of engineered LINE-1 (L1) retrotransposons (Wissing et al., 2012). However, the incidence and functional impact of endogenous retrotransposition in hiPSCs are yet to be established. Here we applied retrotransposon capture sequencing (RC-seq) to 8 hiPSC lines and 3 human embryonic stem cell (hESC) lines, revealing endogenous L1, *Alu* and SVA mobilization during reprogramming and pluripotent cell cultivation. Surprisingly, 4/7 *de novo* L1 insertions were full-length and 6/11 retrotransposition events occurred in protein-coding genes expressed in pluripotent cells. We further demonstrated that an intronic L1 insertion in the *CADPS2* gene was acquired during hiPSC cultivation and disrupted *CADPS2* expression. These experiments elucidate endogenous retrotransposition, and its potential consequences, in hiPSCs and hESCs.

Introduction

Human induced pluripotent stem cells (hiPSCs) hold substantial promise for biomedical applications and as *in vitro* models of disease and development. Unlike human embryonic stem cells (hESCs), hiPSCs are a potential source of autologous cells compatible with the immune system of transplant recipients¹. hiPSCs also circumvent ethical issues associated with the use of human embryos¹. However, genetic and epigenetic aberrations that occur during reprogramming and expansion *in vitro*²⁻⁶ may hinder the use of hiPSCs in regenerative medicine due to, for instance, an elevated risk of tumorigenesis upon implantation⁷. Thus, identifying the full spectrum of aberrant mutational processes occurring in the hiPSC genome, and their functional consequences, is of paramount significance.

LINE-1 (L1) retrotransposons (Fig. 1a) are mobile genetic elements remaining active in nearly all mammals⁸. In humans, 500,000 L1 copies contribute 17% of the genome, though only 80-100 L1s per individual remain transposition competent⁹⁻¹¹. L1 mobilization is thought to primarily occur in germ cells and during early embryonic development and, together with L1-mediated *Alu* and SVA retrotransposition, has caused widespread genome structural variation in human populations^{10,12-14}. *De novo* retrotransposition events can profoundly alter gene structure, expression and function, and drive pathogenesis¹⁵⁻¹⁷. Several intracellular defense mechanisms have consequently evolved to limit L1 mobility, including histone modifications and DNA methylation^{8,18}.

Nonetheless, epigenome-wide remodeling¹⁹ coincident with reprogramming appears to enable L1 promoter hypomethylation and transcriptional activation in hiPSCs^{20,21}. hiPSCs and hESCs also support low level retrotransposition of an engineered L1 reporter^{13,20,22}. These observations indicate that the molecular machinery and substrates required for L1 retrotransposition exist in pluripotent cells. However, genomic analyses of mouse and human derived iPSC populations have to date not identified endogenous L1 mobilization events^{23,24}. It is therefore unclear whether endogenous L1-mediated mobilization occurs during reprogramming or hiPSC cultivation and, as a result, the potential significance of L1 insertional mutagenesis in hiPSCs remains unresolved. Here, we further describe the dynamics of L1 expression associated with reprogramming, elucidate L1, *Alu* and SVA mobilization in hiPSCs, and use an exemplar *de novo* L1 insertion in *CADPS2* to demonstrate the potential impact of endogenous retrotransposition in pluripotent cells.

Results

Dynamic L1 activity in hiPSCs

To elucidate endogenous L1 mobilization associated with hiPSC reprogramming, we first assembled a panel of 8 hiPSC lines and matched parental cells. Briefly, hiPSCs were derived from human fibroblasts and cord blood-derived endothelial cells (hCBECs) using several combinations of reprogramming factors, as well as integrating and non-integrating delivery systems (Table 1). Extensive characterization of these lines is described elsewhere²⁵⁻²⁷ or, as for hFF-iPS4 and hiPS-SB4, was performed here to confirm differentiation potential and expression of pluripotency markers (Supplementary Fig. S1 and S2). Noting that genomic aberrations observed in hiPSCs may occur in small parental cell subpopulations and only rise to prominence after hiPSC cultivation²⁸, we ensured that each hiPSC line used in this study was reprogrammed from a single somatic cell. This lessened the probability that heterogeneous genomic variants in parental cells could be erroneously called as *de novo* in descendant hiPSCs. As additional controls, we used 3 hESC lines as benchmarks of L1 expression and pluripotency (Table 1).

Transcription and translation of functional L1 elements are prerequisites for L1-mediated retrotransposition. To confirm that reports of pronounced L1 expression in hiPSCs by Wissing et al.²⁰ could be extended to the hiPSC lines used here, we measured L1 mRNA abundance, L1 promoter methylation status and L1 ORF1 protein (ORF1p) expression in fibroblast and hCBEC-derived hiPSCs (hFF-iPS4, hiPS-SB4, hiPS-SB5, hCBiPS1, hCBiPS2) and their parental cells (Table 1, Fig. 1b-d). TaqMan qRT-PCR targeting the L1 5'UTR (Fig. 1a, Supplementary Table S1) revealed significantly elevated L1 mRNA levels in each hiPSC line relative to their parental cells ($p < 0.05$ - $p < 0.0001$, ANOVA), that peaked in earlier passages of cell lines hiPS-SB4 and hiPS-SB5 (Fig. 1b)^{20,21}. Northern-blot analyses with an L1 5'UTR-specific probe (Fig. 1a) confirmed elevated expression of full-length L1 transcripts in hiPSCs (Fig. 1c). Notably, extended hiPSC culture led to reduced L1 mRNA abundance (Fig. 1b, left panel; hiPS-SB4, hiPS-SB5; $p < 0.05$ - $p < 0.001$, ANOVA) and resembled levels observed in hESCs (HES-3, Fig. 1b). Bisulfite DNA sequencing of the CpG island present in the canonical L1 promoter revealed strong hypomethylation in all tested hiPSC lines compared to parental cells ($p_{1,2} < 2 \times 10^{-6}$, Fig. 1d; $p_1 = 2.6 \times 10^{-12}$, $p_2 = 1.8 \times 10^{-5}$, Supplementary Fig. S3; chi-square test). Consistently, L1 ORF1p was abundant in hiPSCs, based on immunoblot (Fig. 1e) and immunofluorescence assays (Fig. 2, Supplementary Fig. S4). In agreement with previous observations in human tumors and cancer cell lines²⁹⁻³¹, in hiPSCs we found L1 ORF1p predominantly expressed in cytoplasmic foci (Fig. 2b), though unlike recent studies focused on other cell types^{29,32,33}, we did not resolve whether L1 ORF1p was directed to stress granules in hiPSCs. Finally, quantitative immunoblot analyses revealed a 10-fold increase in L1 ORF1p expression in hiPSCs when compared with parental cells (Supplementary Fig. S5).

Taken together, our results revealed a spike in L1 expression during or immediately after reprogramming, confirming previous findings^{20,21}, followed by attenuation in later hiPSC passages (Fig. 1b,c). To extend these results, we measured L1 mRNA levels upon differentiation of late passage hiPSCs (hiPS-SB4 (p98) and hFF-iPS4 (p50)) into embryoid bodies (EBs). We observed 49% and 58% reductions in L1 mRNA levels after one and ten days of EB differentiation, respectively (Fig. 1b, middle panel). A parallel assay conducted with early passage hiPSCs indicated a gradual decrease of L1 mRNA abundance by up to 65% after

eight days of EB differentiation and a concomitant increase in differentiation markers (Fig. 1b, right panel; Supplementary Fig. S6). Hence, elevated L1 expression in hiPSCs was triggered by reprogramming and attenuated by short-term cultivation, while, in turn, subsequent differentiation gradually reduced L1 expression again.

Endogenous retrotransposition occurs in pluripotent stem cells

To unambiguously determine whether activation of the L1 mobilization machinery produced L1-mediated retrotransposition, we used retrotransposon capture sequencing (RC-seq) to map the genomic integration sites of *de novo* retrotransposon insertions. Briefly, RC-seq involved liquid-phase sequence capture to enrich DNA for the 5' and 3' junctions of recent L1, *Alu* and SVA insertions and the surrounding genome³⁴. Putatively immobile long terminal repeat (LTR) retrotransposons were also probed as negative controls. Multiplexed, paired-end 150mer Illumina sequencing of RC-seq libraries, followed by contig assembly, provided high-fidelity, single nucleotide resolution of insertions absent from the reference genome, even at low read depth³⁴.

We analyzed all 8 hiPSC lines and their matched parental cells by RC-seq. For 5 fibroblast-derived hiPSC lines (Table 1), we included two separate passages each in order to detect mobilization events that may have accumulated during cell culture. Similarly, we analyzed two passages each of 3 hESC lines to evaluate endogenous retrotransposition during hESC cultivation (Table 1). RC-seq detected a total number of 40,608 non-reference retrotransposon insertions including on average 214 L1, 1,411 *Alu*, 53 SVA and 14 LTR non-reference genome insertions per hiPSC and hESC sample (Supplementary Fig. S7, Supplementary Table S3). Insertions were annotated as *de novo* in pluripotent cells if they were not i) reported previously in non-reference retrotransposon insertion databases^{9,12,34-39}, ii) found in parental cells, iii) found in an earlier hESC passage, or iv) found in multiple hiPSC or hESC lines. In total, we detected 8 L1, 7 *Alu* and 2 SVA putative *de novo* insertions (Supplementary Table S3). We found no *de novo* LTR retrotransposon insertions, despite observing profound upregulation of HERV-K group HML-2 transcription in hiPSCs and hESCs (Supplementary Fig. S8).

Five retrotransposon subfamilies (L1-Ta, L1 pre-Ta, *AluYb8*, *AluYa5* and SVA_E) known to be active in humans contributed putative *de novo* insertions^{10,11,40}. These were first validated by genotyping PCR, with 7 L1, 2 *Alu* and 1 SVA insertion confirmed as *de novo* in hiPSCs and a single *Alu* insertion (*Alu-2*) in hESCs (Fig. 3a, Supplementary Fig. S9 and S10, Supplementary Table S4). The remaining 6 putative *de novo* insertions (1 L1, 4 *Alu*, 1 SVA) were detected by PCR in parental cells or an earlier hESC passage, suggesting that these variants were present but were not *de novo*. Next, we determined the entire nucleotide sequence of 10/11 confirmed *de novo* retrotransposon insertions (Supplementary Fig. S9 and S10). For one event, SVA-2, a member of the SVA_E subfamily, we could sequence only the 3' junction, which included a poly-A tail characteristic for L1-mediated *trans* mobilization (Fig. 3a, Supplementary Fig. S9). Our efforts to PCR amplify the matching 5' junction of SVA-2 with multiple primer combinations, intended to detect a possible 5' SVA truncation or a small proximal genomic deletion, were unsuccessful (see Methods). One reasonable explanation for this outcome was the occurrence of a large 5' genomic deletion at the SVA-2 integration site, as reported previously^{13,41-43}. Additional sequence analysis revealed that 9/10 of the remaining

insertions exhibited the canonical hallmarks of L1-mediated target-primed reverse transcription (TPRT)^{8,44} including: i) a target site duplication (TSD), ii) a variable length L1 poly-A tail and iii) an integration site resembling the L1 endonuclease target motif 5'-TTTT/AA-3'^{45,46} (Fig. 3a, Supplementary Fig. S9). The one exception, insertion L1-dn4, was 3' truncated within its poly-A signal and devoid of an L1 endonuclease motif, but nevertheless incorporated an 8-bp TSD. These features were consistent with L1 endonuclease-independent retrotransposition⁴⁷. Insertions L1-dn6 and L1-dn14 presented one and two untemplated G nucleotides at their 5' ends, respectively, as seen elsewhere^{41,43}. Insertions L1-dn3, L1-dn4, L1-dn13 and L1-dn15 exhibited microcomplementarities of one to five nucleotides at their 5' end, a structural feature reported previously for L1 integration sites⁴⁸. L1-dn13 and L1-dn15 were also 5' truncated and inverted, consistent with 'twin-priming'⁴⁹, and in one instance a 5' inversion was displaced from the remaining L1 sequence by a 25-bp DNA fragment of unknown origin (L1-dn15). Thus, L1-mediated retrotransposition in hiPSCs and hESCs occurs via mechanisms described previously in mammalian cells.

The rate of L1-mediated retrotransposition occurring in pluripotent stem cells was difficult to accurately assess given the unknown genomic heterogeneity of each population. However, by estimating the sensitivity of RC-seq, we were able to determine the approximate L1 mobilization rate in hiPSCs. First, we identified that the overall RC-seq false positive rate was 1.5%, based on our recent PCR validation rate of 98.5% for insertions found by RC-seq in a cohort of hepatocellular carcinoma patients³⁴ using the same detection thresholds as used here. Next, we determined that 88.5%, 92.8%, 88.3% and 89.8% of germline L1, *Alu*, SVA and LTR insertions, respectively, found in a parental cell line or early hESC passage were also detected in the matched hiPSC or later hESC passage, indicating an overall RC-seq false negative rate of 7.9%. To then model the sensitivity of RC-seq for *de novo* insertions, we randomly sampled each library and determined the fraction of the total germline events detected in that library as a function of sampling depth (Supplementary Fig. S11). At 50% library sampling depth (i.e. modeling 50% variant allele fraction) 71.4%, 76.2%, 68.4%, 87.3%, respectively, of the germline L1, *Alu*, SVA and LTR insertions found in hiPSC lines were detected, dropping to 5.9%, 5.8%, 7.4% and 27.1% at 5% sampling depth. The estimated overall false negative rates at 50% and 5% variant allele fraction for *de novo* insertions detected in hiPSC lines were therefore 30.5% and 94.4%, respectively. These figures were similar for hESC lines (31.5% and 94.1%). Thus, we concluded that although RC-seq reliably detected high variant allele fraction retrotransposon insertions, a large pool of low variant allele fraction events may have been overlooked at the RC-seq thresholds used here. This would be particularly acute in the chosen hESC lines where, unlike iPSCs, cells had not undergone a recent population bottleneck *in vitro*. Using these parameters and the observed *de novo* L1 insertion counts, we estimated that hiPSC lines carried 3.7 *de novo* L1 insertions with allele frequencies $\geq 5\%$, on average, extrapolating to ~ 1 *de novo* L1 insertion per cell (see Methods). However, the low number of insertions identified precluded similar estimates for hESC lines.

hiPSC cultivation leads to individual L1 copy number variation

Our qualitative L1 insertion site validation PCR experiments (Fig. 3a) indicated that some *de novo* L1 insertions detected by RC-seq were absent from the earlier

hiPSC passage surveyed and therefore may have arisen after reprogramming. To better establish the temporal dynamics of L1 retrotransposition in hiPSCs, we performed multiplex TaqMan quantitative PCR (qPCR) incorporating a 5' junction-spanning probe (Fig. 3b) to quantify L1-dn13 and L1-dn14 copy number variation in hiPS-SB4. We observed an 8-fold increase in L1-dn13 copy number upon extended cultivation (Fig. 3b, left panel) and a ~2-fold decrease in L1-dn14 copy number (Fig. 3b, right panel), indicating the presence of two different hiPS-SB4 subpopulations carrying insertions L1-dn13 or L1-dn14, respectively, with opposite growth dynamics. L1-dn13 and L1-dn14 were not detected in hESCs (HES-3) or the parental fibroblast (HFF-1) population, again showing that L1-dn13 and L1-dn14 were *de novo* insertions. As the hiPS-SB4 line was cultivated from a single cell-derived hiPSC clone, these data showed that either one or both of these insertions occurred during or after reprogramming, confirming our RC-seq and genotyping PCR data. To discriminate whether L1-dn13 arose during hiPSC reprogramming or cultivation, we thawed and extensively cultivated a passage (p19) of hiPS-SB4 isolated well before the later passages analyzed by RC-seq (p43 and p53) (Fig. 3c). L1-dn13 was not detected by qPCR in this second hiPS-SB4 cultivar (Fig. 3d). Hence, L1-dn13 likely arose in the original hiPS-SB4 cultivar between p19 and p43. We concluded that cultivation of hiPSCs, and hESCs, as described above for the *Alu-2* insertion (Fig. 3a), can lead to endogenous retrotransposition.

***De novo* L1 insertions can retain retrotransposition competency**

Intriguingly, 4/7 *de novo* L1 insertions were full-length, a surprising result given that most preexisting genomic L1 retrotransposition events are 5' truncated⁵⁰. Indeed, only ~15% of L1 copies in the reference genome and <1% of somatic L1 insertions found thus far in tumors are full-length^{34,51,52}. PCR amplification and sequencing of three full-length L1s (L1-dn4, L1-dn6, L1-dn14) revealed no deleterious nonsense mutations in their ORFs (Supplementary Fig. S10), suggesting each insertion likely retained retrotransposition competency. As a proof-of-principle, we used an established cell-culture based L1 retrotransposition reporter assay⁵³ to evaluate the mobility of L1-dn6 in HeLa cells. L1-dn6 subclones retrotransposed at a relative efficiency of 20-30% of that obtained for the benchmark L1.3⁵⁴ element (Fig. 4) and were therefore classified as highly active or 'hot'^{9,11}. These data indicated that new, full-length L1 insertions in hiPSCs could retain substantial competence in initiating further rounds of mobilization.

L1 insertional mutagenesis disrupts *CADPS2* expression

Six *de novo* retrotransposition events mapped to introns of protein-coding genes. These included key factors in neuron (*CADPS2*, *NREP*) and nephron (*SLC12A1*) biology, as well as genes with established and predicted roles in cell cycle regulation and oncogenesis (*PTPN9*, *RNF38*, *PLXDC2*). Insertions showed a marked bias for the 5' end of genes, with insertions falling on average in the 20th percentile of gene length measured from the annotated RefSeq transcription start site (TSS), a significant deviation from random expectation (P<0.006, permutation test). Albeit based on a small sample of insertions, this outcome could be explained by L1 endonuclease preference for open chromatin⁵⁵ and increased chromatin accessibility around TSSs⁵⁶.

Given that intronic L1 insertions can disrupt host gene transcription^{8,15,57}, we

noted with interest that all six genes were expressed in hiPSCs and hESCs⁵⁸. For example, L1-dn13 occurred in an intron of *CADPS2* and, as noted above, exhibited copy number variation during hiPSC cultivation (Fig. 3b, d). This afforded us an opportunity to analyze differential *CADPS2* expression with reference to L1-dn13 copy number. First, we measured and compared *CADPS2* mRNA expression in early versus late hiPS-SB4 passages via TaqMan qRT-PCR (Fig. 5a) and observed a 5-fold reduction in *CADPS2* expression in the latter cells (Fig. 5b). Importantly, this assay tested *CADPS2* expression at an exon junction located downstream of the L1-dn13 integration site (Fig. 5a) and indicated opposing changes in L1-dn13 copy number (Fig. 3b, left panel) and *CADPS2* expression for hiPS-SB4 cells in culture, suggesting that L1-dn13 interfered with *CADPS2* expression.

To further test this possibility, we employed a human triose phosphate isomerase (TPI)/Renilla luciferase reporter assay developed to monitor the effects of different introns on mammalian gene expression⁵⁹. We generated three constructs (Supplementary Fig. S12a,b) respectively containing: i) 825bp spanning the empty L1-dn13 target intron of *CADPS2* (pSHM06_01), ii) 423bp spanning the same region but in this case flanking the 389bp L1-dn13 insertion and its TSDs to produce a 825bp sequence (pSHM06_02) and iii) the 423bp sequence on its own (pSHM06_03). We cloned each of these fragments into the TPI/Renilla reporter cassette and quantified their effect on luciferase activity (Supplementary Fig. S12c). Interestingly, the *CADPS2* intron sequence harboring L1-dn13 (pSHM06_03) had the strongest inhibitory effect and reduced luciferase activity by 62%, a significant decrease beyond the constructs lacking L1-dn13 ($p=0.022$).

As further corollary, we isolated two clones from the original hiPS-SB4 culture by single cell-cloning (see Methods) where L1-dn13 was identified by RC-seq, one carrying L1-dn13 (hiPS-SB4_D) and the other not carrying L1-dn13 (hiPS-SB4_B) (Fig. 5c). The identity of each clone was verified by genotyping PCR (Fig. 5d) and qPCR (Fig. 5e). qRT-PCR applied to cytoplasmic RNA extracted from each clone indicated that *CADPS2* expression was ~95% lower in hiPS-SB4_D than in hiPS-SB4_B (Fig. 5f). Consistently, *CADPS2* expression in the original hiPS-SB4 culture, which was heterogeneous for the L1-dn13 allele, was in between expression levels observed for the hiPS-SB4_D and hiPS-SB4_B clones. We then employed endpoint quantitative RT-PCR⁶⁰ with subsequent capillary electrophoresis to compare the relative expression of each *CADPS2* allele in hiPS-SB4_D (Supplementary Fig. S13), as distinguished by a single nucleotide polymorphism located in the 3'UTR of *CADPS2*. Notably, L1-dn13 was associated with complete silencing of the *CADPS2* mutant allele in hiPS-SB4_D while, interestingly, the *CADPS2* wild-type allele was also downregulated by >90% relative to hiPS-SB4_B. Again, expression of each *CADPS2* allele in the hiPS-SB4 culture heterogeneous for L1-dn13 lay between levels observed for the hiPS-SB4_B and hiPS-SB4_D subclones. Altogether, these results conclusively indicate that L1-dn13 interfered with *CADPS2* expression.

Discussion

Here we have demonstrated that endogenous L1-mediated retrotransposition can occur in hiPSCs and hESCs, building upon earlier reports of engineered L1 retrotransposition in stem cells^{13,20,22,61}. By contrast, two previous studies reported an absence of endogenous retrotransposition events in mouse or human iPSCs^{23,24}. A more recent study reported low level L1 mobilization in hiPSCs⁶², though in this

case no insertions could be confirmed by PCR, leaving the validity of the reported putative L1 insertions unclear. We unequivocally demonstrated here through RC-seq, gold-standard PCR validation and capillary sequencing, including L1 integration site structural characterization, that fibroblast-derived hiPSCs clearly can support the mobilization of endogenous non-LTR retrotransposons. We speculate that our use of clonally derived hiPSCs, and the robustness of RC-seq in detecting somatic L1 insertions^{34,35}, enabled us to discover retrotransposition events that may have otherwise remained undetected.

We estimated that hiPSCs each carried ~1 *de novo* L1 insertion, with the notable caveat that this calculation was based on a small number of observed events. Nonetheless, this is a much lower rate than recently found for human hippocampal neurons and glia (13.7 and 6.5 somatic L1 insertions per cell, respectively)⁶³. Our sensitivity calculations suggested that most *de novo* insertions with a variant allele fraction of <5% in hiPSC and hESC populations were overlooked by RC-seq at the detection thresholds used here, and these were not included in the above rate estimate. This is a major consideration in concluding whether parental cell type or choice of reprogramming vector affects endogenous retrotransposition activity in hiPSCs. Notably, low frequency or subclonal retrotransposition may indeed occur in our hCBEC derived hiPSC lines, or reprogrammed via lentiviral systems, and escaped detection by RC-seq here. Therefore, we would propose that additional experiments are required to better define how these and other considerations (e.g. cultivation protocol) affect L1 activity. Indeed, one explanation for the low number of insertions characterized in hESCs is that these cell populations were not clonally derived and were therefore likely to present more extensive genomic heterogeneity than hiPSCs. The lone *Alu* insertion found here in H9 cells is nonetheless the first endogenous retrotransposition event reported in hESCs, reinforcing evidence that L1-mediated mobilization can occur in early human development^{13,64}.

L1 activity was highly dynamic during reprogramming and hiPSC cultivation. Parental cells, early hiPSC passages, later hiPSC passages, and re-differentiated cells presented grossly different levels of L1 expression. As corroborated by RC-seq, genotyping PCR and qPCR, the majority of retrotransposition in hiPSCs likely took place during or immediately after reprogramming, where we observed a peak in expression of the L1 mobilization machinery. As a result, each detected variant could affect substantial hiPSC subpopulations. Interestingly, major induction of L1 mRNA and protein expression, far in excess of that seen in hESCs and neural stem cells^{13,61}, was accompanied by a comparatively modest increase in L1 mobilization rate. Due to drastic epigenetic changes occurring upon reprogramming, it is possible that reprogramming per se may activate the expression of cellular L1 restriction factors such as APOBEC proteins^{22,65} and PIWIL2⁶⁶. Consistently, APOBEC3B and PIWIL2 have been demonstrated to control engineered L1 retrotransposition in hiPSCs^{22,66}. Thus, it is tempting to speculate that the cellular milieu of hiPSCs and hESCs may permit L1 upregulation but also limit L1-mediated mutagenesis.

That 4/7 of the *de novo* L1 insertions reported here were full-length was consistent with 2/3 of the engineered L1 *de novo* insertions resolved by Wissing et al. also being full-length²⁰. This >50% incidence of full-length L1 *de novo* insertions in hiPSCs is unexpected as only ~15% of L1 copies in the human reference genome and <1% of somatic L1 insertions identified in tumors are full-length^{34,37,51,52}.

However, 7/7 engineered L1 retrotransposition events found in hESCs were recently reported to be significantly 5' truncated¹³, suggesting that pluripotency factors common to hiPSCs and hESCs might not play any role in the observed overrepresentation of full-length *de novo* L1 insertions found in hiPSCs. The mechanism of L1 5' truncation is not fully understood. On one hand, the preponderance of 5' truncated L1 copies on the genome has long been explained by an inability of the L1 reverse transcriptase (RT) encoded by L1 ORF2p to copy the entire template L1 RNA, either due to premature dissociation of the L1 RT from its RNA or competition from an unknown cellular RNase that digests the L1 RNA before completion of reverse transcription⁶⁷. Therefore, it is possible that hiPSCs provide a nuclear environment allowing a more stable association of the L1 RT with L1 RNA, or L1 RT does not have to compete with a cellular RNase which might be differentially expressed in hiPSCs. On the other hand, a recent study demonstrated that the DNA-damage signalling protein ATM may control the length or number of *de novo* L1 insertions in human neural stem cells⁶⁸. Thus, it is possible that subtle differences in the DNA repair mechanisms operating in hiPSCs and hESCs could be related to the high frequency of full-length L1 insertions characterized in hiPSCs.

Each *de novo* L1 insertion reported here integrated in a protein-coding gene expressed in pluripotent cells. In one case, we identified an L1 insertion (L1-dn13) that arose during hiPSC cultivation and integrated into an intron of the gene *CADPS2*. It remains to be determined whether acquisition of L1-dn13, and a concurrent reduction in *CADPS2* expression, imbued carrier hiPSCs with a selective advantage *in vitro*. Furthermore, it remains unclear why transcription of the *CADPS2* allele lacking L1-dn13 was reduced by >90%. To speculate, it is possible that *CADPS2* expression involves a direct or indirect positive feedback loop where, for example, transcription from *CADPS2* reinforces open chromatin⁶⁹. A reduction in *CADPS2* expression caused by L1-dn13 could hence have a strongly negative effect on transcription from the wild-type *CADPS2* allele.

In closing, it is notable that intronic L1, *Alu* and SVA insertions can alter cellular phenotype and are associated with numerous instances of human disease⁵⁷. Future in-depth experiments are however required to definitively establish whether endogenous retrotransposition alters the phenotype of hiPSC derivatives sufficiently to impact their use in medical or research applications. We can nevertheless conclude that retrotransposition, in addition to other sources of genetic and epigenetic variation^{2-6,70}, can change the functional landscape of the hiPSC genome.

Methods

Cell lines and culture conditions

hiPSC lines hiPS-SB4 (hFF-T2-OSKM) and hiPS-SB5 (hFF-T2-OSKML)/hiPS-SB5.1 (hiPS-OSKML#6) were generated by reprogramming HFF-1 cells (ATCC-Number: SCRC-1041™) using *Sleeping Beauty* (SB) transposon-based plasmids pT2-OSKM or pT2-OSKML which contain polycistronic OSKM (Oct4, Sox2, Klf4 and c-Myc) or OSKML (OSKM + Lin28) expression cassettes²⁶. Briefly, HFF-1 cells (4 x 10⁵ cells per well) were transfected by nucleofection (Lonza) according to the manufacturer's instructions. In each transfection, 2µg of transposon plasmid (pT2-OSKM or pT2-OSKML) and 0.2µg of CMV- SB100X vector (harboring the enhanced *Sleeping Beauty* transposase gene under control of a CMV promoter²⁶) were used.

After transfection cells were plated onto Matrigel-coated 6-well plates (**hESC-qualified Matrix**, BD Biosciences) and were grown in MEF-conditioned ESC medium used for the cultivation of hESCs and hiPSCs. ESC medium consisted of Knockout™ DMEM (Life Technologies) supplemented with 4ng/ml basic fibroblast growth factor 2 (FGF2, Invitrogen), 20% Knockout™ Serum Replacement (Gibco), 1 mM L-glutamine (Biochrom AG), 50µM β-mercaptoethanol and 0.1mM non-essential amino acids. **The medium was replaced every day. Newly formed hiPSC colonies were picked, transferred to** Matrigel-coated 24-well plates, and expanded for 4-6 days in MEF-conditioned ESC medium. Subsequently, cells were trypsin-dissociated, plated onto feeder cells and cultivated in ESC medium.

In this experiment, cells nucleofected with SB-OSKM gave rise to only one hiPSC colony. Nucleofection with SB-OSKML resulted in several hiPSC colonies. Multiple SB-OSKML colonies were picked and transferred onto the same Matrigel-coated wells. After establishing a mixed culture of hiPSCs generated with either SB-OSKM or SB-OSKML, single cell-derived hiPSC-clones were generated by single-cell dilution using cell sorting (see below) based on their positivity for SSEA4. Six SB-OSKML hiPSC clones and one SB-OSKM hiPSC clone were then characterized for pluripotency and differentiation potential as described²⁶. Two SB-OSKM hiPSC clones (hiPS-SB5 and hiPS-SB5.1) and the only SB-OSKM hiPSC clone (hiPS-SB4) obtained were used in this study.

The lines hiPS-CRL1502²⁵, hiPS-CRL2429²⁵, hCBiPS1²⁷, hCBiPS2²⁷ and hiPS-FB⁷¹ have been described previously. hFF-iPS4 was produced using HFF-1 cells and a lentiviral vector expressing reprogramming factors Oct-4, Sox2, Nanog and Lin28²⁷. Successful reprogramming for the hFF-iPS4 cell line was verified by morphology, pluripotency marker expression (Supplementary Fig. S2), karyotype analysis and the ability to generate teratomas on immunocompromised mice (data not shown).

hESC lines H1, H9 and HES-3 were purchased from the WiCell Research Institute (Madison, WI, USA) and Cythera Inc. (San Diego, CA, USA). The H1 line was used exclusively for the isolation of cell lysate that was loaded as positive control of the immunoblot analysis in Figure 1e/right panel. hESC line HESG (GENEA23) was purchased from GENEA Biocells (<http://www.geneastemcells.com.au>). It formed well defined colonies with compact cells displaying a high nuclear to cytoplasmic ratio and prominent nucleoli. Karyotype analysis (46 chromosomes, XY male) did not uncover any abnormalities at passage 42. HESG cells express pluripotency markers Nanog, Oct-4, Tra1-60 and SSEA4, stain positive for alkaline phosphatase and form teratomas. As for hiPSCs, hESCs were grown on gelatin-coated 6-well plates (Greiner) on inactivated mouse embryonic fibroblasts (MEFs, passage 3, strain CF1; Merck Millipore, Catalogue Number: PMEF-CFL). MEFs were expanded and mitotically inactivated by γ-irradiation with a Cesium source with 30Gy after 3-7 passages, and stored in liquid nitrogen until further use. After thawing, MEFs were seeded at a density of 6×10^5 cells per well of a 6-well plate. hESC medium was replaced daily and cells were passaged at a 1:2 dilution every five days using splitting medium (1mg/ml collagenase IV [Gibco, Darmstadt, Germany] in KO-DMEM). The absence of any mycoplasma contamination of cell lines or cultures used and analyzed was confirmed applying the PCR Mycoplasma Test Kit (Product-No: A3744, AppliChem, GmbH, Darmstadt, Germany) before the respective cultured cell line was included in the study.

Cell sorting

hiPSCs were washed once in PBS containing 0.5% bovine serum albumin, and incubated for 30min with allophycocyanin-conjugated anti-human SSEA4 antibody (R&D Systems). In all samples an anti-mouse Sca-1 (Ly-6A/E) (FITC or PE conjugated, BD Pharmingen) antibody was employed, for gating out the positively labeled mouse feeder cells. Samples were analyzed and sorted using an Aria High Speed Cell Sorter (Becton-Dickinson).

Differentiation of hiPSCs into embryoid bodies and RNA extraction

In all experiments, hiPSCs grown on MEFs were detached from the feeder layer by adding 250µl Collagenase Type IV (1mg/ml; Gibco) per well of a 6-well tissue culture plate. Next, cells were re-suspended in 750µl of ESC medium, transferred to a 15ml conical tube, and centrifuged at 800rpm in a Heraeus Multifuge 4KR for 3min at RT. Subsequently, medium was removed, cells were re-suspended in 3ml of ESC medium without FGF2 and cultured for one to 16 days in T25 flasks (Greiner) containing 10ml of ESC medium without FGF2. At the indicated time, embryoid bodies were harvested and cytoplasmic RNA was isolated as described below. Passage 10 of the hiPSC line hiPS-SB5.1 was cultured in one well of a Geltrex™-coated 6-well culture dish, and treated with collagenase IV (1 mg/ml) for 5min. Cells were washed with warm PBS twice, and fed with 1ml embryoid body (EB) formation medium (Knockout™ DMEM, 20% Knockout™ Serum Replacement, 1 mM L-Glutamine, 1% nonessential amino acids, 0.1 mM β-mercaptoethanol and Primocin™[Invivogen]) and split into small cell clumps. hiPSC colonies were then dissociated with collagenase IV (1 mg/ml) for 5min, and split into small cell clumps. Cell clumps were transferred into three 10-cm low-attachment dishes and fed with EB medium. The medium was changed every two days. EBs were cultured for eight days in total. EBs were collected by sedimentation under gravity from three dishes on day 0 (undifferentiated hiPSCs), 2, 4, 6, and 8, respectively (Fig.1b, right panel; Supplementary Fig. S6). Total RNA was extracted from each well using Trizol® (Invitrogen) following the instructions of the manufacturer.

Analyses of pluripotency marker and L1 expression in EBs by real-time quantitative RT-PCR (qRT-PCR).

0.1µg total RNA per well was used for reverse transcription by using the High Capacity RNA-to-cDNA kit (Applied Biosystems). For each time point and transcript to be quantified, qRT-PCR analyses were done in triplicate. qRT-PCR for pluripotency/differentiation markers was carried out using Power SYBR® Green PCR Master Mix (Applied Biosystems) on the ABI7900HT sequence detector (Applied Biosystems), and data was normalized to GAPDH expression. qRT-PCR for L1 was performed with Absolute QPCR Mix (ABgene), and data was normalized to 18s rRNA expression.

qRT-PCR using TaqMan® fluorogenic probes

Cytoplasmic RNA was extracted from 5×10^6 – 3×10^7 somatic cells, hiPSCs or EB cells using the RNeasy Midi Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Cytoplasmic RNA (0.5-1µg) was incubated with 2U of RNase-free DNaseI (Life Technologies, Darmstadt, Germany) for 30min at room

temperature (RT). DNaseI digestion was stopped by adding 2 μ l of 25mM EDTA and incubation for 10min at 65°C. DNaseI-digested cytoplasmic RNA (0.1-0.5 μ g) was used for cDNA synthesis using the SuperScript III™ First-Strand Synthesis Kit (Invitrogen) in combination with a Random Hexamer Primer (0.5 μ g/ μ l; Invitrogen) according to the manufacturer's instructions. Quantitative real-time PCR was carried out in ABgene plates using an Applied Biosystems 7900HT Fast Real-Time PCR System. The primer and probe combination L1 5'UTR#2⁶¹ was used to quantify transcripts expressed from endogenous L1-Ta copies. Sequences of oligonucleotides and probes used for qRT-PCR are listed in Supplementary Table S1. The probe specific for the L1 5'UTR was labeled with the reporter fluorochrome 6-carboxy-fluorescein (FAM) and a non-fluorescent quencher. 18S rRNA expression was quantified using Eukaryotic 18S rRNA Endogenous Control (VIC/TAMRA Probe, Primer Limited; Part number 4310893E, Applied Biosystems). Transcript levels of the human *CADPS2* gene were monitored using a gene specific assay (Life Technologies, Hs00604528_m1) spanning exon sequences (Fig. 5a). Cycling conditions were the following: 95°C for 15min (one cycle), 95°C for 15s and 60°C for 1min (40 cycles). 1-5 μ l of cDNA per sample were used for the quantification of endogenous L1 and *CADPS2* mRNA levels. Analysis of real-time and end point fluorescence was performed using the software SDS version 2.3 as well as RQ manager 1.2 (Applied Biosystems).

Northern blot analysis

Total RNA was isolated from the cell lines HFF-1, 2102Ep⁷², HES-3 and hiPS-SB4 using TRIzol® (Invitrogen) according to the manufacturer's instructions. Poly(A)+ RNA was isolated applying the Dynabeads® mRNA Purification Kit (Life Technologies) according to the manufacturer's instructions. 2.8 μ g denatured mRNA from each cell line was subjected to denaturing electrophoresis in a horizontal 1% agarose gel containing morpholinepropanesulfonic acid (MOPS) buffer and 6% formaldehyde, and transferred onto a Hybond-N⁺-Nylon membrane (Amersham) by overnight capillary transfer using 10xSSC as transfer buffer. 4 μ l RiboRuler™ High Range RNA ladder (MBI Fermentas, St.Leon-Rot, Germany) were loaded as size marker. After crosslinking the RNA onto the membrane by 'baking' at 80°C for two hours, the membrane was prehybridized overnight in 50% Formamide/4xSSC/1%SDS/2x Denhardt's at 42°C. The full-length L1 mRNA-specific probe was generated by PCR amplification of a 1299-bp L1 fragment ranging from pos. 58-1356 of a full length L1 element by using primers L1_FW1 5'- and L1_RV1 (see Supplementary Table S1) and pJM101/L1_{RP} Δ CMV⁷³ as template. Position numbers (pos.) refer to the L1.3 element⁵⁴ sequence (accession number L19088.1). A 491-bp β -actin mRNA-specific probe was generated by PCR amplification using primers actin_FW and actin_RV_(see Supplementary Table S1) and plasmid 31502 (Addgene⁷⁴) as template. PCR fragments were labeled with [α -³²P]dCTP by applying the Nick Translation System (Invitrogen) according to the manufacturer's instructions.

After denaturing the probe for 10min in boiling water and subsequent incubation for 10min in ice water, the probe was added to the hybridization buffer (50% Formamide/4xSSC/1% SDS/1xDenhardt's) and the membrane was incubated in the probe-containing hybridization buffer overnight at 42°C. Subsequently, the membrane was subjected to two 5min low-stringency washes (2xSSC) at RT and one 30min high-stringency wash (2xSSC/0.5%SDS) at 65°C. The

membrane was stripped by being boiled for 30min in a solution of 10mM tris-HCl (pH 7.5)/1mM EDTA/1mM SDS. The hybridized membrane was exposed to X-ray films for 5-10 days with intensifying screens.

Bisulfite DNA sequencing analyses

Bisulfite DNA sequencing analyses were performed as previously described^{20,61}. Briefly, genomic DNA from hiPSCs and parental cells was isolated at the indicated passage using DNazol® Genomic DNA Isolation Reagent (MRC Inc, Cincinnati, OH, USA) according to the manufacturer's instructions. Next, 2µg of genomic DNA were bisulfite converted using an EpiTect Bisulfite Kit (Qiagen, Hilden, Germany) following manufacturer instructions, with a conversion efficiency of ~95%. To determine the DNA methylation status of L1-Ta promoters, we performed PCR sequencing using primers L1-FW2: 5'-AAGGGGTTAGGGAGTTTTTTT and L1-RV2: 5'-TATCTATACCCTACCCCAAAA. To this end, 300-500ng of converted genomic DNA were used in a 50µl PCR reaction as follows: 2min at 95°C, 35 cycles of 30sec at 94°C followed by 30sec at 54°C and 60sec at 72°C, and a final extension of 10min at 72°C. Amplified products were gel purified (QIAquick gel extraction kit, Qiagen), cloned in pGEM-T Easy (Promega) and at least 30 individual clones were sequenced for each sample. The unique sequence in each clone was analyzed using Repeatmasker at <http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>. Next, the fraction of unmethylated CpG sites was calculated by comparison to a consensus L1-Ta sequence. In addition, each individual sequence was compared to L1.3 and only the sequences with the highest homology to this sequence were used to plot methylation data in single clones (see Supplementary Fig. S3d). The proportion of CpG converted to TpG by bisulfite treatment was compared between samples using the Chi-square test (d.f.= 1; α = 0.05).

Immunoblot analysis

hiPSC colonies were detached from their tissue culture dish by incubation with 250µl of a 1mg/ml collagenase type IV/DMEM and washed subsequently in 1x PBS. Cells were spun down, resuspended in lysis buffer (50mM Tris-HCl (pH 7.4), 150 mM NaCl, 10% Glycerin, 1% Triton X 100; 2mM EDTA, 2mM EGTA, 40mM β -Glycerolphosphate disodium salt hydrate, 50mM NaF, 10mM $\text{Na}_4\text{P}_2\text{O}_7$, 200 μM Na_3VO_4 , 2mM DTT; 1x complete protease inhibitor cocktail [Roche Applied Science]), homogenized by passing the lysate 10 times through a 26G needle, and lysates were cleared by centrifugation. 50µg of each protein lysate were boiled in 3x SDS sample buffer (NEB), loaded on 4-12 % Bis/Tris gels (Invitrogen), subjected to SDS-PAGE, and electroblotted onto nitrocellulose membranes. After protein transfer, membranes were blocked for two hours at RT in a 10% solution of non-fat milk powder in 1x PBS-T (137mM NaCl, 3mM KCl, 16.5mM Na_2HPO_4 , 1.5mM KH_2PO_4 , 0.05% Tween 20 [Sigma-Aldrich Chemie GmbH, Mannheim, Germany]), washed in 1xPBS-T, and incubated overnight with the respective primary antibody at 4°C.

L1 ORF1p and OCT4 proteins were detected using the polyclonal rabbit-anti-L1 ORF1p antibody #984⁴² at a 1:2000 dilution and the Oct-3/4 (C10) antibody (sc-5279, Santa Cruz Biotechnology Inc., Santa Cruz, CA, USA) at a 1:750 dilution, respectively, in 1xPBS-T containing 5% milk powder as primary antibodies. Subsequently, membranes were washed thrice in 1xPBS-T. As secondary antibodies, we used HRP-conjugated donkey anti-rabbit IgG antibody at

a 1:30,000 dilution to detect L1 ORF1p, and HRP-conjugated donkey anti-mouse IgG antibody at a 1:10,000 dilution (Amersham Biosciences) to detect Oct-3/4, in 1xPBS/5% milk powder and incubated the membrane for 2 hours. Subsequently, the membrane was washed thrice for 10 min in 1xPBS-T. β -actin expression was detected using a monoclonal anti- β -actin antibody (clone AC-74, Sigma-Aldrich Chemie GmbH, Steinheim, Germany) at a dilution of 1:30,000 as primary antibody and an anti-mouse HRP-linked species-specific antibody (from sheep) at a dilution of 1:10,000 as secondary antibody. Immunocomplexes were visualized using lumino-based ECL immunoblot reagent (Amersham Biosciences Europe GmbH, Freiburg, Germany). Details of the applied antibodies are listed in Supplementary Table S2.

Immunofluorescence staining

hiPSCs as well as their parental HFF-1 or hCBEC cells were grown on glass cover slips in 12-well plates. Cells were washed with 1xPBS, fixed with 4% paraformaldehyde in 1xPBS (pH 7.4) for 15min at RT and permeabilized with 1% Triton X-100 (Sigma) in 1xPBS for 10min at RT. Subsequently, cells were washed thrice for 2min in 1xPBS. Cells were blocked by incubation with 5% (w/v) BSA/0,1% Triton X-100/1xPBS (pH 7.4) for 30min at RT followed by incubation with the respective primary antibodies, which are listed in Supplementary Table S2, for 1 hour at RT in 5% BSA/1xPBS (pH 7.4). Subsequently, cells were washed three times with 1xPBS for 5min each at RT. Cells were incubated with the appropriate secondary antibody: goat-anti-mouse IgG Alexa 488 or goat-anti-rabbit IgG Alexa 643 (Invitrogen) at 1:1000 dilution in 5% BSA/1xPBS (pH 7.4) for 30min at RT in the dark. Finally, preparations were washed thrice for 5min each at RT using 1xPBS. Subsequently, cells were counterstained with DAPI (4,6-diamidino-2-phenylindole; Sigma-Aldrich), washed thrice with 1xPBS for 10min at RT, embedded in Fluoromount G (Southern Biotech) and kept at 4°C until further analysis. The analysis was performed using an Axio Observer A1 microscope (Carl Zeiss MicroImaging, Goettingen, Germany).

RC-seq library preparation, sequencing and analysis

Genomic DNA was isolated from 1×10^6 cells from each hESC and hiPSC line and their respective parental cells using DNAzol[®] Genomic DNA Isolation Reagent (MRC Inc, Cincinnati, OH, USA) according to the manufacturer's instructions. RC-seq and subsequent computational analyses were performed as described using the hg19 reference genome sequence³⁴. A total of 665,008,770 2x150mer reads were generated from 24 libraries. A complete list of annotated *de novo* insertions supported by at least two unique amplicons separated by ≥ 5 nt (the minimum threshold for reporting) is provided in Supplementary Table S3. To assess the RC-seq false negative rate we randomly sampled each library in increments of 1% (10 samplings per percentile) and determined how many germline insertions were detected at the sampled depth by ≥ 2 unique reads (Supplementary Fig. S11). To approximately assess the rate of L1 mobilization in hiPSCs, we again randomly sampled each RC-seq library to determine the probability of detecting each *de novo* L1 insertion with ≥ 2 unique reads at a given sampling depth, normalized to the corresponding false negative rate identified above and then determined the cumulative sum of this distribution for frequencies of 5%-100%, leading to an estimate of ~ 1 *de novo* L1 insertion

per hiPSC. We did not consider *de novo* L1 insertions carried by fewer than 5% of hiPSCs in this estimate as none of the validated examples were routinely identified at that sampling depth. We also did not analyse the L1 mobilization rate in hESCs or the *Alu* or SVA rate in hiPSCs or hESCs due to the small number of confirmed true positive examples.

A permutation test showing enrichment for validated *de novo* L1 insertions at the 5' end of genes was performed by random sampling of genomic coordinates, with respect to RefSeq annotations. 1×10^6 permutations were performed and in 6000 instances the average position was less than the 20th percentile of gene length, indicating $P < 0.006$.

PCR validation of *de novo* insertions

Seventeen *de novo* insertions (8 L1, 7 *Alu*, 2 SVA) detected by RC-seq were first assayed with PCR using a standard empty site / filled site genotyping assay. Primers were positioned on either side of the insertion site so that the predicted PCR product of the empty site covered <300bp. Additional retrotransposon specific primers were designed and paired with the existing insertion site-specific primers if required. PCR reactions contained 0.125 μ l Crimson Taq (New England Biolabs), 5x PCR-buffer, 10pMol of each Primer, 10mM dNTPs and 10-20ng genomic template DNA in a total volume of 25 μ l. The following cycling conditions were used: 95°C for 2min, then 35 cycles of 95°C for 30sec, 58°C for 30sec, 68°C for 40sec, followed by a single extension step at 68°C for 5min. Optimization in some cases required adjusted annealing temperatures and cycle number. PCR products of the correct size (Fig. 3a) that were obtained with the retrotransposon primer in combination with the genomic primer were TA-cloned and sequenced. The same method was applied to both the 5' and the 3' ends of all *de novo* insertions to fully characterize each, apart from SVA-2. To PCR amplify the 5' junction of the SVA-2 insertion from genomic DNA, we designed three SVA_E-specific primers and three oligonucleotides binding 50-300bp upstream of the SVA-2 integration site. To facilitate the detection of a potentially 5'-truncated SVA, the SVA-specific primers were placed within the sequenced 123bp of the SVA-2 3' end (Supplementary Fig. S9), at the junctions of the SVA_E-specific *Alu*-like and VNTR region, and the (CCCTCT)_n repeat and *Alu*-like region, respectively. Combinatorial use of these genome-SVA primer pairs did either not result in a PCR product or generated non-specific products. For a complete list of used primers see Supplementary Table S4. Eleven *de novo* insertions (7 L1, 3 *Alu*, 1 SVA) were confirmed by PCR as *de novo*. Six additional insertions were determined as germ line insertions, already present in the parental cell line or an early hESC passage. Control genotyping PCR of the single-copy gene GAPDH in genomic DNA preparations of parental and hiPSC lines used for RC-seq and PCR validations of *de novo* insertions is presented in Supplementary Fig. S14. PCR amplification was performed using primers GAPDH- α (5'-CAAAGCTTGTGCCAGACTGTG3') and GAPDH- β (5'-GAGAGCTGGGGAATGGGACT3') which bind in exon 8 (chr12:6646561-6646580) and intron 7 (chr12:6647005-6647026), respectively, resulting in a 466-bp DNA fragment. Cycling conditions were identical to those described above.

Real-Time quantitative PCR (qPCR) analyses of L1-dn13 and L1-dn14 copy numbers using TaqMan® fluorogenic probes

Genomic DNA was isolated using 1ml DNazol® Genomic DNA Isolation Reagent (MRC Inc, Cincinnati, OH, USA) from 1×10^6 cells, according to the manufacturer's instructions. 100ng of genomic DNA was used for quantitative real-time PCR (qPCR). Primer and probe combinations specific to the genomic 5' junctions of the *de novo* insertions L1-dn13 and L1-dn14 (Fig. 3b) were used to quantify the copy number of the respective insertion in hiPSC cultivars. Each probe was labeled with flourochrom6-carboxyfluorescein (FAM) and a non-fluorescent quencher. For normalization the single-copy gene RPP25 (Ribonuclease P/MRP 25kDa subunit; FAM/non-fluorescent quencher, primer limited, HS00706565_S1; Applied Biosystems) was used. Cycling conditions were: 95°C for 15min (one cycle), 95°C for 15sec and 60°C for 1min (40 cycles). For analysis of real-time and end point fluorescence the software SDS version 2.3 as well as RQ manager 1.2 (Applied Biosystems) were used.

Isolation of hiPS-SB4 single cell subclones by limiting dilution and validation of subclones

hiPS-SB4 cells of passage 64 representing a mixed population of cells with and without the L1-dn13 *de novo* retrotransposition event were magnetically separated from feeder cells by applying a Feeder Removal Kit (Miltenyi Biotech GmbH, Bergisch Gladbach, Germany) according to the manufacturer's instructions. hiPSCs were counted and seeded on feeder-coated 96well plates (Cat.No: 167008, Thermo Fisher/Nunc, Roskilde, Denmark) at a cell density of one cell per well or 0.3 cells per well. hiPSCs were grown for 24hr in the presence of 10 μ M ROCK inhibitor (Y-27632, Sigma-Aldrich). Subsequently, cells were cultivated until they formed a single colony per well. Single colonies were transferred to feeder-coated 12well plates (Thermo Fisher/Nunc) and further expanded. To isolate genomic DNA from each clone, cells were harvested after collagenase IV treatment, centrifuged, washed and pelleted again. Genomic DNA was isolated as describe in the previous paragraph. Genotyping PCR conditions applied to screen for the presence of the L1-dn13 insertion and to demonstrate its presence/absence (Fig. 5d) are identical to those described above for insertion PCR validation. Primers used to demonstrate presence/ absence of L1-dn13 are provided in Supplementary Tables S3 and S4. PCR products were visualized on a 1.5 % agarose gel after ethidium bromide staining.

Cloning of full-length L1s and retrotransposition reporter assays

Candidate full-length L1s were amplified from genomic DNA using an Expand Long Template PCR system (Roche) and primers located 50bp upstream/downstream the insertion site (available upon request). For each PCR we used: 0.3 μ l Expand Long Template Taq (Roche), 1x buffer#1, 400 μ M dNTPs, 1 μ M each Primer and 300ng genomic DNA in 50 μ l/tube. Cycling conditions were: 95°C for 5min, then 30 cycles of 95°C for 1min, 56°C for 30sec, 68°C for 6min, followed by a single extension step at 68°C for 10min. To avoid the generation of mutations that may lead to retrotransposition defective elements, we conducted at least 4 independent PCRs per L1. PCR products were resolved on 0.9% agarose gels, and fragments of the expected length of ~6kb representing potential full-length L1 elements were excised and purified using a Qiaquick kit (Qiagen) and cloned in the Topo-XL plasmid (Invitrogen). Each of the cloned PCR products carrying full-length L1 elements L1-dn4, L1-dn6 and L1-dn14 were sequenced (Supplementary Fig. S10).

To evaluate retrotransposition competence of the L1-dn6 *de novo* insertion, two independent genomic PCR amplicons were inserted into the pJJ101/L1.3 backbone after the deletion of the L1.3 sequence by *Not I/BstZ171* restriction^{9,75}. pJJ101/L1.3 contains a full-length active human L1 element (L1.3) tagged with a *mblastI* retrotransposition indicator cassette⁷⁵ cloned in vector pCEP4 (Invitrogen). In total, we generated five JJ101-derived plasmids containing an L1-dn6 element amplified from genomic DNA by PCR (data not shown). For retrotransposition assays, these L1 reporter plasmids were purified using a Qiagen Midiprep system (Qiagen) and only highly supercoiled preparations were used in the following assays.

Retrotransposition assays in HeLa cells were conducted as described previously^{9,47,53,75}. HeLa cells were purchased from ATCC. Cytogenetic authentication of HeLa cells was performed by spectral karyotyping (SKY)-FISH. HeLa cells used in this study were tested for mycoplasma contamination monthly. Briefly, HeLa cells were cultured using DMEM-high glucose (4.5g/l) supplemented with L-glutamine, Penicillin/Streptomycin, and 10% Fetal Bovine Serum (all reagents from GIBCO-Invitrogen) and passaged using Trypsin 0.05% (GIBCO-Invitrogen). 10⁴ HeLa cells/well were plated in triplicate using 6-well tissue culture plates. After 18h, cells were transfected with 1µg/well of plasmid using 3µl of Fugene6 (Promega) following manufacturer's instructions. Next day, medium was replaced and cells cultured for five additional days. Six days after transfection, Blastidin-S (Invitrogen) was added to a final concentration of 10µg/ml and cells were cultured for seven days in the presence of the antibiotic. Next, plates were fixed and stained with crystal violet, and foci counted manually.

Statistical analyses of relative L1 RNA levels

The statistical evaluation of relative L1 mRNA levels determined by qRT-PCR was performed by Analysis of Variance (ANOVA), using Bonferroni correction for multiple comparisons with the same control group. Reduction in full-length transcript levels in the EB time kinetics experiment was evaluated by means of Linear Regression for data from day 0 to day 8 (R-Square=0.79). Analyses were performed with SAS®/STAT software, version 9.2 SAS system for Windows.

References

1. Yamanaka, S. Induced pluripotent stem cells: past, present, and future. *Cell Stem Cell* **10**, 678-684 (2012).
2. Gore, A. et al. Somatic coding mutations in human induced pluripotent stem cells. *Nature* **471**, 63-67 (2011).
3. Hussein, S.M. et al. Copy number variation and selection during reprogramming to pluripotency. *Nature* **471**, 58-62 (2011).
4. Laurent, L.C. et al. Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. *Cell Stem Cell* **8**, 106-118 (2011).
5. Lister, R. et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471**, 68-73 (2011).
6. Mayshar, Y. et al. Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell* **7**, 521-531 (2010).

7. Ben-David, U. & Benvenisty, N. The tumorigenicity of human embryonic and induced pluripotent stem cells. *Nat Rev Cancer* **11**, 268-277 (2011).
8. Levin, H.L. & Moran, J.V. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet* **12**, 615-627 (2011).
9. Beck, C.R. et al. LINE-1 retrotransposition activity in human genomes. *Cell* **141**, 1159-1170 (2010).
10. Mills, R.E., Bennett, E.A., Iskow, R.C. & Devine, S.E. Which transposable elements are active in the human genome? *Trends Genet* **23**, 183-191 (2007).
11. Brouha, B. et al. Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A* **100**, 5280-5285 (2003).
12. Ewing, A.D. & Kazazian, H.H., Jr. Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans. *Genome Res* **21**, 985-990 (2011).
13. Garcia-Perez, J.L. et al. LINE-1 retrotransposition in human embryonic stem cells. *Hum Mol Genet* **16**, 1569-1577 (2007).
14. Kano, H. et al. L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev* **23**, 1303-1312 (2009).
15. Han, J.S., Szak, S.T. & Boeke, J.D. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429**, 268-274 (2004).
16. Hancks, D.C. & Kazazian, H.H., Jr. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* **22**, 191-203 (2012).
17. Beck, C.R., Garcia-Perez, J.L., Badge, R.M. & Moran, J.V. LINE-1 elements in structural variation and disease. *Annu Rev Genomics Hum Genet* **12**, 187-215 (2011).
18. Bourc'his, D. & Bestor, T.H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96-99 (2004).
19. Maherali, N. et al. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1**, 55-70 (2007).
20. Wissing, S. et al. Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. *Hum Mol Genet* **21**, 208-218 (2012).
21. Friedli, M. et al. Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Res* **24**, 1251-1259 (2014).
22. Wissing, S., Montano, M., Garcia-Perez, J.L., Moran, J.V. & Greene, W.C. Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. *J Biol Chem* **286**, 36427-36437 (2011).
23. Cheng, L. et al. Low incidence of DNA sequence variation in human induced pluripotent stem cells generated by nonintegrating plasmid expression. *Cell Stem Cell* **10**, 337-344 (2012).
24. Quinlan, A.R. et al. Genome sequencing of mouse induced pluripotent stem cells reveals retroelement stability and infrequent DNA rearrangement during reprogramming. *Cell Stem Cell* **9**, 366-373 (2011).
25. Briggs, J.A. et al. Integration-free induced pluripotent stem cells model genetic and neural developmental features of down syndrome etiology. *Stem Cells* **31**, 467-478 (2013).

26. Grabundzija, I. et al. Sleeping Beauty transposon-based system for cellular reprogramming and targeted gene insertion in induced pluripotent stem cells. *Nucleic Acids Res* **41**, 1829-1847 (2013).
27. Haase, A. et al. Generation of induced pluripotent stem cells from human cord blood. *Cell Stem Cell* **5**, 434-441 (2009).
28. Abyzov, A. et al. Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. *Nature* **492**, 438-442 (2012).
29. Goodier, J.L., Zhang, L., Vetter, M.R. & Kazazian, H.H., Jr. LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Mol Cell Biol* **27**, 6469-6483 (2007).
30. Doucet, A.J. et al. Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet* **6** (2010).
31. Rodic, N. et al. Long interspersed element-1 protein expression is a hallmark of many human cancers. *Am J Pathol* **184**, 1280-1286 (2014).
32. Goodier, J.L., Pereira, G.C., Cheung, L.E., Rose, R.J. & Kazazian, H.H., Jr. The Broad-Spectrum Antiviral Protein ZAP Restricts Human Retrotransposition. *PLoS Genet* **11**, e1005252 (2015).
33. Moldovan, J.B. & Moran, J.V. The Zinc-Finger Antiviral Protein ZAP Inhibits LINE and Alu Retrotransposition. *PLoS Genet* **11**, e1005121 (2015).
34. Shukla, R. et al. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* **153**, 101-111 (2013).
35. Baillie, J.K. et al. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **479**, 534-537 (2011).
36. Ewing, A.D. & Kazazian, H.H., Jr. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res* **20**, 1262-1270 (2010).
37. Iskow, R.C. et al. Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* **141**, 1253-1261 (2010).
38. Wang, J. et al. dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. *Hum Mutat* **27**, 323-329 (2006).
39. Mir, A.A., Philippe, C. & Cristofari, G. euL1db: the European database of L1HS retrotransposon insertions in humans. *Nucleic Acids Res* **43**, D43-47 (2015).
40. Wang, H. et al. SVA elements: a hominid-specific retroposon family. *J Mol Biol* **354**, 994-1007 (2005).
41. Symer, D.E. et al. Human l1 retrotransposition is associated with genetic instability in vivo. *Cell* **110**, 327-338 (2002).
42. Raiz, J. et al. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Res* **40**, 1666-1683 (2012).
43. Gilbert, N., Lutz, S., Morrish, T.A. & Moran, J.V. Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Mol Cell Biol* **25**, 7780-7795 (2005).
44. Luan, D.D., Korman, M.H., Jakubczak, J.L. & Eickbush, T.H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595-605 (1993).
45. Cost, G.J., Feng, Q., Jacquier, A. & Boeke, J.D. Human L1 element target-primed reverse transcription in vitro. *EMBO J* **21**, 5899-5910 (2002).

46. Jurka, J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc Natl Acad Sci U S A* **94**, 1872-1877 (1997).
47. Morrish, T.A. et al. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet* **31**, 159-165 (2002).
48. Zingler, N. et al. Analysis of 5' junctions of human LINE-1 and Alu retrotransposons suggests an alternative model for 5'-end attachment requiring microhomology-mediated end-joining. *Genome Res* **15**, 780-789 (2005).
49. Ostertag, E.M. & Kazazian, H.H., Jr. Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res* **11**, 2059-2065 (2001).
50. Grimaldi, G., Skowronski, J. & Singer, M.F. Defining the beginning and end of KpnI family segments. *EMBO J* **3**, 1753-1759 (1984).
51. Lee, E. et al. Landscape of somatic retrotransposition in human cancers. *Science* **337**, 967-971 (2012).
52. Solyom, S. et al. Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res* **22**, 2328-2338 (2012).
53. Moran, J.V. et al. High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917-927 (1996).
54. Sassaman, D.M. et al. Many human L1 elements are capable of retrotransposition. *Nat Genet* **16**, 37-43 (1997).
55. Cost, G.J., Golding, A., Schlissel, M.S. & Boeke, J.D. Target DNA chromatinization modulates nicking by L1 endonuclease. *Nucleic Acids Res* **29**, 573-577 (2001).
56. Thurman, R.E. et al. The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82 (2012).
57. Kaer, K. & Speek, M. Retroelements in human disease. *Gene* **518**, 231-241 (2013).
58. Guenther, M.G. et al. Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell* **7**, 249-257 (2010).
59. Nott, A., Meislin, S.H. & Moore, M.J. A quantitative analysis of intron effects on mammalian gene expression. *RNA* **9**, 607-617 (2003).
60. Schmittgen, T.D. et al. Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: comparison of endpoint and real-time methods. *Anal Biochem* **285**, 194-204 (2000).
61. Coufal, N.G. et al. L1 retrotransposition in human neural progenitor cells. *Nature* **460**, 1127-1131 (2009).
62. Arokium, H. et al. Deep sequencing reveals low incidence of endogenous LINE-1 retrotransposition in human induced pluripotent stem cells. *PLoS One* **9**, e108682 (2014).
63. Upton, K.R. et al. Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* **161**, 228-239 (2015).
64. van den Hurk, J.A. et al. L1 retrotransposition can occur early in human embryonic development. *Hum Mol Genet* **16**, 1587-1592 (2007).
65. Richardson, S.R., Narvaiza, I., Planegger, R.A., Weitzman, M.D. & Moran, J.V. APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. *Elife* **3**, e02008 (2014).

66. Marchetto, M.C. et al. Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* **503**, 525-529 (2013).
67. Ostertag, E.M. & Kazazian, H.H., Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet* **35**, 501-538 (2001).
68. Coufal, N.G. et al. Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proc Natl Acad Sci U S A* **108**, 20382-20387 (2011).
69. Kapranov, P. et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484-1488 (2007).
70. Draper, J.S. et al. Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nat Biotechnol* **22**, 53-54 (2004).
71. Becherel, O.J. et al. A new model to study neurodegeneration in ataxia oculomotor apraxia type 2. *Hum Mol Genet* **24**, 5759-5774 (2015).
72. Andrews, P.W., Bronson, D.L., Benham, F., Strickland, S. & Knowles, B.B. A comparative study of eight cell lines derived from human testicular teratocarcinoma. *Int J Cancer* **26**, 269-280 (1980).
73. Kimberland, M.L. et al. Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. *Hum Mol Genet* **8**, 1557-1560 (1999).
74. Watanabe, N. & Mitchison, T.J. Single-molecule speckle analysis of actin filament turnover in lamellipodia. *Science* **295**, 1083-1086 (2002).
75. Kopera, H.C., Moldovan, J.B., Morrish, T.A., Garcia-Perez, J.L. & Moran, J.V. Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase. *Proc Natl Acad Sci U S A* **108**, 20345-20350 (2011).
76. Richards, M., Fong, C.Y., Chan, W.K., Wong, P.C. & Bongso, A. Human feeders support prolonged undifferentiated growth of human inner cell masses and embryonic stem cells. *Nat Biotechnol* **20**, 933-936 (2002).
77. Thomson, J.A. et al. Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145-1147 (1998).
78. Kapitonov, V.V. & Jurka, J. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat Rev Genet* **9**, 411-412; author reply 414 (2008).

Acknowledgements

We thank O. Weichenrieder for the generous gift of an anti-ORF1p antibody, R. Löwer and Csaba Miskey for helpful discussions, K.-M Hanschmann for statistical analyses, as well as S. Heras and F. Sanchez-Luque for technical suggestions on RNA extraction and manipulation. This work was supported by the LOEWE Center for Cell and Gene Therapy Frankfurt (funded by the Hessian Ministry of Higher Education, Research and the Arts; funding reference number: III L 4- 518/17.004 [2010]) (to S.K. and G.G.S.), grants SCHU1014/8-1 (to G.G.S.) and MA2331/11-1 (to U.M.) from the *Deutsche Forschungsgemeinschaft*, and EU-FP6/CliniGene-NoE-

Flexibility fund FLEXI7-*m* (to G.G.S. and Z.Iz.), EU-FP7 InduStem (grant number 230675 to Z.Iv.), Hungarian stem cell project TÁMOP-4.2.2-08/1-2008-0015 (to A.S., B.S., Z.Iz. and Z.Iv.), OTKA 83533 (B.S.), the *Bundesministerium für Bildung und Forschung* (ReGene, grant number 01GN1003A, to Z.Iv.) and REBIRTH Cluster of Excellence, EXC 62/3 (to U.M.). G.J.F. acknowledges the support of an NHMRC Career Development Fellowship (GNT1045237) and NHMRC Project Grants GNT1042449, GNT1045991, GNT1067983 and GNT1068789. J.L.G.P.'s lab is supported by CICE-FEDER-P09-CTS-4980, CICE-FEDER-P12-CTS-2256, Plan Nacional de I+D+I 2008-2011 and 2013-2016 (FIS-FEDER-PI11/01489 and FIS-FEDER-PI14/02152), PCIN-2014-115-ERA-NET NEURON II, the European Research Council (ERC-Consolidator ERC-STG-2012-233764) and by an International Early Career Scientist grant from the Howard Hughes Medical Institute (IECS-55007420). M.M.-L. was supported by PeS-FEDER-PI-0224-2011.

Author contributions

S.K., J.L.G.-P., G.J.F. and G.G.S. designed the experiments. G.G.S. coordinated the project. G.J.F. and G.G.S. wrote the manuscript. S.K., N.V.F., J.W., I.G., A.B., A.S., B.S., S.M., A.W., A.H., E.J.W., M.G.-C., C.L.-R., J.A.P., U.M. and Z.Iv. generated hiPSC lines, cultivated hiPSC and hESC lines, characterized their pluripotency and/or extracted DNA, RNA and protein from these cells. K.R.U., R.S., D.J.G. performed RC-seq experiments. G.J.F. led the bioinformatic analyses. N.V.F., K.R.U., G.J.F., M.M.L. and J.L.G.-P. validated de novo insertions by genotyping PCR. S.K., N.V.F., J.W., J.L. and Z.Iz. performed qRT-PCR and qPCR experiments. N.V.F. and M.G.-C. performed immunoblot analyses. U.H. and N.V.F. performed Northern Blot analyses and luciferase reporter assays. M.M.L., P.G. and J.L.G.-P. were responsible for bisulfite sequencing and L1 retrotransposition reporter assays. Each author contributed to manuscript editing.

Additional information

Accession codes: RC-seq FASTQ files are available from the Sequence Read Archive (SRA) with accession PRJEB3191.

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at
<http://npg.nature.com/reprintsandpermissions/>

How to cite this article: [TBA]

Tables

Table 1: Analyzed pluripotent stem cell lines and their characteristics

Stem cell line	Parental cells		Reprogramming factors	Factor delivery by	Reference	Passages (p) assayed by RC-seq	
	Name	Description					
hiPS-SB4	HFF-1	Foreskin fibroblasts (male)	OCT4, SOX2, KLF4, c-MYC	<i>Sleeping Beauty</i> transposon	²⁶	43, 53	
hiPS-SB5			OCT4, SOX2, KLF4, c-MYC, LIN28			40, 59	
hiPS-CRL1502	CRL-1502	Dermal fibroblasts (female)	OCT4, SOX2, NANOG, LIN28, KLF4, c-MYC	oriP/EBNA1-based pCEP4 episomal vector	²⁵	15, 40	
hiPS-CRL2429	CRL-2429	Dermal fibroblasts (male)				11, 40	
hiPS-FB	FB	Dermal fibroblasts (female)	OCT4, SOX2, KLF4, c-MYC	Lentiviral vector	⁷¹	7, 23	
hFF-iPS4	HFF-1	Foreskin fibroblasts (male)	OCT4, SOX2, NANOG, LIN28			unpublished	58
hCBiPS1	hCBEC	Cord blood-derived endothelial cells (male)	OCT4, SOX2, NANOG, LIN28			²⁷	30
hCBiPS2				23			
HES-3	hESC lines		n/a	n/a	⁷⁶	92, 102	
H9						⁷⁷	30, 60
HESG						unpublished	10, 23

Figure legends

Figure 1: Reprogramming induced expression of the L1 retrotransposition machinery is abrogated during embryoid body formation.

(a) Schematic representation of the organization and expression of a functional human L1 element. A 363-bp region of the 5'-UTR including the CpG island (Pos. 232-491 of L1 reference sequence L1.3⁵⁴) was analyzed for its methylation status by bisulfite-PCR sequencing. Binding sites of the TaqMan primer/probe combinations (small convergent arrows) on the L1 cDNA used for qRT-PCR analyses and of the 1299-bp [α -³²P]dCTP-labelled PCR product in the 5'UTR region (black bar) for Northern analysis are shown. Open circles indicate CpG residues.

(b) Relative full-length L1 mRNA transcript levels were assessed by qRT-PCR using cytoplasmic RNA extracted from early passage HFF-1- and hCBEC-derived hiPSC lines (hiPSC cultivation, left panel), and after differentiation of hFF-iPS4 (passage 50) and hiPS-SB4 (passage 98) lines into embryoid bodies (EB differentiation, middle panel). L1 mRNA levels accumulated upon culturing relative to the parental cell line HFF-1 or hCBEC cells by 4-25,000-fold or 48-1940-fold, respectively, depending on passage number (p) after reprogramming (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). hiPS-SB5.1 cells (p10) were differentiated into EBs and L1 transcript levels were quantified on day 0 prior to the initiation of differentiation, and after 2, 4, 6 and 8 days of differentiation by qRT-PCR on total RNA (EB time kinetics, right panel). Between day 0 and day 8, a statistically significant reduction of L1 full-length RNA levels was observed (*** $p < 0.001$; Linear Regression t-test). Primer/probes correspond to the L1 5'UTR. Values are normalized to 18s rRNA levels and to the endogenous L1 mRNA / transcript level of HES-3, which was set as 1. Each bar represents the arithmetic mean \pm SD of data obtained from experiments that were performed as technical duplicates of biological triplicates.

(c) Northern Blot analysis of cytoplasmic poly A mRNA with a 1299-bp L1 5'UTR-specific probe (see Methods) confirmed exceeding activation of full-length L1 transcription (FL-L1) during hiPSC cultivation. Each lane of the 1% agarose gel was loaded with 2.8 μ g polyA+ mRNA isolated from HFF-1, HES-3, 2102Ep and hiPS-SB4 (passage 27) cells. β -actin mRNA (1.8 kb) expression served as loading control.

(d) Endogenous L1 promoter sequences were significantly hypomethylated in hiPSC lines relative to their parental cell lines HFF-1 and hCBEC. Each column

represents the overall percentage methylation of the CpG island located in the L1 promoter (Fig. 1a) for parental HFF-1 and hCBEC cells (n=29 CpG islands) versus five hiPSC lines (n=95 CpG islands). The p-value was calculated applying the χ^2 test. Samples were obtained from two parental and their five derived hiPSC lines. None of the available samples was excluded from the analysis. Individual CpG dinucleotide methylation frequencies for each cell line are presented in Supplementary Figure S3. Error bars indicate SEM *** $p < 0.001$;

(e) Immunoblot analysis of cell lysates from parental HFF-1 and hCBEC cells and the respective hiPSC lines (Table 1) was used to measure L1 protein expression. L1 ORF1p (40 kDa) and Oct-3/4 (A isoform, 45 kDa; B isoform 33 kDa) were detected with anti-ORF1p and anti-Oct3/4 antibodies, respectively. ORF1p was present in all pluripotent cell lines and absent from the respective parental lines. Oct-3/4 expression was used as pluripotency marker. Shorter (exp.1) and longer exposures (exp.2) of the α Oct-3/4 immunoblot are presented to demonstrate both differences in Oct3/4 abundance in the various cell lines and precise molecular weight of the Oct3/4 signal in hiPS-SB4 and hiPS-SB5 cells. Lysates from hESC lines HES-3 (left panel) and H1 (right panel) served as positive control for L1 ORF1p and Oct-3/4 expression. β -actin (42 kDa) was used as loading control.

Figure 2: Immunofluorescence staining of hiPSC colonies and their parental cells to monitor induction of endogenous L1 ORF1p expression in HFF-1 derived hiPSCs.

(a) ORF1p staining indicates activation of endogenous L1 expression after reprogramming of HFF-1 cells into lines hiPS-SB4, hiPS-SB5, and hFF-iPS4. Cells were analyzed at passages (p) 23, 16 and 60, respectively. Oct3/4 staining confirmed the pluripotent status of the analyzed stem cell colonies. Mesenchymal stem cell marker CD105 (endoglin) is reported to be expressed in HFF-1 cells but not expressed in pluripotent stem cells.

(b) Enlarged areas indicated by boxed dashes in **(a)** demonstrate cytoplasmic localization of endogenous L1 ORF1p and its accumulation in foci. Scale bars represent 20 μ m.

Figure 3: RC-seq reveals endogenous *de novo* L1, *Alu* and SVA retrotransposition in pluripotent stem cells.

(a) Schematic representations of validated *de novo* L1, *Alu* and SVA insertions. Each illustration depicts the structure of a retrotransposition event (red: untranslated region, white: open reading frame, green diamond: TSD) and includes the name of the insertion (e.g L1-dn13), and a gene name (e.g *CADPS2*) or chromosomal position for intergenic examples. RC-seq reads are aligned above the respective insertion. Nucleotide positions of the 5' ends of L1 and *Alu* insertions refer to the reference sequences for L1.3⁵⁴ and *AluYb8*⁷⁸ and are provided below each insertion. In cases where an insertion was detected by RC-seq at one terminus only, PCR and capillary sequencing were applied to the remaining end to resolve integration site structure. Validation primers are indicated by α and β . Corresponding PCR results are provided to the right of each schematic. Further details are provided in Supplementary Fig. S9. Note: retrotransposons not drawn to scale.

(b) Increased L1-dn13 copy number coincides with decreased L1-dn14 copy number during hiPS-SB4 cultivation. Relative L1-dn13 and L1-dn14 copy numbers (relative L1-dn13/L1-dn14 content) at hiPS-SB4 passages 43 and 53 were determined by qPCR using genomic DNA isolated from both passages and primer/probe combinations specific for the 5' junctions of insertions L1-dn13 and L1-dn14, respectively. Binding sites of the TaqMan primer/probe combinations are shown (Top panels, red arrows and lines). Genomic DNAs from HFF-1, and HES-3 cells lacking both L1-dn13 and L1-dn14 served as negative controls. For normalization, a primer/probe combination specific for the human single-copy gene *RPP25* was used. $\Delta\Delta C_t$ values measured the relative quantity of L1-dn13 and L1-dn14, respectively. Results are presented as relative insertion content and normalized to the parental cell line HFF-1. Bars indicate arithmetic means \pm SEM of technical triplicates. Due to the minimal SEMs observed in the L1-dn14-specific qPCR experiments (right panel), error bars fell below the presentable threshold and are therefore not visible.

(c) Passaging scheme of hiPSC line hiPS-SB4 harboring L1-dn13. After the generation of the hiPS-SB4 line by reprogramming of HFF-1 cells, the culture was continuously cultivated for 60 passages (culture 1). Genomic DNA (gDNA) was

isolated from culture 1 at passages 43, 53 (see Fig.3a), 56, 58 and 60 (red lettering). gDNAs from passages 43 and 53 were used to identify *de novo* retrotransposition events (including L1-dn13) by RC-seq, PCR (insertion validation) and qPCR (Fig. 3b). gDNAs of passages 43, 56, 58 and 60 (red lettering) were used to quantify the relative L1-dn13 content of the respective culture passage by qPCR. Cells of passage 19 of culture 1 were split and half of the culture was stored by cryo-preservation and cultivated again after several weeks of cryo-preservation (culture 2) until hiPS-SB4 culture 2 reached passage 49 (blue lettering). gDNA was isolated from passages 28, 34, 43 and 49 (blue lettering) of culture 2.

d) The L1-dn13 retrotransposition event occurred after reprogramming, between passages 19 and 43 of the original hiPS-SB4 culture 1. Relative L1-dn13 copy numbers (L1-dn13 content) at passages 43, 56, 58 and 60 of culture 1 (red lettering) and at passages 28, 34, 43 and 49 of culture 2 (blue lettering) (Fig. 3c) were quantified by qPCR using primer/probe combinations specific for the 5' junction of the L1-dn13 insertion (see Fig. 3b). While L1-dn13 was present in passages 43 to 60 of culture 1, it was absent from all analyzed passages of culture 2. Since splitting of both cultures occurred at passage 19, L1 mobilization resulting in L1-dn13 *de novo* insertion must have occurred in culture 1 between passages 19 and 43.

Figure 4: *De novo* full-length L1 insertions retain retrotransposition competency *in vitro*.

Intact, full-length L1 insertions L1-dn6-2.2 and L1-dn6-5.4 were obtained from two independent genomic PCR reactions amplifying the L1-dn6 *de novo* insertion (Fig. 3a, Supplementary Fig. S9 and S10). They were tagged with an *mblast1* retrotransposition indicator cassette⁴⁷, inserted into an episomal expression plasmid and set under transcriptional control of the CMV promoter. To test for the retrotransposition-competence of the L1-dn6 derived PCR products, the resulting L1 reporter plasmids pJJ101/L1-dn6_2.2 and pJJ101/L1-dn6_5.4 were submitted to an L1 retrotransposition reporter assay described previously^{47,53}. Only cells in which mobilization of the L1 reporter takes place can express the functional BLAST gene and grow in the presence of Blastidicin-S. HeLa cells were transfected

with the L1-dn6 reporter plasmids or with positive and negative control L1 reporter plasmids pJJ101/L1.3⁷⁵ and pJJ101/L1.3-D702A^{53,75}, respectively. pJJ101/L1.3 was used for normalization (100% activity). pJJ101/L1.3-D702A contains a single point mutation in the L1 reverse transcriptase domain^{47,53}. The bar diagram depicts arithmetic mean \pm s.d of three independent retrotransposition reporter assays of the engineered L1-dn6 elements relative to L1.3. Black hexagon, SV40 polyadenylation signal; grey arrows, TSDs flanking a 5'-truncated *de novo* L1 insertion. Blast(s), Blastidicin-S sensitive; Blast(r), Blastidicin-S resistant; SD, splice donor; SA, splice acceptor.

Figure 5: L1-dn13 affects *CADPS2* expression

a) Schematic of the human *CADPS2* allele of the hiPS-SB4 line harboring insertion L1-dn13 (red box). A *CADPS2* transcript including nucleotide sequences of exons 7, 8, 27 and 28 is presented. Binding sites of the TaqMan primer/probe combinations spanning the exon27/exon28 junction on the *CADPS2* cDNA used for qRT-PCR analysis are shown (red arrows and line).

b) Reduction in *CADPS2* expression in a late versus early passage hiPS-SB4 culture characterized by an increase in the cell population harboring L1-dn13 (Fig. 3b). Relative *CADPS2* mRNA levels in early (p16) and late passage (p50) hiPS-SB4 cells were assessed by qRT-PCR using cytoplasmic mRNA. HES-3 and hFF-iPS4 cells were used as positive controls due to the established expression of *CADPS2* in human pluripotent stem cells⁵⁸. qRT-PCR results were normalized to 18s rRNA using *CADPS2* expression in parental HFF-1 cells as the control.

c) Genomic structure of the L1-dn13 integration site in *CADPS2* gene in hiPS-SB4 subclones. hiPS-SB4_D differs from hiPS-SB4_B by the presence of the L1-dn13 *de novo* insertion (red box) in intron 7 of the *CADPS2* gene. Binding sites of L1-dn13-specific validation PCR primers OP1, ISP1 and ISP2 (Supplementary Tables S3 and S4) and the expected lengths of the resulting PCR products are indicated. Black diamonds, TSDs flanking L1-dn13 insertion.

d) Genotyping PCR validating the presence of L1-dn13 in subclone hiPS-SB4_D and its absence from hiPS-SB4_B. gDNAs used were isolated from HFF-1, hiPS-SB4_B, and hiPS-SB4_D cells and from the original mixed population of the hiPS-SB4 culture (hiPS-SB4 [Mix]). Primer combinations used are indicated in blue;

H₂O.ISP1/ISP2 and H₂O.ISP1/OP1, negative control PCRs using H₂O instead of gDNA; 100-bp ladder, size marker.

e) qPCR analyses confirming the absence of L1-dn13 from cultivated hiPS-SB4_B and HFF-1 cells, and its presence in hiPS-SB4_D cells and in the hiPS-SB4 culture. The same gDNA preparations used for genotyping PCR (Fig. 5d) were utilized. Applied primer/probe combinations are specific for the 5' junction of L1-dn13 and described in Figure 3b. Genomic DNAs from HFF-1 cells lacking L1-dn13 and from hiPS-SB4 cells including a subpopulation of L1-dn13 carrying cells served as negative and positive controls, respectively. For normalization, a primer/probe combination specific for the human single-copy gene RPP25 was used. $\Delta\Delta C_t$ values measured the relative quantity of L1-dn13. gDNA from HFF-1 cells served as negative control. Bars indicate arithmetic means \pm SEM of technical triplicates.

f) L1-dn13 in hiPS-SB4_B interferes with *CADPS2* expression. Relative *CADPS2* mRNA levels in hiPS-SB4_B, hiPS-SB4_D and hiPS-SB4 (Mix) cells were determined by qRT-PCR using cytoplasmic RNA and primer/probe combinations spanning the exon 27/exon 28 junction of the *CADPS2* cDNA as outlined in the schematic of Fig. 5a. Only a subpopulation of hiPSCs of the hiPS-SB4 line carried L1-dn13. Bars indicate arithmetic means \pm SEM of technical triplicates.

REFERENCIAS BIBLIOGRÁFICAS

1. Lodish H, Berk A, Matsudaira P, Kaiser CA, Krieger M, Scott MP *et al.* *Molecular Cell Biology*, 2003.
2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001; **409**(6822): 860-921.
3. Kersanach R, Brinkmann H, Liaud MF, Zhang DX, Martin W & Cerff R. Five identical intron positions in ancient duplicated genes of eubacterial origin. *Nature* 1994; **367**(6461): 387-389.
4. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG *et al.* The sequence of the human genome. *Science* 2001; **291**(5507): 1304-1351.
5. Mc CB. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A* 1950; **36**(6): 344-355.
6. Munoz-Lopez M & Garcia-Perez JL. DNA transposons: nature and applications in genomics. *Curr Genomics* 2010; **11**(2): 115-128.
7. Richardson JM, Dawson A, O'Hagan N, Taylor P, Finnegan DJ & Walkinshaw MD. Mechanism of Mos1 transposition: insights from structural analysis. *EMBO J* 2006; **25**(6): 1324-1334.
8. Smit AF. The origin of interspersed repeats in the human genome. *Curr Opin Genet Dev* 1996; **6**(6): 743-748.
9. Ivics Z, Hackett PB, Plasterk RH & Izsvak Z. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* 1997; **91**(4): 501-510.
10. Ray DA, Feschotte C, Pagan HJ, Smith JD, Pritham EJ, Arensburger P *et al.* Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res* 2008; **18**(5): 717-728.
11. Mitra R, Li X, Kapusta A, Mayhew D, Mitra RD, Feschotte C *et al.* Functional characterization of piggyBat from the bat *Myotis lucifugus* unveils an active mammalian DNA transposon. *Proc Natl Acad Sci U S A* 2013; **110**(1): 234-239.
12. Pace JK, 2nd & Feschotte C. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res* 2007; **17**(4): 422-432.

13. Cameron JR, Loh EY & Davis RW. Evidence for transposition of dispersed repetitive DNA families in yeast. *Cell* 1979; **16**(4): 739-751.
14. Copeland CS, Mann VH, Morales ME, Kalinna BH & Brindley PJ. The Sinbad retrotransposon from the genome of the human blood fluke, *Schistosoma mansoni*, and the distribution of related Pao-like elements. *BMC Evol Biol* 2005; **5**: 20.
15. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B *et al.* A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* 2007; **8**(12): 973-982.
16. Boeke JD. LINEs and Alus--the polyA connection. *Nat Genet* 1997; **16**(1): 6-7.
17. Kazazian HH, Jr. Mobile elements and disease. *Curr Opin Genet Dev* 1998; **8**(3): 343-350.
18. Bannert N & Kurth R. The evolutionary dynamics of human endogenous retroviral families. *Annu Rev Genomics Hum Genet* 2006; **7**: 149-173.
19. Malik HS, Henikoff S & Eickbush TH. Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res* 2000; **10**(9): 1307-1318.
20. Medstrand P & Mager DL. Human-specific integrations of the HERV-K endogenous retrovirus family. *J Virol* 1998; **72**(12): 9782-9787.
21. Korenberg JR & Rykowski MC. Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. *Cell* 1988; **53**(3): 391-400.
22. Smit AF, Toth G, Riggs AD & Jurka J. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J Mol Biol* 1995; **246**(3): 401-417.
23. Eickbush TH. Repair by retrotransposition. *Nat Genet* 2002; **31**(2): 126-127.
24. Khan H, Smit A & Boissinot S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res* 2006; **16**(1): 78-87.
25. Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE *et al.* LINE-1 retrotransposition activity in human genomes. *Cell* 2010; **141**(7): 1159-1170.
26. Sassaman DM, Dombroski BA, Moran JV, Kimberland ML, Naas TP, DeBerardinis RJ *et al.* Many human L1 elements are capable of retrotransposition. *Nat Genet* 1997; **16**(1): 37-43.

27. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV *et al.* Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A* 2003; **100**(9): 5280-5285.
28. Skowronski J, Fanning TG & Singer MF. Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol Cell Biol* 1988; **8**(4): 1385-1397.
29. Boissinot S, Chevret P & Furano AV. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol Biol Evol* 2000; **17**(6): 915-928.
30. Beck CR, Garcia-Perez JL, Badge RM & Moran JV. LINE-1 elements in structural variation and disease. *Annu Rev Genomics Hum Genet* 2011; **12**: 187-215.
31. Dombroski BA, Mathias SL, Nanthakumar E, Scott AF & Kazazian HH, Jr. Isolation of an active human transposable element. *Science* 1991; **254**(5039): 1805-1808.
32. Scott AF, Schmeckpeper BJ, Abdelrazik M, Comey CT, O'Hara B, Rossiter JP *et al.* Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. *Genomics* 1987; **1**(2): 113-125.
33. Bourc'his D & Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 2004; **431**(7004): 96-99.
34. Swergold GD. Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol Cell Biol* 1990; **10**(12): 6718-6729.
35. Speek M. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol* 2001; **21**(6): 1973-1985.
36. Li J, Kannan M, Trivett AL, Liao H, Wu X, Akagi K *et al.* An antisense promoter in mouse L1 retrotransposon open reading frame-1 initiates expression of diverse fusion transcripts and limits retrotransposition. *Nucleic Acids Res* 2014; **42**(7): 4546-4562.
37. Ciaudo C, Jay F, Okamoto I, Chen CJ, Sarazin A, Servant N *et al.* RNAi-dependent and independent control of LINE1 accumulation and mobility in mouse embryonic stem cells. *PLoS Genet* 2013; **9**(11): e1003791.
38. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD & Kazazian HH, Jr. High frequency retrotransposition in cultured mammalian cells. *Cell* 1996; **87**(5): 917-927.

39. Feng Q, Moran JV, Kazazian HH, Jr. & Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 1996; **87**(5): 905-916.
40. Denli AM, Narvaiza I, Kerman BE, Pena M, Benner C, Marchetto MC *et al.* Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. *Cell* 2015; **163**(3): 583-593.
41. Hohjoh H & Singer MF. Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J* 1996; **15**(3): 630-639.
42. Kolosha VO & Martin SL. High-affinity, non-sequence-specific RNA binding by the open reading frame 1 (ORF1) protein from long interspersed nuclear element 1 (LINE-1). *J Biol Chem* 2003; **278**(10): 8112-8117.
43. Martin SL & Bushman FD. Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol Cell Biol* 2001; **21**(2): 467-475.
44. Holmes SE, Singer MF & Swergold GD. 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* 1992; **267**(28): 19765-19768.
45. Ergun S, Buschmann C, Heukeshoven J, Dammann K, Schnieders F, Lauke H *et al.* Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues. *J Biol Chem* 2004; **279**(26): 27753-27763.
46. Doucet AJ, Hulme AE, Sahinovic E, Kulpa DA, Moldovan JB, Kopera HC *et al.* Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet* 2010; **6**(10).
47. Mathias SL, Scott AF, Kazazian HH, Jr., Boeke JD & Gabriel A. Reverse transcriptase encoded by a human transposable element. *Science* 1991; **254**(5039): 1808-1810.
48. Fanning T & Singer M. The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. *Nucleic Acids Res* 1987; **15**(5): 2251-2260.
49. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM *et al.* The regulated retrotransposon transcriptome of mammalian cells. *Nat Genet* 2009; **41**(5): 563-571.
50. Kazazian HH, Jr. & Moran JV. The impact of L1 retrotransposons on the human genome. *Nat Genet* 1998; **19**(1): 19-24.

51. Kurose K, Hata K, Hattori M & Sakaki Y. RNA polymerase III dependence of the human L1 promoter and possible participation of the RNA polymerase II factor YY1 in the RNA polymerase III transcription system. *Nucleic Acids Res* 1995; **23**(18): 3704-3709.
52. Schumann G, Zundorf I, Hofmann J, Marschalek R & Dingermann T. Internally located and oppositely oriented polymerase II promoters direct convergent transcription of a LINE-like retroelement, the Dictyostelium repetitive element, from Dictyostelium discoideum. *Mol Cell Biol* 1994; **14**(5): 3074-3084.
53. Wilkinson DA, Mager DL & Leong J-AC. Endogenous Human Retroviruses. *Springer US* 1994(9): 465-535.
54. Mizrokhi LJ, Georgieva SG & Ilyin YV. jockey, a mobile Drosophila element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. *Cell* 1988; **54**(5): 685-691.
55. Dmitriev SE, Andreev DE, Terenin IM, Olovnikov IA, Prassolov VS, Merrick WC *et al.* Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5' untranslated region of the human retrotransposon LINE-1 mRNA is strictly cap dependent rather than internal ribosome entry site mediated. *Mol Cell Biol* 2007; **27**(13): 4685-4697.
56. Kojima KK, Matsumoto T & Fujiwara H. Eukaryotic translational coupling in UAAUG stop-start codons for the bicistronic RNA translation of the non-long terminal repeat retrotransposon SART1. *Mol Cell Biol* 2005; **25**(17): 7675-7686.
57. Alisch RS, Garcia-Perez JL, Muotri AR, Gage FH & Moran JV. Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev* 2006; **20**(2): 210-224.
58. Seleme MC, Disson O, Robin S, Brun C, Teninges D & Bucheton A. In vivo RNA localization of I factor, a non-LTR retrotransposon, requires a cis-acting signal in ORF2 and ORF1 protein. *Nucleic Acids Res* 2005; **33**(2): 776-785.
59. Kinsey JA. Tad, a LINE-like transposable element of Neurospora, can transpose between nuclei in heterokaryons. *Genetics* 1990; **126**(2): 317-323.
60. Kubo S, Seleme MC, Soifer HS, Perez JL, Moran JV, Kazazian HH, Jr. *et al.* L1 retrotransposition in nondividing and primary human somatic cells. *Proc Natl Acad Sci U S A* 2006; **103**(21): 8036-8041.
61. Christensen SM, Ye J & Eickbush TH. RNA from the 5' end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. *Proc Natl Acad Sci U S A* 2006; **103**(47): 17602-17607.

62. Yang J, Mohr G, Perlman PS & Lambowitz AM. Group II intron mobility in yeast mitochondria: target DNA-primed reverse transcription activity of a1 and reverse splicing into DNA transposition sites in vitro. *J Mol Biol* 1998; **282**(3): 505-523.
63. Zingler N, Willhoeft U, Brose HP, Schoder V, Jahns T, Hanschmann KM *et al.* Analysis of 5' junctions of human LINE-1 and Alu retrotransposons suggests an alternative model for 5'-end attachment requiring microhomology-mediated end-joining. *Genome Res* 2005; **15**(6): 780-789.
64. Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G *et al.* Human l1 retrotransposition is associated with genetic instability in vivo. *Cell* 2002; **110**(3): 327-338.
65. Blesa D & Martinez-Sebastian MJ. bilbo, a non-LTR retrotransposon of *Drosophila subobscura*: a clue to the evolution of LINE-like elements in *Drosophila*. *Mol Biol Evol* 1997; **14**(11): 1145-1153.
66. Malik HS, Burke WD & Eickbush TH. The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol* 1999; **16**(6): 793-805.
67. Kurzynska-Kokorniak A, Jamburuthugoda VK, Bibillo A & Eickbush TH. DNA-directed DNA polymerase and strand displacement activity of the reverse transcriptase encoded by the R2 retrotransposon. *J Mol Biol* 2007; **374**(2): 322-333.
68. Han JS. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. *Mob DNA* 2010; **1**(1): 15.
69. Szak ST, Pickeral OK, Makalowski W, Boguski MS, Landsman D & Boeke JD. Molecular archeology of L1 insertions in the human genome. *Genome Biol* 2002; **3**(10): research0052.
70. Dewannieux M, Esnault C & Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 2003; **35**(1): 41-48.
71. Hancks DC & Kazazian HH, Jr. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev* 2012; **22**(3): 191-203.
72. Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retrotransposons. *Proc Natl Acad Sci U S A* 1997; **94**(5): 1872-1877.
73. Richardson SR, Doucet AJ, Kopera HC, Moldovan JB, Garcia-Perez JL & Moran JV. The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. *Microbiol Spectr* 2015; **3**(2).

74. Batzer MA & Deininger PL. Alu repeats and human genomic diversity. *Nat Rev Genet* 2002; **3**(5): 370-379.
75. Bennett EA, Keller H, Mills RE, Schmidt S, Moran JV, Weichenrieder O *et al.* Active Alu retrotransposons in the human genome. *Genome Res* 2008; **18**(12): 1875-1883.
76. Rubin CM, Houck CM, Deininger PL, Friedmann T & Schmid CW. Partial nucleotide sequence of the 300-nucleotide interspersed repeated human DNA sequences. *Nature* 1980; **284**(5754): 372-374.
77. Ullu E & Tschudi C. Alu sequences are processed 7SL RNA genes. *Nature* 1984; **312**(5990): 171-172.
78. Sarrowa J, Chang DY & Maraia RJ. The decline in human Alu retroposition was accompanied by an asymmetric decrease in SRP9/14 binding to dimeric Alu RNA and increased expression of small cytoplasmic Alu RNA. *Mol Cell Biol* 1997; **17**(3): 1144-1151.
79. Chang DY, Sasaki-Tozawa N, Green LK & Maraia RJ. A trinucleotide repeat-associated increase in the level of Alu RNA-binding protein occurred during the same period as the major Alu amplification that accompanied anthropoid evolution. *Mol Cell Biol* 1995; **15**(4): 2109-2116.
80. Raiz J, Damert A, Chira S, Held U, Klawitter S, Hamdorf M *et al.* The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Res* 2011; **40**(4): 1666-1683.
81. Ostertag EM, Goodier JL, Zhang Y & Kazazian HH, Jr. SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am J Hum Genet* 2003; **73**(6): 1444-1451.
82. Hancks DC & Kazazian HH, Jr. SVA retrotransposons: Evolution and genetic instability. *Semin Cancer Biol* 2010; **20**(4): 234-245.
83. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA *et al.* SVA elements: a hominid-specific retroposon family. *J Mol Biol* 2005; **354**(4): 994-1007.
84. Hancks DC, Ewing AD, Chen JE, Tokunaga K & Kazazian HH, Jr. Exon-trapping mediated by the human retrotransposon SVA. *Genome Res* 2009; **19**(11): 1983-1991.
85. Ono M, Kawakami M & Takezawa T. A novel human nonviral retroposon derived from an endogenous retrovirus. *Nucleic Acids Res* 1987; **15**(21): 8725-8737.

86. Shen L, Wu LC, Sanlioglu S, Chen R, Mendoza AR, Dangel AW *et al.* Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region. Molecular cloning, exon-intron structure, composite retroposon, and breakpoint of gene duplication. *J Biol Chem* 1994; **269**(11): 8466-8476.
87. Savage AL, Bubb VJ, Breen G & Quinn JP. Characterisation of the potential function of SVA retrotransposons to modulate gene expression patterns. *BMC Evol Biol* 2013; **13**: 101.
88. Jacobs FM, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S *et al.* An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* 2014; **516**(7530): 242-245.
89. Maestre J, Tchenio T, Dhellin O & Heidmann T. mRNA retroposition in human cells: processed pseudogene formation. *EMBO J* 1995; **14**(24): 6333-6338.
90. Esnault C, Maestre J & Heidmann T. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* 2000; **24**(4): 363-367.
91. Vanin EF. Processed pseudogenes: characteristics and evolution. *Annu Rev Genet* 1985; **19**: 253-272.
92. Wei W, Gilbert N, Ooi SL, Lawler JF, Ostertag EM, Kazazian HH *et al.* Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol* 2001; **21**(4): 1429-1439.
93. Snider L, Geng LN, Lemmers RJ, Kyba M, Ware CB, Nelson AM *et al.* Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genet* 2010; **6**(10): e1001181.
94. Parker HG, VonHoldt BM, Quignon P, Margulies EH, Shao S, Mosher DS *et al.* An expressed *fgf4* retrogene is associated with breed-defining chondrodysplasia in domestic dogs. *Science* 2009; **325**(5943): 995-998.
95. Buzdin A, Gogvadze E, Kovalskaya E, Volchkov P, Ustyugova S, Illarionova A *et al.* The human genome contains many types of chimeric retrogenes generated through in vivo RNA recombination. *Nucleic Acids Res* 2003; **31**(15): 4385-4390.
96. Weber MJ. Mammalian small nucleolar RNAs are mobile genetic elements. *PLoS Genet* 2006; **2**(12): e205.
97. Branciforte D & Martin SL. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Mol Cell Biol* 1994; **14**(4): 2584-2592.

98. Seleme MC, Busseau I, Malinsky S, Bucheton A & Teninges D. High-frequency retrotransposition of a marked I factor in *Drosophila melanogaster* correlates with a dynamic expression pattern of the ORF1 protein in the cytoplasm of oocytes. *Genetics* 1999; **151**(2): 761-771.
99. Kidd SJ & Glover DM. *Drosophila melanogaster* ribosomal DNA containing type II insertions is variably transcribed in different strains and tissues. *J Mol Biol* 1981; **151**(4): 645-662.
100. Kazazian HH, Jr. Mobile elements: drivers of genome evolution. *Science* 2004; **303**(5664): 1626-1632.
101. Kazazian HH, Jr. An estimated frequency of endogenous insertional mutations in humans. *Nat Genet* 1999; **22**(2): 130.
102. Martin SL & Branciforte D. Synchronous expression of LINE-1 RNA and protein in mouse embryonal carcinoma cells. *Mol Cell Biol* 1993; **13**(9): 5383-5392.
103. Ostertag EM, DeBerardinis RJ, Goodier JL, Zhang Y, Yang N, Gerton GL *et al.* A mouse model of human L1 retrotransposition. *Nat Genet* 2002; **32**(4): 655-660.
104. Prak ET, Dodson AW, Farkash EA & Kazazian HH, Jr. Tracking an embryonic L1 retrotransposition event. *Proc Natl Acad Sci U S A* 2003; **100**(4): 1832-1837.
105. Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV & Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 2005; **435**(7044): 903-910.
106. Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM *et al.* L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev* 2009; **23**(11): 1303-1312.
107. Freeman P, Macfarlane C, Collier P, Jeffreys AJ & Badge RM. L1 hybridization enrichment: a method for directly accessing de novo L1 insertions in the human germline. *Hum Mutat* 2011; **32**(8): 978-988.
108. Malki S, van der Heijden GW, O'Donnell KA, Martin SL & Bortvin A. A role for retrotransposon LINE-1 in fetal oocyte attrition in mice. *Dev Cell* 2014; **29**(5): 521-533.
109. Garcia-Perez JL, Marchetto MC, Muotri AR, Coufal NG, Gage FH, O'Shea KS *et al.* LINE-1 retrotransposition in human embryonic stem cells. *Hum Mol Genet* 2007; **16**(13): 1569-1577.
110. Garcia-Perez JL, Morell M, Scheys JO, Kulpa DA, Morell S, Carter CC *et al.* Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature* 2010; **466**(7307): 769-773.

111. Macia A, Munoz-Lopez M, Cortes JL, Hastings RK, Morell S, Lucena-Aguilar G *et al.* Epigenetic control of retrotransposon expression in human embryonic stem cells. *Mol Cell Biol* 2011; **31**(2): 300-316.
112. Inoue H, Nagata N, Kurokawa H & Yamanaka S. iPS cells: a game changer for future medicine. *EMBO J* 2014; **33**(5): 409-417.
113. Wissing S, Munoz-Lopez M, Macia A, Yang Z, Montano M, Collins W *et al.* Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. *Hum Mol Genet* 2012; **21**(1): 208-218.
114. Macia A, Blanco-Jimenez E & Garcia-Perez JL. Retrotransposons in pluripotent cells: Impact and new roles in cellular plasticity. *Biochim Biophys Acta* 2014; **1849**(4): 417-426.
115. Klawitter S, Fuchs N, Upton KR, Muñoz-Lopez M, Shukla R, Wang J *et al.* Reprogramming triggers endogenous L1, Alu and SVA retrotransposition in human induced pluripotent stem cells. *Nat Commun* 2015; **Accepted**.
116. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* 1998; **282**(5391): 1145-1147.
117. van den Hurk JA, van de Pol DJ, Wissinger B, van Driel MA, Hoefsloot LH, de Wijs IJ *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* 2003; **113**(3): 268-275.
118. Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW *et al.* Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res* 1992; **52**(3): 643-645.
119. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT *et al.* L1 retrotransposition in human neural progenitor cells. *Nature* 2009; **460**(7259): 1127-1131.
120. Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, Lehmann HS *et al.* Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* 2012; **151**(3): 483-496.
121. Upton KR, Gerhardt DJ, Jesuadian JS, Richardson SR, Sanchez-Luque FJ, Bodea GO *et al.* Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* 2015; **161**(2): 228-239.
122. Chenais B, Caruso A, Hiard S & Casse N. The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. *Gene* 2012; **509**(1): 7-15.

123. Hulme AE, Kulpa DA, Perez JLG & Moran JV. The Impact of LINE-1 Retro transposition on the Human Genome. *Humana Press* 2006: 35-55.
124. Biemont C & Vieira C. Genetics: junk DNA as an evolutionary force. *Nature* 2006; **443**(7111): 521-524.
125. Miller WJ, McDonald JF & Pinsker W. Molecular domestication of mobile elements. *Genetica* 1997; **100**(1-3): 261-270.
126. Moran JV, DeBerardinis RJ & Kazazian HH, Jr. Exon shuffling by L1 retrotransposition. *Science* 1999; **283**(5407): 1530-1534.
127. Cowley M, Wood AJ, Bohm S, Schulz R & Oakey RJ. Epigenetic control of alternative mRNA processing at the imprinted Herc3/Nap115 locus. *Nucleic Acids Res* 2012; **40**(18): 8917-8926.
128. Taube JH, Allton K, Duncan SA, Shen L & Barton MC. Foxa1 functions as a pioneer transcription factor at transposable elements to activate Afp during differentiation of embryonic stem cells. *J Biol Chem*; **285**(21): 16135-16144.
129. Levin HL & Moran JV. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet* 2011; **12**(9): 615-627.
130. Han JS & Boeke JD. A highly active synthetic mammalian retrotransposon. *Nature* 2004; **429**(6989): 314-318.
131. Simons C, Pheasant M, Makunin IV & Mattick JS. Transposon-free regions in mammalian genomes. *Genome Res* 2006; **16**(2): 164-172.
132. Sironi M, Menozzi G, Comi GP, Cereda M, Cagliani R, Bresolin N *et al.* Gene function and expression level influence the insertion/fixation dynamics of distinct transposon families in mammalian introns. *Genome Biol* 2006; **7**(12): R120.
133. Ustyugova SV, Lebedev YB & Sverdlov ED. Long L1 insertions in human gene introns specifically reduce the content of corresponding primary transcripts. *Genetica* 2006; **128**(1-3): 261-272.
134. Chen J, Rattner A & Nathans J. 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* 2006; **15**(13): 2146-2156.
135. Goodier JL, Mandal PK, Zhang L & Kazazian HH, Jr. Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion. *Hum Mol Genet* 2010; **19**(9): 1712-1725.

136. Tubio JM, Li Y, Ju YS, Martincorena I, Cooke SL, Tojo M *et al.* Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* 2014; **345**(6196): 1251343.
137. Pickeral OK, Makalowski W, Boguski MS & Boeke JD. Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res* 2000; **10**(4): 411-415.
138. Zhang XH & Chasin LA. Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. *Proc Natl Acad Sci U S A* 2006; **103**(36): 13427-13432.
139. Sorek R. The birth of new exons: mechanisms and evolutionary consequences. *RNA* 2007; **13**(10): 1603-1608.
140. Sela N, Mersch B, Gal-Mark N, Lev-Maor G, Hotz-Wagenblatt A & Ast G. Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. *Genome Biol* 2007; **8**(6): R127.
141. Sela N, Mersch B, Hotz-Wagenblatt A & Ast G. Characteristics of transposable element exonization within human and mouse. *PLoS One* 2010; **5**(6): e10907.
142. Han JS & Boeke JD. LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? *Bioessays* 2005; **27**(8): 775-784.
143. Morgan HD, Sutherland HG, Martin DI & Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* 1999; **23**(3): 314-318.
144. Nigumann P, Redik K, Matlik K & Speek M. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics* 2002; **79**(5): 628-634.
145. Singer T, McConnell MJ, Marchetto MC, Coufal NG & Gage FH. LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends Neurosci* 2010; **33**(8): 345-354.
146. Wheelan SJ, Aizawa Y, Han JS & Boeke JD. Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome Res* 2005; **15**(8): 1073-1078.
147. Perepelitsa-Belancio V & Deininger P. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat Genet* 2003; **35**(4): 363-366.

148. Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, de Rooij DG *et al.* MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* 2007; **12**(4): 503-514.
149. Soper SF, van der Heijden GW, Hardiman TC, Goodheart M, Martin SL, de Boer P *et al.* Mouse maelstrom, a component of nuage, is essential for spermatogenesis and transposon repression in meiosis. *Dev Cell* 2008; **15**(2): 285-297.
150. Slotkin RK & Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 2007; **8**(4): 272-285.
151. Yoder JA, Walsh CP & Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 1997; **13**(8): 335-340.
152. Bestor TH & Bourc'his D. Transposon silencing and imprint establishment in mammalian germ cells. *Cold Spring Harb Symp Quant Biol* 2004; **69**: 381-387.
153. Wood AJ, Schulz R, Woodfine K, Koltowska K, Beechey CV, Peters J *et al.* Regulation of alternative polyadenylation by genomic imprinting. *Genes Dev* 2008; **22**(9): 1141-1146.
154. Rangwala SH, Zhang L & Kazazian HH, Jr. Many LINE1 elements contribute to the transcriptome of human somatic cells. *Genome Biol* 2009; **10**(9): R100.
155. Puszyk W, Down T, Grimwade D, Chomienne C, Oakey RJ, Solomon E *et al.* The epigenetic regulator PLZF represses L1 retrotransposition in germ and progenitor cells. *EMBO J* 2013; **32**(13): 1941-1952.
156. Munoz-Lopez M, Macia A, Garcia-Canadas M, Badge RM & Garcia-Perez JL. An epi [c] genetic battle: LINE-1 retrotransposons and intragenomic conflict in humans. *Mob Genet Elements* 2011; **1**(2): 122-127.
157. Shen Y, Chow J, Wang Z & Fan G. Abnormal CpG island methylation occurs during in vitro differentiation of human embryonic stem cells. *Hum Mol Genet* 2006; **15**(17): 2623-2635.
158. Reichmann J, Reddington JP, Best D, Read D, Ollinger R, Meehan RR *et al.* The genome-defence gene *Tex19.1* suppresses LINE-1 retrotransposons in the placenta and prevents intra-uterine growth retardation in mice. *Hum Mol Genet* 2013; **22**(9): 1791-1806.
159. Kuwabara T, Hsieh J, Muotri A, Yeo G, Warashina M, Lie DC *et al.* Wnt-mediated activation of *NeuroD1* and retro-elements during adult neurogenesis. *Nat Neurosci* 2009; **12**(9): 1097-1105.

160. Muotri AR, Marchetto MC, Coufal NG, Oefner R, Yeo G, Nakashima K *et al.* L1 retrotransposition in neurons is modulated by MeCP2. *Nature* 2010; **468**(7322): 443-446.
161. Lorincz MC, Dickerson DR, Schmitt M & Groudine M. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat Struct Mol Biol* 2004; **11**(11): 1068-1075.
162. Buckley RM & Adelson DL. Mammalian genome evolution as a result of epigenetic regulation of transposable elements. *Biomol Concepts* 2014; **5**(3): 183-194.
163. Smith ZD & Meissner A. DNA methylation: roles in mammalian development. *Nat Rev Genet* 2013; **14**(3): 204-220.
164. Hajkova P, Ancelin K, Waldmann T, Lacoste N, Lange UC, Cesari F *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* 2008; **452**(7189): 877-881.
165. Chen J, Liu H, Liu J, Qi J, Wei B, Yang J *et al.* H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat Genet* 2013; **45**(1): 34-42.
166. Aravin AA, Sachidanandam R, Bourc'his D, Schaefer C, Pezic D, Toth KF *et al.* A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* 2008; **31**(6): 785-799.
167. Girard A & Hannon GJ. Conserved themes in small-RNA-mediated transposon control. *Trends Cell Biol* 2008; **18**(3): 136-148.
168. Hammond SM. Dicing and slicing: the core machinery of the RNA interference pathway. *FEBS Lett* 2005; **579**(26): 5822-5829.
169. Faulkner GJ. Retrotransposon silencing during embryogenesis: dicer cuts in LINE. *PLoS Genet* 2013; **9**(11): e1003944.
170. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE & Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 1998; **391**(6669): 806-811.
171. Tabara H, Sarkissian M, Kelly WG, Fleenor J, Grishok A, Timmons L *et al.* The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* 1999; **99**(2): 123-132.
172. Ketting RF, Haverkamp TH, van Luenen HG & Plasterk RH. Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD. *Cell* 1999; **99**(2): 133-141.

173. Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B *et al.* The small RNA profile during *Drosophila melanogaster* development. *Dev Cell* 2003; **5**(2): 337-350.
174. Chung WJ, Okamura K, Martin R & Lai EC. Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. *Curr Biol* 2008; **18**(11): 795-802.
175. Hutvagner G & Simard MJ. Argonaute proteins: key players in RNA silencing. *Nat Rev Mol Cell Biol* 2008; **9**(1): 22-32.
176. Siomi MC, Sato K, Pezic D & Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* 2011; **12**(4): 246-258.
177. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M *et al.* DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev* 2008; **22**(7): 908-917.
178. Heras SR, Macias S, Caceres JF & Garcia-Perez JL. Control of mammalian retrotransposons by cellular RNA processing activities. *Mob Genet Elements* 2014; **4**: e28439.
179. Heras SR, Macias S, Plass M, Fernandez N, Cano D, Eyraas E *et al.* The Microprocessor controls the activity of mammalian retrotransposons. *Nat Struct Mol Biol* 2013; **20**(10): 1173-1181.
180. Moran VA, Perera RJ & Khalil AM. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res* 2012; **40**(14): 6391-6400.
181. Hall LL, Carone DM, Gomez AV, Kolpa HJ, Byron M, Mehta N *et al.* Stable COT-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell* 2014; **156**(5): 907-919.
182. Denoeud F, Kapranov P, Ucla C, Frankish A, Castelo R, Drenkow J *et al.* Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions. *Genome Res* 2007; **17**(6): 746-759.
183. Belancio VP, Hedges DJ & Deininger P. LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res* 2006; **34**(5): 1512-1521.
184. Goodier JL, Cheung LE & Kazazian HH, Jr. Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. *Nucleic Acids Res* 2013; **41**(15): 7401-7419.

185. Peddigari S, Li PW, Rabe JL & Martin SL. hnRNPL and nucleolin bind LINE-1 RNA and function as host factors to modulate retrotransposition. *Nucleic Acids Res*; **41**(1): 575-585.
186. Dai L, Taylor MS, O'Donnell KA & Boeke JD. Poly(A) binding protein C1 is essential for efficient L1 retrotransposition and affects L1 RNP formation. *Mol Cell Biol* 2012; **32**(21): 4323-4336.
187. Zhang A, Dong B, Doucet AJ, Moldovan JB, Moran JV & Silverman RH. RNase L restricts the mobility of engineered retrotransposons in cultured human cells. *Nucleic Acids Res* 2014; **42**(6): 3803-3820.
188. Wissing S, Montano M, Garcia-Perez JL, Moran JV & Greene WC. Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. *J Biol Chem* 2011; **286**(42): 36427-36437.
189. Chiu YL & Greene WC. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. *Annu Rev Immunol* 2008; **26**: 317-353.
190. Mangeat B, Turelli P, Caron G, Friedli M, Perrin L & Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* 2003; **424**(6944): 99-103.
191. Zhang Z, Harrison PM, Liu Y & Gerstein M. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. *Genome Res* 2003; **13**(12): 2541-2558.
192. Dutko JA, Schafer A, Kenny AE, Cullen BR & Curcio MJ. Inhibition of a yeast LTR retrotransposon by human APOBEC3 cytidine deaminases. *Curr Biol* 2005; **15**(7): 661-666.
193. Kinomoto M, Kanno T, Shimura M, Ishizaka Y, Kojima A, Kurata T *et al.* All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res* 2007; **35**(9): 2955-2964.
194. MacDuff DA, Demorest ZL & Harris RS. AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. *Nucleic Acids Res* 2009; **37**(6): 1854-1867.
195. Bogerd HP, Wiegand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV *et al.* Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc Natl Acad Sci U S A* 2006; **103**(23): 8780-8785.
196. Richardson SR, Narvaiza I, Planegger RA, Weitzman MD & Moran JV. APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. *Elife* 2014; **3**: e02008.

197. Goodier JL, Cheung LE & Kazazian HH, Jr. MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. *PLoS Genet* 2012; **8**(10): e1002941.
198. Luo K, Wang T, Liu B, Tian C, Xiao Z, Kappes J *et al.* Cytidine deaminases APOBEC3G and APOBEC3F interact with human immunodeficiency virus type 1 integrase and inhibit proviral DNA formation. *J Virol* 2007; **81**(13): 7238-7248.
199. Izumi T, Burdick R, Shigemi M, Plisov S, Hu WS & Pathak VK. Mov10 and APOBEC3G localization to processing bodies is not required for virion incorporation and antiviral activity. *J Virol* 2013; **87**(20): 11047-11062.
200. Zheng K, Xiol J, Reuter M, Eckardt S, Leu NA, McLaughlin KJ *et al.* Mouse MOV10L1 associates with Piwi proteins and is an essential component of the Piwi-interacting RNA (piRNA) pathway. *Proc Natl Acad Sci U S A* 2010; **107**(26): 11841-11846.
201. Frost RJ, Hamra FK, Richardson JA, Qi X, Bassel-Duby R & Olson EN. MOV10L1 is necessary for protection of spermatocytes against retrotransposons by Piwi-interacting RNAs. *Proc Natl Acad Sci U S A* 2010; **107**(26): 11847-11852.
202. Stetson DB, Ko JS, Heidmann T & Medzhitov R. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* 2008; **134**(4): 587-598.
203. Zhao K, Du J, Han X, Goodier JL, Li P, Zhou X *et al.* Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutieres syndrome-related SAMHD1. *Cell Rep* 2013; **4**(6): 1108-1115.
204. Karimi MM, Goyal P, Maksakova IA, Bilenky M, Leung D, Tang JX *et al.* DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* 2011; **8**(6): 676-687.
205. Marchetto MC, Narvaiza I, Denli AM, Benner C, Lazzarini TA, Nathanson JL *et al.* Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* 2013; **503**(7477): 525-529.
206. Matthaei KI, Andrews PW & Bronson DL. Retinoic acid fails to induce differentiation in human teratocarcinoma cell lines that express high levels of a cellular receptor protein. *Exp Cell Res* 1983; **143**(2): 471-474.
207. Buchholz F, Angrand PO & Stewart AF. Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat Biotechnol* 1998; **16**(7): 657-662.

208. Wei W, Morrish TA, Alisch RS & Moran JV. A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. *Anal Biochem* 2000; **284**(2): 435-438.
209. Pulgarin AJ, Cano D, Morell S, Amador S & L.Garcia-Perez J. Characterization of engineered L1 retrotransposition events: the recovery method. *Springer* 2015.
210. Benavides FJ, Guenet, J.L. . Manual de genética de roedores de laboratorio. Principios básicos y aplicaciones. Universidad de Alcalá *Universidad de Alcalá* (2003): 227-262.
211. Fiering S, Epner E, Robinson K, Zhuang Y, Telling A, Hu M *et al.* Targeted deletion of 5'HS2 of the murine beta-globin LCR reveals that it is not essential for proper regulation of the beta-globin locus. *Genes Dev* 1995; **9**(18): 2203-2213.
212. Dymecki SM. Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice. *Proc Natl Acad Sci U S A* 1996; **93**(12): 6191-6196.
213. Freshney RI. *Culture of Animal Cell:A Manual of Basic Technique*, 2005.
214. Gilbert N, Lutz-Prigge S & Moran JV. Genomic deletions created upon LINE-1 retrotransposition. *Cell* 2002; **110**(3): 315-325.
215. Gilbert N, Lutz S, Morrish TA & Moran JV. Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Mol Cell Biol* 2005; **25**(17): 7780-7795.
216. Meyer TJ, Srikanta D, Conlin EM & Batzer MA. Heads or tails: L1 insertion-associated 5' homopolymeric sequences. *Mob DNA* 2010; **1**(1): 7.
217. Le TN, Miyazaki Y, Takuno S & Saze H. Epigenetic regulation of intragenic transposable elements impacts gene transcription in *Arabidopsis thaliana*. *Nucleic Acids Res* 2015; **43**(8): 3911-3921.
218. Yajima I, Sato S, Kimura T, Yasumoto K, Shibahara S, Goding CR *et al.* An L1 element intronic insertion in the black-eyed white (Mitf[mi-bw]) gene: the loss of a single Mitf isoform responsible for the pigimentary defect and inner ear deafness. *Hum Mol Genet* 1999; **8**(8): 1431-1441.
219. Smith BF, Yue Y, Woods PR, Kornegay JN, Shin JH, Williams RR *et al.* An intronic LINE-1 element insertion in the dystrophin gene aborts dystrophin expression and results in Duchenne-like muscular dystrophy in the corgi breed. *Lab Invest* 2011; **91**(2): 216-231.

220. Schwahn U, Lenzner S, Dong J, Feil S, Hinzmann B, van Duijnhoven G *et al.* Positional cloning of the gene for X-linked retinitis pigmentosa 2. *Nat Genet* 1998; **19**(4): 327-332.
221. Woods-Samuels P, Wong C, Mathias SL, Scott AF, Kazazian HH, Jr. & Antonarakis SE. Characterization of a nondeleterious L1 insertion in an intron of the human factor VIII gene and further evidence of open reading frames in functional L1 elements. *Genomics* 1989; **4**(3): 290-296.
222. Hollister JD & Gaut BS. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* 2009; **19**(8): 1419-1428.
223. Rakyán VK, Chong S, Champ ME, Cuthbert PC, Morgan HD, Luu KV *et al.* Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proc Natl Acad Sci U S A* 2003; **100**(5): 2538-2543.

