



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**Doctorado en Ciencias Bioquímicas**

REGULACIÓN TRANSCRIPCIONAL DE UN ARN LARGO NO-CODIFICANTE  
INTERGÉNICO POR UNA SECUENCIA REPETIDA TIPO ALU

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

ROSARIO PÉREZ MOLINA

TUTOR PRINCIPAL

DR. FÉLIX RECILLAS TARGA

INSTITUTO DE FISIOLÓGÍA CELULAR, UNAM

MIEMBROS DEL COMITÉ

DR. JESÚS SANTA-OLALLA

Facultad de Medicina, UAEM

DR. JAIME MAS OLIVA

INSTITUTO DE FISIOLÓGÍA CELULAR, UNAM

CIUDAD UNIVERSITARIA, CD.MX.

MARZO 2023



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

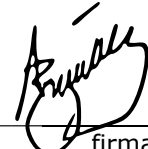

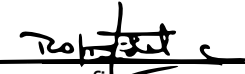
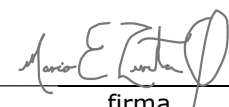
El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición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

**SINODALES DESIGNADOS**  
**Presente**

Estimado académico:

Los miembros del Subcomité Académico en reunión extraordinaria de **27 de octubre** del presente año, conocieron la solicitud de asignación de **JURADO DE EXAMEN** para optar por el grado de **Doctora en Ciencias** de la estudiante **Pérez Molina Rosario**, con la tesis **“Regulación transcripcional de un ARN largo no-codificante intergénico por una secuencia repetida tipo Alu”**, dirigida por la Dra. **Recillas Targa Félix**.

De su análisis se acordó nombrar el siguiente jurado en el que se encuentra usted incluido:

		Acepto			
PRESIDENTE	Zarain Herzberg Ángel Alfonso (FM)	Si <input checked="" type="checkbox"/>	No <input type="checkbox"/>	<u>11 / 11 / 22</u> fecha	<u></u> firma
VOCAL	Romero Camarena David René (CCG)	Si <input checked="" type="checkbox"/>	No <input type="checkbox"/>	<u>11 / 11 / 22</u> fecha	<u></u> firma
VOCAL	Rojas del Castillo Emilio (IIB)	Si <input checked="" type="checkbox"/>	No <input type="checkbox"/>	<u>14 / 11 / 22</u> fecha	<u></u> firma
VOCAL	Zurita Ortega Mario (IBT)	Si <input checked="" type="checkbox"/>	No <input type="checkbox"/>	<u>14 / 11 / 22</u> fecha	<u></u> firma
SECRETARIO	Escalante Alcalde Diana María (IFC)	Si <input checked="" type="checkbox"/>	No <input type="checkbox"/>	<u>11 / 11 / 22</u> fecha	<u></u> firma

Sin otro particular por el momento, aprovecho la ocasión para enviarle un cordial saludo.

**Atentamente**  
**“POR MI RAZA, HABLARÁ EL ESPÍRITU”**  
 Cd. Universitaria, Cd. Mx., a 28 de octubre de 2022  
**COORDINADORA**



**Dra. Claudia Lydia Treviño Santa Cruz**

## RECONOCIMIENTOS ACADÉMICOS

Esta tesis de Doctorado se realizó bajo la tutoría del Dr. Félix Recillas Targa del Departamento de Genética Molecular del Instituto de Fisiología Celular en la Universidad Nacional Autónoma de México.

El comité Tutorial que dirigió el desarrollo de esta tesis estuvo formado por:

Dr. Félix Recillas Targa                      Instituto de Fisiología Celular, UNAM

Dr. Jaime Mas Oliva                            Instituto de Fisiología Celular, UNAM

Dr. Jesús Santa-Olalla Tapiat            Facultad de Medicina, UAEM

Se reconoce la asesoría técnica de la Biol. Georgina Guerrero Avendaño durante la realización de diversos experimentos para el estudio de la función de la secuencia repetida AluSx, así como su labor en la implementación del protocolo de transducción de células por lentivirus en el laboratorio, que permitió llevar a cabo la eliminación *in vivo* de la secuencia repetida por la técnica de CRISPR-Cas9.

Se reconoce el apoyo de la laboratorista Gianelli Cortés por todo el apoyo para la preparación de material y algunas soluciones necesarias para la realización de muchos de los experimentos realizados durante este proyecto.

Se agradece el apoyo de la Dra. Laura Ongay Larios, la Bióloga Guadalupe Códiz Huerta y la Q.F.B. Minerva Mora Cabrera de la Unidad de Biología Molecular, en la secuenciación de ADN y síntesis de oligonucleótidos necesario para la realización de varios de los experimentos.

También se reconoce el apoyo del Ing. Juan Manuel Barbosa Castillo de la Unidad de Cómputo quién facilitó durante mi programa de Doctorado, el software y asesoría técnica para la instalación y uso de estos recursos utilizados en el análisis de los resultados.

Finalmente agradezco a la Dra. Sandra Moncada Hernández, coordinadora de la Biblioteca “Armando Gómez Puyou” del IFC por la asesoría en el uso del programa Mendeley para la anotación de manera eficiente y metódica las fuentes bibliográficas revisadas para la escritura de este manuscrito.

El proyecto fue financiado particularmente por la Dirección de Asuntos del Personal Académico (DGAPA)- PAPIIT de la UNAM (donativos IN203811, IN201114 y IN203620), así como el Consejo Nacional de Ciencia y Tecnología (CONACyT: 128464 y 220503; y la beca doctoral 173754). Con el apoyo adicional del Posgrado



en Ciencias Bioquímicas y el Instituto de Fisiología Celular de la Universidad Nacional Autónoma de México.

Agradezco también al “Programa de Apoyo a los Estudios del Posgrado” (PAEP) ya que durante mi doctorado asistí al congreso internacional “Chromatin: Structure & Function 2011” para la exposición del cartel: “Characterization of the different mechanisms involved in the epigenetic silencing of the human Retinoblastoma gene promoter” 2011 Aruba, Países Bajos.

# ÍNDICE

## RESUMEN

## ABREVIATURAS

### I. INTRODUCCIÓN

Las regiones no-codificantes del genoma y su papel en la regulación transcripcional

#### A. Elementos Regulatorios en *cis*

1. Promotores
2. *Enhancers* y *Superenhancers*
3. Silenciadores
4. *Insulators*

#### B. Los ARNs no-codificantes

1. Identificación en el genoma y su función regulatoria
2. Tipos de ARNs no-codificantes (miRNAs, siRNAs, lncRNAs y lincRNAs)

#### C. Los elementos transponibles (ETs) y secuencias repetidas

1. El mundo inexplorado de las secuencias repetidas
2. Clasificación y sus características
3. Su función en el tiempo y espacio

### II. ANTECEDENTES

Estructura del promotor del gen de *Retinoblastoma* humano (*RB1*) y su actividad bidireccional

### III. PLANTEAMIENTO DEL PROBLEMA

### IV. HIPÓTESIS

### V. OBJETIVO

### VI. MATERIALES Y MÉTODOS

### VII. ESTRATEGIA EXPERIMENTAL

1. Determinar la actividad de la secuencia repetida AluSx río arriba del promotor del gen *RB1*.
2. Caracterización de la función de la repetida AluSx en su contexto endógeno

3. Análisis genómico del efecto de secuencias repetidas tipo Alu en lincRNAs

VIII. RESULTADOS

1. Análisis de la distribución de las secuencias repetidas tipo Alu en el locus del gen *RB1*
2. La secuencia repetida Alu se comporta como un elemento *enhancer* y protege contra el silenciamiento epigenético en construcciones reporteras
3. La eliminación de la secuencia de la repetida Alu promueve la regulación positiva del *LINC00441* ocasionando alteraciones en la proliferación celular
4. Las secuencias de Alu intragénicas correlacionan con niveles más bajos de transcripción de lincRNAs en diferentes tejidos humanos

IX. DISCUSIÓN

X. CONCLUSIONES

XI. PERSPECTIVAS

XII. BIBLIOGRAFÍA

XIII. ANEXOS

## RESUMEN

Los elementos transponibles (ETs) son secuencias que están presentes en todos los organismos y su presencia ha tenido un impacto en la evolución de los genomas y la biodiversidad. Estas secuencias han sido propuestas como los ancestros de varios de los elementos reguladores que conforman la gran red de regulación transcripcional en el genoma humano. Los elementos Alu son secuencias repetidas específicas de primates y representan el tipo más abundante de ETs en el humano. Su paisaje cromatínico dado por el enriquecimiento de modificaciones postraduccionales en las histonas, así como la accesibilidad al ADN, sugiere que las secuencias Alu humanas podrían funcionar como *enhancers* potenciando su transcripción; sin embargo, ningún experimento funcional ha evaluado el papel de las secuencias Alu en el control de la transcripción. El presente trabajo se enfoca en analizar la actividad reguladora de una secuencia Alu humana de la familia AluSx ubicada en el segundo intrón del ARN no-codificante intergénico largo *LINC00441*, que se encuentra en orientación divergente al gen *Retinoblastoma (RB1)*. Observamos que la secuencia Alu actúa como un elemento tipo *enhancer* basados en ensayos utilizando un gen reportero. Eliminaciones realizadas mediante el sistema CRISPR-Cas9 de la secuencia Alu endógena en células eritroleucémicas K562, dieron como resultado una regulación transcripcional positiva del *LINC00441* y una disminución en la proliferación celular. Finalmente, mediante análisis de genoma completo, se determinó que los lincRNAs con Alus intrónicos muestran niveles generales de transcripción más bajos que aquellos que no las contienen.

Los resultados sugieren que una secuencia Alu intragénica con actividad de *enhancer* puede actuar como un atenuador transcripcional del lincRNA huésped y esta nueva función regulatoria sea una evidencia que sustente el proceso de domesticación que plantea que las secuencias repetidas experimentan para preservarse evolutivamente en el genoma.

## ABSTRACT

Transposable elements (TEs) are present in all organisms and their presence has had a major impact on genome evolution and biodiversity. TEs have been proposed as the ancestors of several of the regulatory elements that make up the large transcriptional regulatory network in the human genome. Alu elements are primate-specific repeats and represent the most abundant type of TEs in the human genome. Genome-wide analysis of the enrichment of histone post-translational modifications, as well as the accessibility to DNA, suggests that human Alu sequences could function as transcriptional enhancers; however, no functional experiments have evaluated the role of Alu sequences in the control of transcription *in situ*. The present study focuses on analyzing the regulatory activity of a human Alu sequence from the AluSx family located in the second intron of the long intergenic non-coding RNA *LINC00441*, found in divergent orientation to the *Retinoblastoma* (RB1) gene. We observed that the Alu sequence acts as an enhancer element based on reporter gene assays. CRISPR-Cas9 deletions of the endogenous Alu sequence in K562 erythroleukemic cells, resulted in a marked transcriptional upregulation of *LINC00441* and a decrease in cell proliferation. Finally, through genome-wide analysis, it was determined that lincRNAs with intronic Alus show lower overall transcription levels than those that lack it. The results suggest that an intragenic Alu sequence with enhancer activity can act as a transcriptional attenuator of its host lincRNA and this new regulatory function is evidence that supports the domestication process that repeat sequences have to go through to preserve themselves evolutionarily in the genome.

## ABREVIATURAS

ET: Elemento Transponible

SNPs: Polimorfismos de un solo nucleótido

FT: Factor de Transcripción

TFBS: (*Transcription factor binding site*) Sitio de unión a factor de transcripción

FAIRE: (*Formaldehyde assisted isolation of Regulatory Elements*) Aislamiento de elementos regulatorios asistido con formaldehído

ATAC: (*Assay for Transposase-Accessible Chromatin*) Ensayo de accesibilidad a la cromatina por Transposasa

CTCF: *CCCTC binding factor*

pb : pares de bases

TSS: (*Transcription start site*) Sitio de inicio de la Transcripción

mRNA: (*messenger RNA*) ARN mensajero

eRNA: (*enhancer RNA*) ARN *enhancer*

lncRNA: (*long non coding RNA*) ARN no codificante largo

CpGs: Citosina fostato Guanina

ESC: (*embryonic stem cell*) célula troncal embrionaria

ChIP: (*chromatin immunoprecipitation*) inmunoprecipitación de la cromatina

qRT-PCR: (*quantitative Real-time Polymerase Chain Reaction*) Reacción en cadena de la polimerasa cuantitativa en tiempo real

RNApolIII: RNA polimerasa II

siRNA: (*small interfering RNA*) ARN pequeño de interferencia

LTR: (*long terminal repeat*) repetición terminal larga

SINEs: (*Short Interspersed nuclear elements*) Elementos nucleares dispersos cortos

LINEs: (*Long Interspered nuclear elements*) Elementos nucleares dispersos largos

FACS: (*Fluorescence-activated cell sorting*) Clasificación de células activadas por fluorescencia

PCR: (*Polymerase chain reaction*) Reacción en cadena de la Polimerasa

GFP: (*Green fluorescent protein*) Proteína verde fluorescente

RB1: Gen de Retinoblastoma 1

DSB: (*Double strand break*) Corte de doble cadena

sgRNA: (*single-guide RNA*) ARN de guía única

## I. INTRODUCCIÓN

Durante las últimas décadas posteriores a la secuenciación del genoma humano (Consortio Internacional de Secuenciación del Genoma Humano 2004) y de la secuenciación de varias de las especies eucariontes que utilizamos como organismos modelo en la biología, ha surgido el interés por entender en su totalidad el significado de todo el material genético contenido en nuestros genomas. Uno de los hallazgos más enigmáticos fue el descubrimiento de que gran parte de los genomas en organismos complejos, a pesar de ser transcritos, no codifican para proteínas. La caracterización funcional de este tipo de transcritos, ha demostrado que son importantes en la fisiología celular; particularmente al nivel de la regulación de la expresión génica. Por lo anterior, el presente trabajo tiene como objetivo contribuir en el entendimiento de la función de este tipo de secuencias. Para ello, considero necesario introducirnos en el tema de los diferentes tipos de secuencias no-codificantes en el genoma y lo que se conoce acerca de su función.

### **Las regiones no-codificantes del genoma y su papel en la regulación transcripcional**

Todas las células de un organismo poseen la misma información genética y es a través de una regulación coordinada de la expresión génica como un genoma confiere identidad y función específica a todas las células de cada especie. Por lo tanto, la regulación génica es crítica para la organización y coordinación de las funciones celulares y el desarrollo de un organismo (Vaquerizas *et al.*, 2009). Además de las secuencias codificantes, el genoma está compuesto mayoritariamente por secuencias no-codificantes con características muy variadas, pero sobre todo con una abundancia en elementos transponibles (ETs) y secuencias repetidas. Muestra de la relevancia de las secuencias no-codificantes, es la asociación de enfermedades complejas a variantes genéticas (polimorfismos de un solo nucleótido, SNPs) y mutaciones localizadas en regiones no-codificantes del



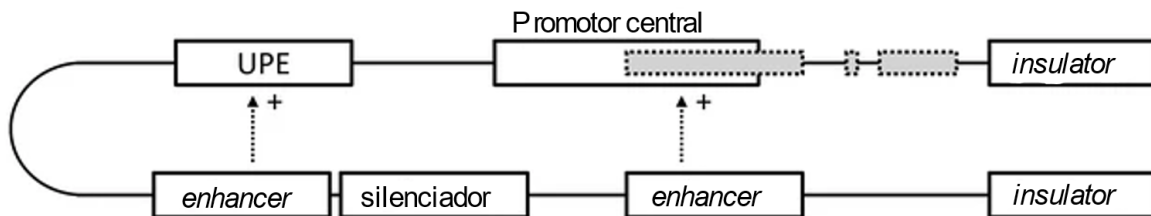
genoma humano, demostrando la importancia del ADN regulatorio en regiones intergénicas y en el desarrollo de un gran número de patologías (Welter *et al.*, 2014; Huang, 2015). Como resultado de las evidencias funcionales y correlativas, actualmente el genoma no-codificante funcional se divide en elementos regulatorios en *cis*, en ARNs no-codificantes y en elementos transponibles o secuencias repetidas.

Durante las últimas décadas, a partir de la secuenciación de genomas de diferentes organismos así como la disponibilidad de perfiles cromatínicos (basados en la accesibilidad del ADN, marcas postraduccionales en las histonas y modificaciones epigenéticas), se han definido como secuencias funcionales y que probablemente estén relacionadas con regulación, a una gran cantidad de secuencias no-codificantes en regiones intergénicas en el genoma humano ( $>8 \times 10^6$ ). Este número es mucho mayor al compararlas con el número de secuencias codificantes (20,000) (Lander *et al.*, 2001) que corresponde únicamente al 1-2% del total del genoma humano.

Por lo anterior, grandes consorcios como ENCODE y el NIH, a través del uso de tecnologías de secuenciación masiva han generado e integrado mapas epigenómicos (marcas de histonas y sitios de unión de factores de transcripción) y mapas de accesibilidad al ADN en más de 150 líneas celulares humanas y de diversos tejidos (Dunham *et al.*, 2012; Roadmap Epigenomics Consortium *et al.*, 2015). Actualmente el acceso a esta información nos puede guiar en la identificación y estudio del funcionamiento de estas regiones potencialmente regulatorias en el genoma. Cabe resaltar que existen limitaciones para determinar la función de estas secuencias, ya que a pesar de que una región contenga ciertas características en su cromatina que correlacionan con una función regulatoria, requiere de una validación funcional directa que establezca su papel en la regulación genética. Por lo anterior es que estrategias experimentales como las que se abordaron en este trabajo, son necesarias para determinar la función regulatoria de diversos tipos de secuencias, incluyendo secuencias repetidas, sobre su gen blanco.

## A. Elementos Regulatorios en *cis*

Una región regulatoria se define por la actividad que deriva de la asociación de proteínas de unión al ADN, la cual requiere que en el contexto de la cromatina, esta región se encuentre accesible para su reconocimiento y unión. Es por ello que metodologías que determinan la accesibilidad al ADN como la hipersensibilidad a la DNasa I (DNase-seq), el aislamiento de elementos reguladores asistido por formaldehído (FAIRE) o las mediciones de accesibilidad a la cromatina con el uso de una transposasa (ATAC) (Giresi *et al.*, 2007; Boyle *et al.*, 2008; Bulger & Groudine, 2011), han sido útiles para la identificación de elementos proximales como son los promotores; y distales, como los *enhancers*, *insulators* y silenciadores (Elkon & Agami, 2017). En la Figura 1, se representa la unidad regulatoria transcripcional de un gen en eucariontes. Un promotor central o *core* que traslapa al primer exón y elementos promotores proximales río arriba (UPE). El bucle de ADN muestra cómo las secuencias que no se encuentran en cercanía de manera lineal con su promotor, como los *enhancers* y/o silenciadores se puedan acercar físicamente al promotor central o al UPE y regular su actividad.

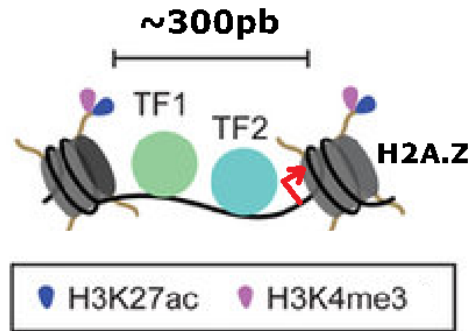


**Figura 1. Representación de una unidad regulatoria transcripcional.** Los exones se representan por cuadros grises con borde de líneas discontinuas (figura modificada de Riethoven, 2010).

### 1. Promotores

Se definen como la región de ADN que controla y da inicio a la transcripción de una unidad genómica codificante o no-codificante. Los promotores activos generalmente contienen de 80-300 pares de bases (pb) libres de nucleosomas delimitados por un par de nucleosomas bien posicionados en cada uno de sus extremos. Estos

nucleosomas se caracterizan por contener a la variante de histona H2A.Z y un enriquecimiento de la trimetilación en la lisina 4 de la histona H3 (H3K4me3) (Barski *et al.*, 2007); (Figura 2).



**Figura 2. Región cromatínica definida como promotor.** Región cromatínica accesible a la unión de factores de transcripción (FTs) que se caracteriza por la incorporación de la variante de histonas H2A.Z en sus nucleosomas extremos y el enriquecimiento de la marca H3K4me3.

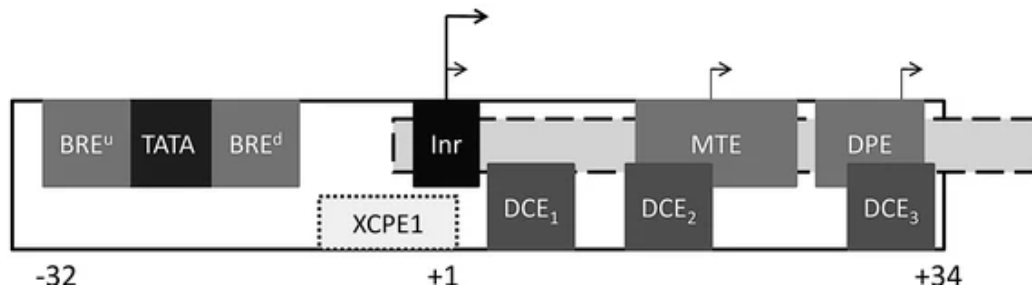
Estos elementos están compuestos por una región central alrededor del sitio del inicio de la transcripción (TSS +1) que contiene las secuencias indispensables para la unión de las distintas subunidades que componen la RNA polimerasa II (RNAPolII) y cofactores necesarios para el inicio de la transcripción. En estas regiones existe en muchos casos una sobrerrepresentación de dinucleótidos CpGs definidas como islas CpG.

En eucariontes superiores existe una mayor diversidad de secuencias descritas como sitios de unión que definen a los promotores. Entre las secuencias más frecuentemente halladas y bien caracterizadas está el elemento iniciador (Inr) que es quien define al *core* del promotor, el cual se extiende a ambos lados del TSS y está presente entre el 40-60% de los promotores por encima de la caja TATA que solo se encuentra entre el 10-20% (Gershenzon *et al.*, 2006; Yang *et al.*, 2007). A diferencia del Inr, entre las cajas más conservadas encontramos a la caja TATA, la cual une a la TBP (por sus siglas en inglés TATA-Binding Protein), que junto con la unión de factores asociados a la caja TATA (TAFs), forman el complejo general de la transcripción TFIID; primer paso para la creación de un complejo de inicio de la

transcripción (PIC) compuesto por más de 100 proteínas como son los TFIIA-J, entre otros (Smale & Kadonaga, 2003).

Sin embargo, desde hace ya varios años se sabe que existen otras secuencias importantes presentes en promotores con una menor frecuencia pero que de igual manera contribuyen para la formación del PIC (Cooper *et al.*, 2006). Entre ellas está, el elemento de reconocimiento del factor TFIIIB (BRE, ascendente y descendente), y el elemento promotor río abajo (DPE) (Yang *et al.*, 2007). Otras secuencias asociadas al *core* de promotores son el motivo de diez elementos (MTE), el elemento central río abajo (DCE, subunidades 1, 2 y 3) y el elemento X del promotor central 1 (XCPE1) (Riethoven, 2010).

En la Figura 3 se representa la distribución de todas estas secuencias en un promotor, ejemplificando la existencia de promotores con un solo TSS (negrita) y otros con varios TSS alternativos (más pequeños) que también son una variante más para la transcripción de un gen. En resumen, estas secuencias cortas pero altamente específicas, le dan al *core* de los promotores, una característica única y representan un primer nivel de regulación para la acción de la maquinaria transcripcional; en particular para el posicionamiento regulado del complejo de inicio de la transcripción.



**Figura 3. Arquitectura detallada del promotor central.** Varios elementos, muchos de ellos activos a partir de distintas combinatorias, se muestran aproximadamente a escala. Los elementos promotores sombreados más oscuros son los más frecuentemente hallados en promotores (figura modificada de Riethoven, 2010).

## 2. *Enhancers* y *Superenhancers*

Los potenciadores o mejor conocidos por su anglicismo, *enhancers*, son regiones que contienen sitios de unión a factores de transcripción y cofactores que aumentan los niveles transcripcionales de sus genes blanco. Su papel es fundamental para una regulación diferencial tejido específica.

Su identificación ha sido difícil debido a la baja conservación en sus secuencias, ya que en ocasiones el mismo patrón de expresión de un gen en distintas especies, se obtiene por un control instrumentado por secuencias con poca similitud entre sí (Hare *et al.*, 2008), además que su localización no necesariamente tiene que ser cercana a los genes sobre los que actúan, ni siquiera en el mismo cromosoma. Por lo mismo, son secuencias que presentan una gran capacidad de acción ya que pueden actuar en *cis* o *trans* (Tolhuis *et al.*, 2002; Lomvardas *et al.*, 2006). Son elementos que actúan independientemente de su orientación y distancia (de Kb a Mb) en relación a su gen blanco. Tienen una localización genómica muy variada ya que se pueden hallar comúnmente en regiones intergénicas e incluso en intrones de regiones codificantes y no-codificantes. Por todo lo anterior, sus características a nivel epigenético es una de las formas más eficientes para identificarlas en el genoma.

Los *enhancers* son secuencias que ocupan alrededor de 200-500 pb con una cromatina abierta, lo cual permite que sean regiones sensibles al corte por DNasa I (He *et al.*, 2010). Están delimitadas por regiones ricas en la variante de histonas H3.3 y H2AZ, en modificaciones como la mono y/o dimetilación de la lisina 4 de la histona H3 (H3K4me1/H3K4me2) y la acetilación de la lisina 27 en la histona H3 (H3K27ac). Además de estas marcas de histonas clásicas, se han descrito un par de marcas más (H3K27me3 o H3K36me3) que, dependiendo de su enriquecimiento en un *enhancer*, se clasifican en activos, intermedios o en pausa; en relación a su estado y función durante el desarrollo y diferenciación así como de su contexto celular (Zentner *et al.*, 2011).

Generalmente son regiones a las cuales se asocian proteínas como la acetiltransferasa de histonas p300, subunidades del complejo Mediador, la helicasa

CHD7 dependiente de ATP, cohesina y/o CTCF (Heintzman *et al.*, 2009; Zentner *et al.*, 2011) y para *enhancers* canónicos se observa además la unión de la RNAPolIII (De Santa *et al.*, 2010). Y ya de manera particular, factores de transcripción tejido específicos que varían dependiendo del gen blanco (Soufi *et al.*, 2012, 2015; Spitz & Furlong, 2012). Por ejemplo, uno de los contextos celulares más estudiados es el de las células troncales embrionarias (ESCs) donde se ha visto que la unión de los llamados factores pioneros o de enucleación (FOXA, PU.1, OCT4 o GATA-1), son aquellos que se unen a *enhancers* que se encuentran parcialmente accesibles (Zaret, 2020). Éstos a su vez, alteran el estado de la cromatina a través de la unión de cofactores que modifican (e.g. acetiltransferasas de histonas) o reclutan complejos remodeladores de cromatina ATP-dependientes que favorecen la unión de factores tejido específicos necesarios para la activación de la transcripción (You *et al.*, 2011).

Y es así como la implementación de nuevas tecnologías de secuenciación y el hallazgo de varias de estas características en determinadas regiones, ha sido posible identificar sorprendentemente a más de 400,000 *enhancers* en el genoma humano (Dunham *et al.*, 2012).

Adicional a la caracterización de estos elementos, la forma en que los *enhancers* estimulan la transcripción es una pregunta central que sigue sin entenderse por completo. Por lo que, se han propuesto dos modelos para explicar su funcionamiento: el modelo binario y el modelo progresivo o reostático (Ko *et al.*, 1990; Walters *et al.*, 1995; García-González *et al.*, 2016). El modelo binario propone que los *enhancers* aumentan la probabilidad que un mayor número de células activan la transcripción en un locus específico dentro de una población celular. En contraste, el modelo progresivo propone que los *enhancers* aumentan el número de moléculas de ARN transcritas del gen blanco, pero no tienen un efecto en el número de células que inician la transcripción (García-González *et al.*, 2016).

Además de estas dos propuestas en cuanto a la forma en la que operan, actualmente muchos de los estudios para entender el mecanismo por el cual estos elementos de regulación funcionan; considera la forma en la que el ADN se

encuentra distribuido y organizado dentro del núcleo de los organismos eucariontes. La cromatina, estructura compuesta por ADN e histonas, a través de la formación de asas (“loops”) desempeña un papel al acercar físicamente los *enhancers* a la región promotora proximal o central de un gen blanco (Miele & Dekker, 2008). Estas interacciones han sido descritas a través de ensayos de captura conformacional de la cromatina (3C) y protocolos derivados de éste (4C, 5C, Hi-C; entre otros) (Simonis *et al.*, 2007; Lieberman-Aiden *et al.*, 2009). Aún no es clara la forma en la que se regula la formación de estas estructuras y todas las proteínas involucradas, pero derivado de algunas evidencias se han propuesto diferentes modelos de acción para este tipo de reguladores de la transcripción. Por ejemplo, el modelo de barrido plantea la hipótesis de que el complejo protéico de *enhancer* activo recorre de alguna manera la hebra de cromatina hasta que encuentra la región promotora (Kolesky *et al.*, 2002). En el modelo de relocalización donde se basa en el descubrimiento de que los *enhancers* son fundamentales para cambiar la posición subnuclear de los genes diana y acercarlos a una fuente lista de RNAPolIII: los loci o fábricas transcripcionales de RNAPolIII, aumentando así su transcripción (Lancôt *et al.*, 2007; Cook, 2010).

Otra de las características que se ha ligado con el estado de activación de un *enhancer* es que desde hace una década se ha reportado que los *enhancers* se transcriben en ARNs no-codificantes (eRNAs) tejido-específicos (De Santa *et al.*, 2010; Kim *et al.*, 2010; de Lara *et al.*, 2019). En general se ha descrito que esta transcripción puede ocurrir bidireccionalmente dando a lugar a moléculas de ARN no-poliadeniladas ni procesadas (Andersson *et al.*, 2014), por lo que son rápidamente degradadas por el exosoma. La mayoría de los eRNAs se transcriben en menor abundancia que los ARN mensajeros (mRNAs) y otros ARNs no-codificantes como los ARNs largos no-codificantes (lncRNAs) (Pefanis *et al.*, 2015). Es por ello que su inestabilidad y poca abundancia ha sido un factor limitante para determinar los mecanismos a través de los cuales los eRNAs están implicados como componentes críticos de los *enhancers* activos. Se ha demostrado que los eRNAs pueden reclutar proteínas estructurales como la cohesina (RAD21 y SMC3) para la formación de asas de cromatina y se propone que estas moléculas actúan como

andamiaje para la formación de estructuras que estabilizan la interacción *enhancer*-promotor (Kim *et al.*, 2010; D. Wang *et al.*, 2011; Li *et al.*, 2013). Además, se ha visto la unión de eRNAs a varios factores de transcripción donde se sugiere pueden estar promoviendo un mayor tiempo de residencia de los FTs en los *enhancers* y esto a su vez la asociación de RNAPolIII a estas regiones. Otro de los hallazgos relacionados con su función, se plantea al identificar que existen eRNAs que interactúan con la proteína que une a CREB (CBP), una acetiltransferasa que, junto con la p300; se encarga de la acetilación en la histona H3 en la lisina 27 (K27) entre otros residuos de histonas. En este mismo estudio, se demostró que la unión de los eRNAs está relacionada con la actividad de acetiltransferasa promoviendo así, el establecimiento de una cromatina permisiva apta para la función de los *enhancers* (Bose *et al.*, 2017). También se ha visto que los eRNA pueden facilitar la liberación de la RNAPolIII pausada, a través de la activación del complejo de elongación p-TEFb (Zhao *et al.*, 2016). Y por último, adicional a estas evidencias, los eRNA también pueden actuar como señuelos para secuestrar cofactores que reprimen la transcripción en los genes blanco (Schaukowitch *et al.*, 2014).

Por otra parte, varios de los estudios que se enfocan en el entendimiento de los elementos *enhancers*, han descrito al complejo Mediador como un factor esencial para el reclutamiento de la RNAPolIII a través de FTs para el mantenimiento de las células madre y su diferenciación. Y fue que por primera vez en ESCs donde se identificaron regiones de hasta 50 kb que muestran un gran enriquecimiento de los componentes del complejo Mediador así como la histona H3K27ac, las cuales denominaron *superenhancers* (Whyte *et al.*, 2013). Estas regiones se encuentran enriquecidas de motivos de unión de factores de transcripción “maestros” específicos para cada tipo celular por lo que se asocian a genes que controlan y definen la biología de estas células (Hnisz *et al.*, 2013).

En resumen, el modo de acción de los *enhancers* es complejo y variado. Requieren de la acción coordinada de varios elementos incluyendo a los cambios en su estructura de la cromatina, accesibilidad a un conjunto de FTs, eRNAs y su estructuración óptima a nivel tridimensional mediante la formación de asas. Sin



embargo, pueden existir otros mecanismos, incluso aquellos en los cuales las secuencias repetidas pueden jugar un papel (ver más adelante).

### 3. Silenciadores

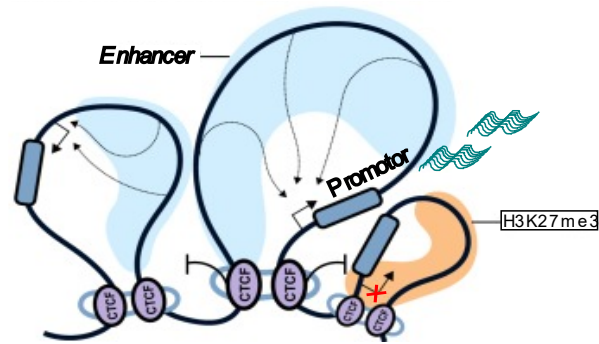
Son elementos reguladores distales que reducen la actividad de su promotor blanco. A pesar de que los silenciadores han sido menos estudiados, varios grupos de investigación utilizando técnicas como ensayos reporteros, ensayos de inmunoprecipitación de la cromatina (ChIP) y de hipersensibilidad a la DNasa I, han podido conocer algunas de las siguientes características acerca de estos elementos. Se sabe que son elementos modulares (es decir, que conservan su actividad represiva aun cuando se retiran de su contexto genómico nativo) y al igual que los *enhancers*, pueden actuar independiente de la posición y la orientación (Zheng *et al.*, 2004; Qi *et al.*, 2015). Reclutan represores transcripcionales como SNAI1 y 2, REST, KLF12 y MeCP2; además de participar activamente en el silenciamiento modificando el estado de la cromatina (Ogbourne & Antalis, 1998; Amir *et al.*, 1999; Westbrook *et al.*, 2008).

Una de las características que ha sorprendido a varios grupos, es el hecho de que existen algunos ejemplos de silenciadores tejido específicos que actúan como elementos reguladores en *cis* bifuncionales, es decir que también actúan como *enhancers* en diferentes contextos celulares incluso hay elementos que actúan sobre un gen como silenciador y en otro como *enhancer* en un mismo tipo celular (Gisselbrecht *et al.*, 2020; Bandara *et al.*, 2021). Por lo anterior es que su identificación, así como el definir características únicas de la cromatina en estos elementos ha sido una tarea difícil. Entre las marcas de histonas más relacionadas con estos elementos, están la H4K20me y la H3K27me3 (Pang & Snyder, 2020). Por lo anterior es que algunos silenciadores se han relacionado con los complejos represores Polycomb (PRC1 y PRC2) encargados de depositar y mantener a la trimetilación de la K27 en la histona H3 y fomentar una cromatina compacta (Ngan *et al.*, 2020).

Existen distintos tipos de silenciadores, por lo que aún no hay una firma de cromatina característica de todos, o incluso la mayoría, de los silenciadores. De hecho, parece probable que no haya una sola vía para la función del silenciador. En cambio, todos los fenómenos que conducen a la represión transcripcional (incluida la heterocromatina tanto facultativa como constitutiva, el silenciamiento de elementos móviles en el ADN mediado por siRNA y la partición del genoma en compartimentos activos e inactivos) podrían aportar mecanismos de silenciamiento adicionales a uno o más subconjuntos de silenciadores funcionalmente definidos (Segert *et al.*, 2021). El mecanismo de acción de estos elementos se ha estudiado a nivel de la interacción y efecto que tienen sobre sus genes blanco, pero es importante recordar que todos estos circuitos regulatorios funcionan en una organización tridimensional en el espacio nuclear y es ahí donde se plantea un modelo donde la cercanía o lejanía con “hot spots” de transcripción nucleoplásmica juega un papel determinante para su función (Kolovos *et al.*, 2012).

#### **4. Elementos frontera o *insulators***

La organización tri-dimensional del genoma ha requerido de estrategias sofisticadas para su formación. Como se puede observar en la Figura 4, los elementos frontera o *insulators* se localizan frecuentemente en la base de las asas cromatínicas, contribuyendo a su formación y estructura (Arzate-Mejía *et al.*, 2018). Estos elementos son secuencias que unen a las denominadas proteínas arquitectónicas entre las cuales se encuentran YY1 y principalmente CTCF en asociación con cohesinas (Sun *et al.*, 2022). Algunos *insulators* por lo tanto contribuyen a la organización topológica del genoma y participan, de manera indirecta en el control de la expresión diferencial de los genes. Sin embargo, los *insulators* tienen otras funciones entre las cuales se ha caracterizado la actividad de barrera. Esta actividad separa y/o contiene zonas de heterocromatina evitando así su propagación a zonas más relajadas de cromatina (eucromatina) donde se localizan usualmente los genes transcripcionalmente activos.



**Figura 4. Esquema de asas cromatínicas formadas por elementos frontera.** La unión de proteínas arquitectónicas como CTCF, permite poner en cercanía física a un *enhancer* con su promotor blanco, así como aislar al gen de señales regulatorias de regiones vecinas (figura modificada de Arzate-Mejía, 2018).

En el contexto de este trabajo de tesis, la función de barrera junto con la formación de asas resulta relevante dado que la presencia de elementos que influyen en la actividad transcripcional de genes vecinos como son las secuencias repetidas cercanas a unidades transcripcionales, requieren estar contenidas sus características represoras para evitar la propagación de heterocromatina (De La Rosa-Velázquez *et al.*, 2007; Arzate-Mejía *et al.*, 2018). Este es un aspecto de los *insulators* poco estudiado y que merece ser analizado en detalle dada la abundancia de secuencias repetidas en el genoma (Schmidt *et al.*, 2012; Modzelewski *et al.*, 2022). Otro aspecto relevante acerca de su función, se ha descrito en ensayos donde la disrupción genética de estos elementos, tiene por lo general, efectos adversos para la expresión de los genes contenidos en las asas de cromatina que se encuentran enmarcadas por los *insulators* (Lupiáñez *et al.*, 2015; Arzate-Mejía *et al.*, 2020). Por lo tanto, la función estructural y su relación con el estado de la cromatina, constituye otro nivel de regulación que influye en múltiples procesos, incluyendo la expresión regulada de los genes.

Los elementos regulatorios clásicos en *cis* descritos anteriormente son componentes fundamentales de una red de regulación transcripcional de un gen, pero dada su complejidad es evidente que aún faltan piezas por descubrir. Durante las últimas dos décadas, el interés por conocer la función de la fracción del genoma

no-codificante sobre todo aquellas regiones que forman parte del transcriptoma humano, ha dado como resultado el encontrar más elementos relacionados con la regulación transcripcional. Es así como una gran variedad de moléculas de ARN, así como de secuencias repetidas, se han sumado a la lista de elementos que de manera coordinada actúan para regular la expresión de un gen. Los mecanismos de acción aún no son claros y sobre todo es de particular interés, el conocer en qué contextos están presentes y la forma en que un gen interpreta toda esta información.

## **B. Los ARNs no-codificantes**

### **1. Identificación en el genoma y su función regulatoria**

La relevancia funcional del ARN ha tenido un crecimiento exponencial desde 1961 cuando Jacob y Monod establecieron el concepto del ARN mensajero (mRNA) como la molécula resultante de la transcripción de un gen, la cual da origen a la proteína codificante respectiva (Jacob & Monod, 1961). Y no fue hasta varios años después que diversos estudios reportaron la presencia de grandes cantidades de ARN que era transcrito pero no se traducía a proteínas (Weinberg & Penman, 1968; Paul & Duerksen, 1975; Salditt-Georgieff *et al.*, 1981). Algunas funciones de estos ARNs se explicaron más tarde mediante el proceso de edición génica (*splicing* o procesamiento) y la participación de ARN en la maquinaria de traducción y su regulación como el ARN ribosomal (rRNA) y los ARN de transferencia (tRNA).

Y fue a partir de la secuenciación del Genoma Humano y análisis transcriptómicos más finos acoplados a técnicas de secuenciación masiva (RNA-seq), que demostraron que la mayor parte del genoma eucarionte presenta una transcripción ubicua, concluyendo que existe una gran proporción de ARNs no-codificantes mucho mayor de lo antes pensado (Djebali *et al.*, 2012). Es de esperarse que aún se desconozca la funcionalidad de todas estas moléculas de ARN, incluso si todas ellas resulten funcionales (Guttman & Rinn, 2012). A pesar de ello, actualmente existen varios estudios que demuestran que los ARNs no-codificantes son de

relevancia a nivel celular (Penny *et al.*, 1996; Martianov *et al.*, 2007; Rinn *et al.*, 2007; Bartel, 2009; Wang *et al.*, 2011). Esta caracterización ha logrado identificar también ARNs estructurales y regulatorios a través de su localización celular y similitud de secuencias (Paul & Duerksen, 1975; Maison *et al.*, 2002; Bernstein *et al.*, 2006; Bartel, 2009). Y es así como existen muchos ejemplos de ARNs no-codificantes que participan en la regulación de diversos procesos biológicos importantes como el mantenimiento de la pluripotencia de células troncales (Loewer *et al.*, 2010), regulación del desarrollo a través de la diferenciación celular (Dinger *et al.*, 2008), regulación del ciclo celular (Huarte *et al.*, 2010) así como algunos ARNs involucrados en impronta e inactivación del cromosoma X (Xist, H19, AIR) (Bernstein & Allis, 2005; Nagano & Fraser, 2011).

## **2. Tipos de ARNs no-codificantes (miRNAs, siRNAs, lncRNAs y lincRNAs)**

En un primer intento por clasificar los nuevos transcritos no-codificantes provenientes de estudios globales del estado de transcripción celular, se hizo una distinción a partir de su tamaño ya que varían desde los ~22 nucleótidos (nt) hasta 200 nt, definiendo así a los ARNs pequeños; y aquellos con una longitud a partir de los ~200 nts hasta miles de nucleótidos denominados genéricamente como lncRNAs (por sus siglas en inglés: *long noncoding RNAs*).

Dentro de la clasificación de ARNs pequeños encontramos a los miRNAs (microRNAs), los siRNAs (ARNs pequeños de interferencia), los snoRNAs (ARNs pequeños nucleolares), los rasiRNAs (ARNs pequeños interferentes asociados a secuencias repetidas), los snRNAs (ARNs pequeños nucleares), los piRNAs (ARNs que interactúan con PIWI) los cuales se han asociado a la formación de heterocromatina; entre otros (Gavazzo *et al.*, 2013). Los snoRNAs es el grupo más abundante de los ARNs no-codificantes pequeños presentes en eucariontes, los cuales tienen una función central conservada en la biogénesis ribosomal, formando parte de los complejos ribonucleoproteínicos, snoRNPs (Dieci *et al.*, 2009).

Por otra parte, los lncRNAs son los transcritos más largos que se han descrito hasta el momento de hasta 100 Kb (Seidl *et al.*, 2006; Pandey *et al.*, 2008), los cuales se ha visto que ejercen diversas funciones reguladoras dentro de la célula (Ponting *et al.*, 2009; Mercer & Mattick, 2013). Resulta interesante el hecho de que estas moléculas comparten muchas características bioquímicas y estructurales con los mRNAs codificantes. Por ejemplo son transcritos por la RNAPolIII, la mayoría presenta 5'-cap y se poliadenilan en su extremo 3', además de tener una estructura exón-intrón por lo que también son procesados post-transcripcionalmente (Guttman *et al.*, 2009; Cabili *et al.*, 2011; Derrien *et al.*, 2012). En contraste, estos transcritos presentan patrones de expresión tejido-específico pero en muy bajos niveles comparados con transcritos codificantes, son inestables y están poco conservados entre las distintas especies (Pang *et al.*, 2006; Ravasi *et al.*, 2006; Dinger *et al.*, 2009; Ponting *et al.*, 2009; Cabili *et al.*, 2011).

Existen varias propuestas de la forma en que los lncRNAs están implicados en la regulación génica en diferentes procesos celulares que van desde la formación de “andamios moleculares” para la formación de complejos multiprotéicos y ribonucleicos (Maison *et al.*, 2002; Wutz *et al.*, 2002; Rinn *et al.*, 2007; Guttman *et al.*, 2011; Rinn & Chang, 2012) hasta fungir como elementos “guía” para la unión de moléculas de ADN o de proteínas encargadas de promover y/o reprimir la transcripción ya sea a través de su reclutamiento o modulando su actividad regulatoria (Willingham *et al.*, 2005; Yao *et al.*, 2010). Además de los ejemplos anteriores donde se ha visto que estos transcritos son los que propiamente ejercen esta función regulatoria, se han sumado evidencias que demuestran que la transcripción de estas moléculas influye en el estado de la cromatina (deposición de nucleosomas, marcas de cromatina, entre otros) teniendo un efecto en la regulación transcripcional del o los genes cercanos a su vecindad (Núñez-Martínez & Recillas-Targa, 2022).

Por lo anterior es que resulta interesante que derivado de la caracterización y anotación global de los lncRNAs, se ha propuesto a su vez, una clasificación de los mismos basada en su localización relativa con respecto al gen codificante más

cercano (Derrien *et al.*, 2012). Ya que puede ser génico si traslapa con otros genes o intergénico con respecto a los genes codificantes previamente anotados (lincRNAs).

En conclusión, aunque se ha podido determinar que muchos lncRNAs son importantes en la regulación génica, hasta hace poco se desconocían los mecanismos mediante los cuales estas moléculas llevaban a cabo su función reguladora. En la actualidad, podemos decir que el número de lncRNAs que se han estudiado y caracterizado funcionalmente es mínimo con respecto a los miles de transcritos no-codificantes presentes en el genoma (más de 15,000) (Sun *et al.*, 2018). Sin embargo, proyectos como el que se expone en esta tesis contribuye al entendimiento del grado en que la transcripción no-codificante ubicua es relevante en la fisiología celular. A continuación, se describirán las principales características y las evidencias que se tienen a la fecha acerca del papel que tiene el último tipo de secuencias no-codificantes: las secuencias repetidas.

## **C. Los elementos transponibles (ETs) o secuencias repetidas**

### **1. El mundo inexplorado de las secuencias repetidas**

A partir del descubrimiento de los elementos transponibles (ETs) en maíz por parte de Barbara McClintock, se abrió una puerta oscura e intrigante sobre sus características y funciones (McClintock, 1950). Al lo largo de los años fue cada vez más claro que los ETs y distintos tipos de secuencias repetidas fueron colonizando genomas cada vez más complejos y a falta de herramientas para entender sus funciones; fueron nombradas como ADN parásito (Modzelewski *et al.*, 2022). Sin embargo los análisis genómicos iniciales, alejados de los alcances de la biología computacional de hoy en día, sugirieron que la inserción exógena de secuencias repetidas y la transposición contribuyen a la evolución de los genomas (Haws *et al.*, 2022; Modzelewski *et al.*, 2022). Con el transcurso de los años y el refinamiento de los métodos de secuenciación de genomas completos, ha sido posible proponer que

entre el 50-60% del genoma humano está compuesto por diversos tipos de secuencias repetidas (Tomilin, 1999; Haws *et al.*, 2022; Hoyt *et al.*, 2022). Incluso utilizando tecnologías que permiten la secuenciación de regiones mucho más largas, se ha logrado recientemente completar en su totalidad la secuencia del genoma humano, ensamblando todas las secuencias repetidas en una versión (T2T-CHM13) que antes no había sido posible incluirlas por completo (Haws *et al.*, 2022).

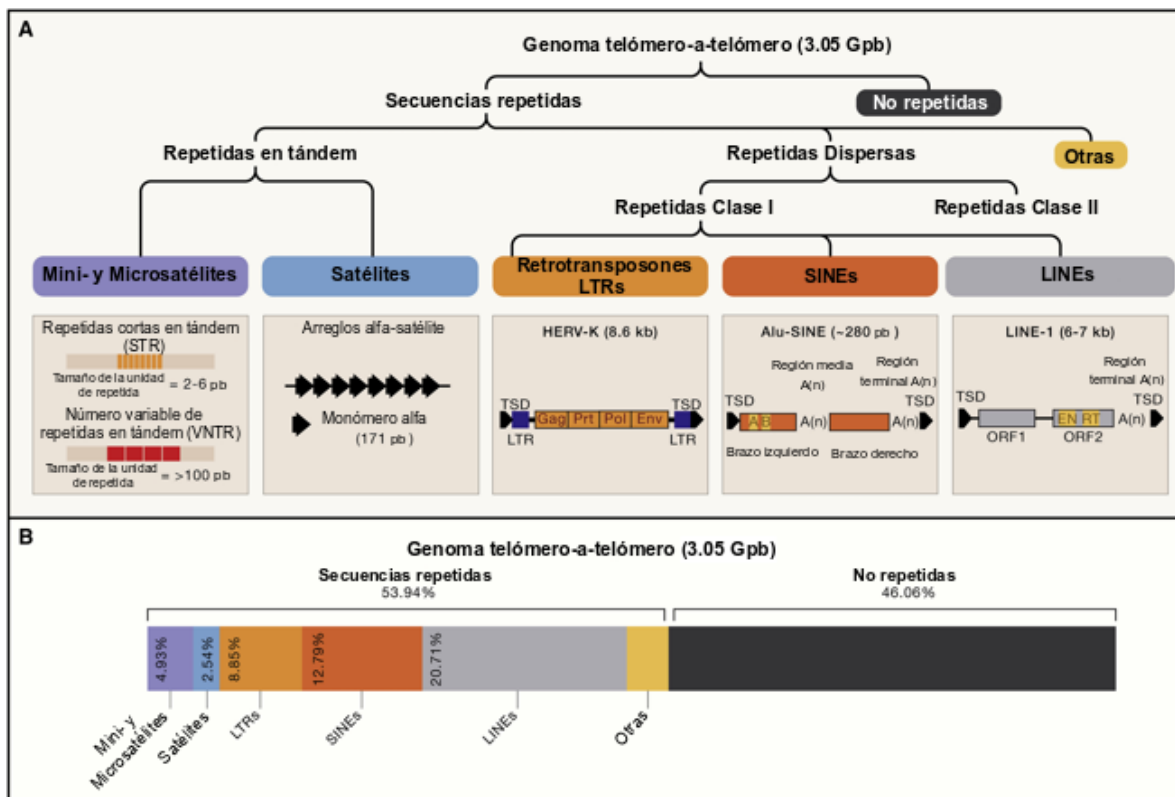
## **2. Clasificación y sus características**

Derivado del gran número de secuencias repetidas en el genoma humano se han dividido en dos grandes grupos basados en su distribución y características particulares, las repetidas en tándem y las repetidas dispersas (Figura 5). Cada una se ha visto con frecuencias de distribución distintas en el genoma humano asociándose a elementos genómicos relacionados con su posible función (Figura 5B). Las repetidas en tándem son secuencias de ADN no móviles en las que las copias de uno o más nucleótidos se repiten de forma continua. Se dividen en secuencias repetidas satelitales o mini/microsatelitales. Por ejemplo, en ratón y humano las regiones pericentroméricas y centroméricas están formadas por repetidas satelitales. Las secuencias centroméricas en el genoma humano están formadas por segmentos de secuencias repetidas  $\alpha$ -satélite de 171 pb, mientras que en la región pericentromérica existen 6 distintas clases de secuencias repetidas satelitales (Haws *et al.*, 2022).

Dentro de las repetidas dispersas se subdividen las repetidas Clase I y II, que a su vez la Clase I comprende a los LTRs (Long Terminal Repeats), los elementos o repetidas nucleares cortas dispersas o SINEs y las repetidas nucleares largas dispersas o LINEs que son las más abundantes (~20% de ocupación en el genoma). Por ejemplo, las LINE-1, tienen una longitud de 6 Kb y codifican su propia maquinaria necesaria para su retrotranscripción y transposición (Figura 5A).



Cabe resaltar que, dentro de las SINEs, la clase más común son las repetidas Alu, las cuales se han clasificado a su vez en tres grandes subfamilias basadas en su edad evolutiva (AluJ, AluS y AluY). Son secuencias repetidas no autónomas que significa que dependen de la maquinaria codificada en LINE-1 para facilitar su transposición. El tipo de secuencia repetida en la cual nos enfocamos para su estudio en el presente trabajo de tesis doctoral fue una de la subfamilia AluSx y cuya selección justificaremos más adelante (Giordano *et al.*, 2007).



**Figura 5. Tipos y distribución de elementos repetidos en el genoma humano.** **A.** El genoma humano consiste en un 54 % de secuencias repetidas. Las secuencias repetidas se dividen por su distribución en dos grandes grupos: las que se encuentran dispersas o las distribuidas en tándem. A su vez, las secuencias dispersas de Clase I comprenden elementos retrotransponibles, que incluyen las LINE (gris), las SINE (naranja oscuro) y los retrotransposones LTR (ERV, naranja claro). El otro tipo de secuencias repetidas son las que se encuentran en tándem que se distribuyen por todo el genoma y son más frecuentes como repeticiones de satélites en regiones de cromatina pericentroméricas y centroméricas. **B.** El reciente ensamblaje del genoma humano de telómero a telómero sin espacios (T2T) consta de ensamblajes para los 22 autosomas humanos y el cromosoma X y comprende 3.05 pares de gigabases (Gbp) de ADN nuclear (CHM13v1). El diagrama de barras, representan las estimaciones recientes del porcentaje por cada clase de repetidas en el genoma humano. TSD, duplicación del sitio blanco; ORF, marco abierto de lectura; EN, endonucleasa; RT, transcriptasa reversa; A y B, partes del promotor de ARN de Pol III; LTR, secuencias terminales repetidas largas; Gag, antígenos de grupo; Prt, proteasa; Pol2, transcriptasa reversa; Env, proteína de la envoltura. (modificada de Haws, 2022).

Un aspecto importante que ha distinguido a los dos grupos de repetidas, es el mecanismo de propagación por medio del cual se han insertado en los genomas a lo largo de la evolución. Las repetidas en tándem que incluyen a las secuencias satelitales y las repetidas cortas en tándem, son secuencias repetidas no-movibles. Mientras que los transposones utilizan un mecanismo de “cut-and-paste” (corte y pegado) donde secuencias propias del transposón (las “terminal internal repeats”) son reconocidas y unidas por la transposasa para facilitar su transposición e integración a un nuevo sitio del genoma huésped.

### **3. Su función en el tiempo y espacio**

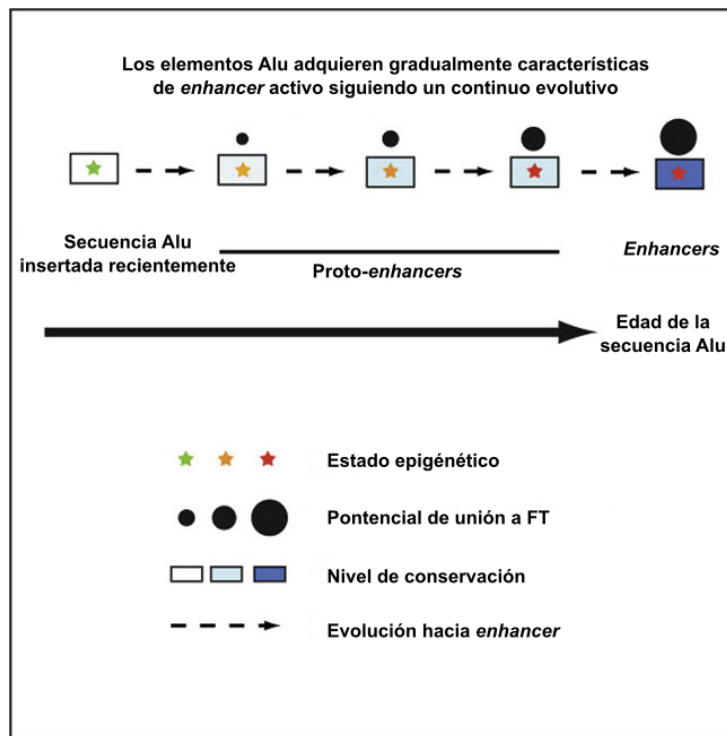
Por mucho tiempo, las secuencias repetidas habían sido prácticamente ignoradas, en gran medida por desconocer sus características genómicas, pero sobre todo por las dificultades técnicas que afrontaron muchos grupos de investigación para su manipulación y estudio. Hoy en día los avances en diversos campos de la biología celular y molecular específicamente a nivel epigenético; han permitido los recientes hallazgos acerca de la relevancia y función de los distintos tipos de secuencias repetidas.

Una de las propuestas que durante años se ha planteado es que la transposición y los cambios derivados de la integración de secuencias repetidas, ha contribuido en gran medida a la evolución de los genomas. El interés creciente por entender el papel de estas secuencias en el genoma, ha motivado el proponer y buscar nuevas y múltiples funciones. Por ejemplo, esta evolución en los genomas se ve reflejada en la creación de nuevos sitios de reconocimiento a factores de transcripción, nuevos mecanismos de regulación de la expresión génica y hasta modificaciones en la organización tridimensional del genoma (Friedli & Trono, 2015; Haws *et al.*, 2022). Además, las secuencias repetidas contribuyen a la formación de regiones muy extensas de heterocromatina, con distintas características, participando así, en la estabilidad del genoma y la organización de su epigenoma.

En base a lo anterior y el hecho de que este tipo de secuencias se encuentran presentes desde bacterias hasta eucariontes, existe la propuesta que plantea que la domesticación de ETs ha permitido la selección de elementos regulatorios que le otorgan un beneficio evolutivo a su huésped a distintos niveles (Feschotte, 2008; Jangam *et al.*, 2017).

De manera particular las secuencias repetidas SINEs tipo Alu, se han relacionado mayoritariamente con actividades regulatorias en el genoma. Estas secuencias que derivan del ARN 7SL, tienen un tamaño de ~300 pb y se localizan mayoritariamente en regiones ricas en genes. Son secuencias específicas de primates ocupando el 11% de su genoma, lo cual equivale a más de 1 millón de copias (McPherson *et al.*, 2001; Batzer & Deininger, 2002). Son elementos que su transcripción y su subsecuente retrotransposición, puede ser regulada por metilación del ADN debido a que poseen una alta densidad de CpGs. Prueba de ello es que en procesos patológicos como es el cáncer o durante el envejecimiento, se observa que los niveles de metilación en las secuencias repetidas se alteran, relacionándose con la inestabilidad genómica presente en ambos escenarios (Liu & Schmid, 1993; Ade *et al.*, 2013; Luo *et al.*, 2014). Interesantemente este tipo de secuencias contiene un gran número de sitios de unión a factores de transcripción, los cuales no solo podrían estar implicados en la regulación de la transcripción de sí mismos sino influir por igual, la transcripción de genes aledaños (Polak & Domany, 2006; Deininger, 2011). Durante los últimos años concretamente, estas secuencias han sido caracterizadas a nivel genómico y se ha visto que son regiones con hipersensibilidad a la DNasa I además de presentar enriquecimiento de las marcas en la histona H3K4me1, H3K4me2, H3K27me1, H3K9ac, H3K36me3 y otras modificaciones asociadas con cromatina abierta y *enhancers* (Barski *et al.*, 2007; Zhang *et al.*, 2019). Además, se ha determinado que este tipo de repetidas forman contactos con promotores o regiones cercanas a sitios de inicio de la transcripción. Al analizar este tipo de características tomando en cuenta su edad evolutiva, se ha visto que las subfamilias más viejas (AluSx, AluJo y AluJb específicamente) presentan modificaciones de histonas típicas de *enhancers*, ganancia de motivos de unión a FTs y un alto grado de conservación. Características que apoyan la teoría

de que este tipo de repetidas representan los ancestros evolutivos de secuencias regulatorias tipo *enhancer* (Su *et al.*, 2014) (Figura 6).



**Figura 6. Modelo propuesto de elementos Alu como “proto-*enhancers*” que siguen un continuo evolutivo después de la inserción en el genoma que evolucionan hacia *enhancers* funcionales.** El nivel de conservación se refiere a la conservación entre humanos y chimpancés, donde el color más oscuro representa un mayor nivel de conservación. El estado epigenético se refiere al estado de la modificación de histonas en un elemento Alu (verde, inactivo; amarillo, medio; rojo, activo). El tamaño de los círculos representa la posibilidad de unión de reguladores transcripcionales a los elementos Alu (figura modificada de Su *et al.*, 2014).

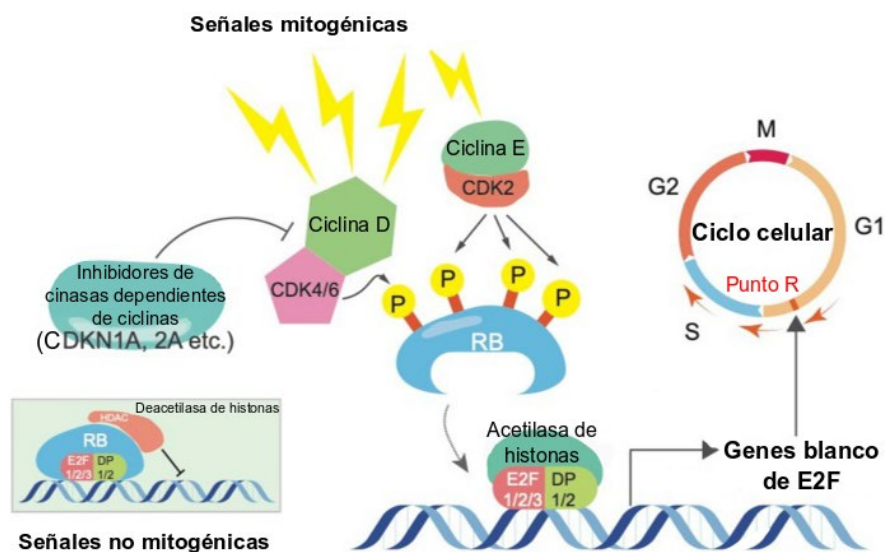
Con respecto a la función de este tipo de repetidas se han relacionado con la regulación a diferentes niveles, ya que no solo se ha visto que intervienen a nivel transcripcional sino también en la edición de mensajeros y traducción (Häsler & Strub, 2006). Existen ejemplos de regulación transcripcional positiva (*enhancer* específico de células T del gen humano de CD8a, *enhancer* del receptor de estrógenos, promotor de la haptoglobina, el gen de la cadena Fc(RI<sub>V</sub>) así como ejemplos de regulación negativa (*APOA5*, *WT1*, *ε-globin*, *GH*, *poly(ADP-ribosil)transferasa* (Wu *et al.*, 1990; Hewitt *et al.*, 1995; Sharan *et al.*, 1999; Trujillo *et al.*, 2006).

A pesar que en la literatura se cuenta con una gran cantidad de evidencias indirectas como son las observaciones correlativas tanto de secuencia como de perfiles cromatínicos de las repetidas Alu y su posible función, existen en la actualidad muy pocas evidencias experimentales directas de las repetidas Alu en la regulación del genoma humano (Feschotte, 2008).

Por lo anterior, en el presente trabajo nos interesa conocer el papel regulatorio de este tipo de repetidas, de manera particular de secuencias tipo Alu localizadas en genes relevantes como el gen de *Retinoblastoma* humano (*RB1*). Así como investigar la función regulatoria de las Alu a nivel genómico y su relación con la expresión de regiones cercanas a ellas.

## II. ANTECEDENTES

El gen de *Retinoblastoma* (*RB1*) ha sido uno de los primeros genes supresores de tumores descritos y estudiados con múltiples funciones en células normales. De manera recurrente, se encuentra genética y/o epigenéticamente inactivo en células tumorales de diversos tejidos. Mutaciones y deleciones en él, han sido asociados a varios tipos de cáncer, pero sobre todo al retinoblastoma. Se sabe que la proteína Rb (pRb) interactúa con factores de la familia E2F y remodeladores de la cromatina para asociarse a promotores de genes relevantes para la regulación y progresión del ciclo celular (Hatakeyama & Weinberg, 1995; Giacinti & Giordano, 2006; Sanidas *et al.*, 2022). Por otra parte, señales mitogénicas activan la función de cinasas dependientes de ciclinas quienes fosforilan a pRb, favoreciendo la activación de genes dependientes de E2F y la subsecuente progresión del ciclo celular (Knudsen & Knudsen, 2008). Por lo tanto, parte del control en la función de pRb está dada, en su forma activa por un estado no-fosforilado, mientras que su estado inactivo es hiperfosforilado (Figura 7).



**Figura 7. La función canónica de RB en el control de la progresión del ciclo celular.** Los complejos CDK dependientes de ciclinas regulan la actividad de RB a través de la fosforilación, mientras que los inhibidores de cinasas dependientes de ciclinas suprimen la actividad de los complejos ciclina-CDK. La RB fosforilada libera E2F, que estimula a las células a entrar en las fases S y M a través de la expresión de genes blanco de E2F, como la ciclina A y la E (Figura modificada de (Kitajima *et al.*, 2020))

Sin embargo, la visión clásica de cómo pRb contribuye al control del ciclo celular ha sido, por muchos años, algo sin completar por dos razones. La primera razón tiene que ver con que la fosforilación de pRb no es la única forma de regular sus funciones. Además remodeladores de la cromatina como SUV39H1/2 (metiltransferasa específica de H3K9me3) y SWI/SNF (complejo remodelador de la cromatina ATP-dependiente) se asocian a pRb y modifican la estructura de la cromatina de genes asociados a la transición G1/S y G2/M, respectivamente (Giacinti & Giordano, 2006). Y la segunda razón recientemente descrita, es que pRb tiene diversas funciones que han sido descubiertas gracias a estrategias experimentales como el ChIP-seq, con lo cual se ha redefinido la distribución de pRb a lo largo de todo el genoma humano (Sanidas *et al.*, 2022). De esta forma, se ha demostrado que pRb se asocia a diferentes categorías de genes mediante su asociación indirecta a promotores, *enhancers* y un subconjunto de regiones genómicas vinculadas al factor nuclear CTCF (Sanidas *et al.*, 2022). Y es así como se propone que en la fase G1 del ciclo celular, pRb junto con E2F, factores y cofactores asociados; verifican la integridad y la adecuada progresión del ciclo celular. Una vez ocurrido lo anterior, pRb se redistribuye a otros promotores, *enhancers* e *insulators* (CTCF) para regular programas génicos que complementan el control del ciclo celular y otras funciones celulares.

En resumen, pRb es un supresor tumoral clave en el control del ciclo celular. Su regulación transcripcional es de suma importancia para sintetizar cantidades de proteína pRb suficientes para llevar a cabo sus múltiples funciones. Por esta razón, en este trabajo de tesis, hemos escogido al gen *RB1* humano como locus modelo para el estudio de su regulación epigenética y transcripcional, con particular interés en el papel que tienen las secuencias repetidas tipo Alu.

Desde hace varios años, hemos realizado en nuestro grupo de investigación, estudios sobre la regulación del gen del retinoblastoma humano (De La Rosa-Velázquez *et al.*, 2007; Recillas-Targa *et al.*, 2011; Dávalos-Salas *et al.*, 2011). El gen *RB1* a nivel de secuencia contiene en su promotor varios sitios de unión a diferentes factores de transcripción así como a CTCF, un factor *insulator*

involucrado en la regulación epigenética y la estructura tridimensional de la cromatina (Ohlsson *et al.*, 2001; Recillas-Targa, 2002; Ling *et al.*, 2006).

Además en trabajos previos, se reportó acerca de los mecanismos involucrados en la regulación de este *locus*, particularmente el papel de CTCF como un componente epigenético necesario para la estructura de la cromatina y su integridad funcional (De La Rosa-Velázquez *et al.*, 2007). Así como el análisis de los niveles de metilación y de las marcas de cromatina asociadas a su silenciamiento en diferentes líneas celulares (Dávalos-Salas *et al.*, 2011).

A pesar de los hallazgos obtenidos sobre la regulación de este promotor, se ha planteado que existan otros mecanismos involucrados en su regulación epigenética que expliquen el silenciamiento observado en procesos patológicos como el cáncer. Por lo tanto, la relevancia de estudiar la función de las repetidas como elementos de regulación en este *locus*, no solo permitirá el caracterizar la función de estas secuencias tan abundantes en el genoma sino de conocer aún mejor la regulación de un gen maestro que participa en el control del ciclo celular.



### **III. PLANTEAMIENTO DEL PROBLEMA**

El genoma humano está compuesto en una alta proporción por elementos transponibles, particularmente por las repetidas tipo Alu quienes se han seleccionado como la clase más abundante en primates. A pesar de que han sido propuestas como elementos que dieron origen a la gran diversidad de secuencias regulatorias, existen pocas evidencias directas que demuestren su actividad funcional sobre elementos regulatorios, estructurales o incluso secuencias codificantes aledañas a este tipo de secuencias repetidas. Por lo cual en este trabajo de tesis propongo el estudio de las secuencias repetidas tipo Alu para entender su contribución en el contexto de la región genómica río arriba del gen retinoblastoma humano.

### **IV. HIPÓTESIS**

Secuencias repetidas tipo Alu localizadas río arriba del gen *RB1* humano, presentan una actividad regulatoria sobre el *locus* que comprende a este gen.

### **V. OBJETIVO**

Determinar la actividad de una secuencia repetida tipo Alu y su participación en la regulación de la expresión del gen *RB1* humano.

## VI. MATERIALES Y MÉTODOS

### Construcciones de plásmidos

Se utilizó el plásmido *pRB1prom-GFP* que contiene el promotor del gen *RB1* caracterizado y clonado en el plásmido *pEGFP* (*pEGFP-1*; Clontech, Palo Alto, CA; De La Rosa-Velázquez *et al.*, 2007). Este fragmento de 401 pb es la región intergénica entre *LINC00441* y el gen *RB1* humano (chr13: 48,877,623-48,878,023; versión h19). La secuencia repetida tipo *AluSx* de 287 pb, localizada río arriba del promotor del gen *RB1* (secuencia repetida *AluSx1* chr13: 48,874,474-48,874,760) fue amplificada utilizando unos oligonucleótidos específicos para esta secuencia repetida (Tabla 1) utilizando como templado ADN genómico de linfocitos humanos y subclonado en el *pRB1prom+GFP* en dos orientaciones para generar plásmidos *pAlu(5'-3')+RB1prom+GFP* y *pAlu(3'-5')+RB1prom+GFP*.

Todos los plásmidos contienen un *cassette* de resistencia a la neomicina, que permite la selección por G418 de células transfectadas de forma estable. La integridad de todas las construcciones de plásmidos se confirmó mediante secuenciación de ADN.

### Cultivo de células

Se cultivaron células eritroleucémicas humanas K562 en medio ISCOVE (Invitrogen). Las células K562 (K562 ATCC®CCL-243™) fueron proporcionadas por Gary Felsenfeld (NIH, Bethesda, Maryland, EE. UU.) y se cultivaron en DMEM. Todos los medios contenían 10% (v/v) de suero fetal bovino (FBS) y 1% penicilina/estreptomicina y se mantuvieron en una incubadora a 37°C con 5% CO<sub>2</sub>. Se obtuvieron linfocitos humanos de sangre periférica de un donante sano, aislado con Ficoll-Paque Plus 2 (Amersham) siguiendo las instrucciones recomendadas por el fabricante. Se obtuvo el consentimiento informado por escrito de este donante sano.

## **Transfección estable de células K562**

Las células K562 se lavaron dos veces con solución amortiguadora de fosfatos (PBS) y resuspendieron en DMEM. Se utilizaron un total de  $3 \times 10^5$  células K562 por cada ensayo de transfección. A continuación, las células se transfirieron a una placa de 6 pozos y se transfectaron con  $1 \mu\text{g}$  ( $1 \mu\text{g}/\mu\text{l}$ ) de los plásmidos linearizados correspondientes utilizando Lipofectamine 2000 (Invitrogen, MA, EE. UU.) según las instrucciones del fabricante. Después de 6 hrs, se agregaron 4 ml de medio no selectivo a las células transfectadas. Transcurridas 48 hrs de recuperación de las células, se transfirieron a medios que contenían 0,9 mg/ml de G-418 (Geneticin, Calbiochem) para su selección. Las transfecciones resistentes a la geneticina fueron analizadas por citometría de flujo (FACS) en diferentes tiempos (día 0, día 15, día 25, día 40 y día 60) para obtener el porcentaje (%) de células positivas para *GFP* y la media de intensidad de fluorescencia para cada construcción por cada punto en el tiempo analizado. Los datos fueron registrados y compilados con el software BD CellQuest Pro (BD Biosciences). Se realizaron cuatro experimentos independientes, cuyos datos fueron analizados estadísticamente con el software Graphpad Prisma 7.0. Las diferencias estadísticamente significativas de las medias de intensidad de fluorescencia entre las construcciones que contenían a la secuencia repetida Alu y las que no la tenían, fueron calculadas utilizando la prueba de Mann-Whitney de dos colas ( $p < 0,05$ ).

## **Aislamiento de clones independientes**

Después de 48 hrs de recuperación celular,  $\sim 5 \times 10^5$  ( $500 \mu\text{l}$ ) de células transfectadas se transfirieron a una matriz de celulosa (Methocel, FLuka) que contiene 0,9 mg/ml de G-418. Después de 2-3 semanas, se fueron aislando colonias individuales en cultivos independientes las cuales se expandieron en medio líquido DMEM que contenía G-418 para realizar experimentos posteriores. Para cada construcción, aislamos y analizamos 14 clones independientes en los diferentes puntos del ensayo curso-temporal realizado (día 0, día 15, día 30, día 45, día 60, día 80 y día 100) durante 100 días de cultivo celular continuo. La integridad del

transgén se verificó mediante PCR y Southern blot. Para las PCRs se utilizaron tres pares de oligonucleótidos diseñados para corroborar por tamaño, la presencia de los elementos en el transgén que tenían que poseer las colonias seleccionadas para su análisis curso-temporal. El par de oligonucleótidos denominado GFP1, determinaba si la construcción poseía al promotor del gen *RB1*, dando un producto de 520 pb; o la secuencia Alu+promotor, amplificando un producto de 820 pb. El segundo par denominado GFP2, determinaba por tamaño si la clona poseía en su totalidad a la región codificante al gen reportero *GFP*, amplificando un producto de 955 pb. Y el tercer par de oligonucleótidos para determinar si este transgén se encuentra integrado en una sola copia o multicopia, ya que al ser oligonucleótidos que se dirigen hacia extremos opuestos, únicamente en las clonas donde se ha integrado más de una copia del transgén, se logra la amplificación de un fragmento de 346 pb (ver Figura 14 y Tabla 1).

### **Eliminación dirigida mediada por CRISPR-Cas9**

En la línea celular eritroleucémica K562, fue eliminado el elemento AluSx (chr13: 48 874 474-48 874 760, Figura 8) localizado río arriba del locus del promotor del gen *RB1*. Se realizó mediante la cotransfección de dos plásmidos que contienen Cas9, el plentiCRISPR-v2 (Addgene Cat. 52961; Cambridge, MA, EE. UU.). Para cada plásmido se sintetizaron ARN guías (sgRNA) que reconocen cada uno de los límites (5' y 3') de la secuencia Alu específicamente en su contexto endógeno. El diseño de los sgRNAs para incorporar al sistema de CRISPR-Cas9 se hizo con el software CRISPOR1 tal y como lo describe el desarrollador (Haeussler *et al.*, 2016) para minimizar el reconocimiento de secuencias inespecíficas ("off-targets"). También se agregaron en las guías, sitios de restricción para BsmB1 para facilitar su integración al plásmido plentiCRISPR-v2. Los oligonucleótidos fueron alineados siguiendo un protocolo estándar (Sambrook & Russell, 2006), ligados al vector y confirmados por secuenciación para su posterior transfección a la línea celular K562 (Tabla 1).

Para la transfección de los vectores se utilizó Lipofectamine 2000 (Invitrogen) basados en el protocolo del fabricante. Después de permanecer las transfecciones

en cultivo durante 3-4 días en medio selectivo con puromicina (5 µg/ml; Sigma), obtuvimos ADN genómico de los “pools” de células transfectadas con el sistema CRISPR-Cas9. Se utilizó este ADN genómico como templado para analizar por PCR, la eliminación del fragmento esperado, la secuencia repetida AluSx. Los oligonucleótidos utilizados fueron diseñados fuera de la región identificada por las guías (sitio de ruptura “double strand break, DSB”) y en caso de no haber ocurrido una eliminación, se obtendría un fragmento de ADN de 844 pb. Cabe aclarar que dada la heterogeneidad de moléculas mutantes al evaluar las eliminaciones en “pool”, resulta difícil obtener enriquecimiento en un solo fragmento mutante y solo se observará un barrido por debajo del fragmento silvestre. Por lo tanto, para poder aislar células monoclonales de una misma eliminación, se sembraron los “pools” de las transfecciones CRISPR-Cas9 a muy baja densidad para que mediante dilución en serie en una placa de 96 pozos pudiéramos aislar eventos independientes. Cuarenta y ocho colonias resistentes a la puromicina fueron seleccionadas al azar, individualmente expandidas y divididas para cultivo. Se aisló su ADN genómico para su posterior análisis del fragmento escindido en cada una de ellas. Cien nanogramos de ADN genómico de estos clones o el ADN genómico de la línea celular silvestre K562 (control) se examinaron mediante PCR con pares de oligonucleótidos descritos anteriormente (Tabla 1) y utilizados también en los “pools” de transfección.

Los fragmentos de PCR fueron evaluados en geles de agarosa al 1%. Se cortaron las bandas del gel y se purificaron usando el QIAquick Kit (QIAGEN). La lesión molecular exacta de la eliminación fue confirmada por secuenciación de Sanger. Seleccionamos tres clones homocigotas para dos deleciones diferentes de la secuencia repetida AluSx, para los análisis posteriores ( $\Delta$ Alu-C1,  $\Delta$ Alu-C2 y  $\Delta$ Alu-C3).

## **Extracción de ARN y RT-PCR cuantitativa (RT-qPCR)**

El ARN total se extrajo de las células K562 con el reactivo TRIzol (Invitrogen) de acuerdo a las instrucciones del fabricante. Las concentraciones del ARN se determinaron usando el espectrofotómetro NanoDrop ND-1000 (NanoDrop Inc., DE, EE. UU.). Se realizó PCR cuantitativa en tiempo real (RT-qPCR) utilizando el Kit KAPA SYBR® FAST One-Step RT-qPCR y oligonucleótidos específicos para los transcritos de *RB1*, *LINC00441* y  *$\alpha$ -tubulina* como control de normalización endógena (Tabla 1). Las reacciones de qPCR fueron realizadas en el sistema de detección StepOne (Applied Biosystems) a 42°C durante 5 min, 95°C durante 30 s, seguido de 40 ciclos de tres pasos de 95°C por 3 s, 62°C por 10 s, y 72°C por 10 s, en triplicado para cada muestra. Se calcularon los niveles relativos de ARN utilizando el método comparativo  $\Delta\Delta C_t$  (Schmittgen & Livak, 2008). Diferencias estadísticamente significativas en la expresión génica entre los mutantes por CRISPR y la línea silvestre, se calcularon utilizando una prueba de t ( $p < 0,05$ ) y Graphpad Prisma Software 7.0.

## **Ensayo de proliferación celular**

Utilizamos en el ensayo azul de tripano para determinar la viabilidad celular mediante el conteo del número de células viables con un microscopio de la siguiente manera: la línea celular K562 WT frente a  $\Delta$ Alu-C1,  $\Delta$ Alu-C2 y  $\Delta$ Alu-C3 (Strober, 2001). El día 0 empezamos con  $1 \times 10^5$  células en un volumen de 3 ml por triplicado. Contamos con ayuda de un hemocitómetro el número de células para cada condición cada 24 hrs durante 4 días. Finalmente, graficamos los valores promedio de los triplicados del número celular en función del tiempo para las diferentes líneas celulares.

## **Análisis Bioinformático**

### **Análisis de datos de motivos**

La identificación de los sitios de unión para los FTs se realizó utilizando la base de datos de vertebrados del programa JASPAR (Fornes *et al.*, 2020) utilizando un umbral de  $p < 0,0001$ .

### **Análisis de la distribución de secuencias repetidas tipo Alu en lincRNAs y su correlación con los niveles de expresión**

Los análisis bioinformáticos se realizaron con la versión del Genoma Humano hg19/NCBI 37. Tanto las coordenadas cromosómicas de todas las secuencias repetidas anotadas para el genoma humano en Repeat Masker (Jurka, 2000) como los valores de expresión de todos los genes derivados de los diferentes experimentos de RNA-seq del consorcio GTEx (G. TEx Consortium, 2013), fueron descargados usando la herramienta Table Browser del UCSC Genome Browser (Karolchik *et al.*, 2004). Las coordenadas cromosómicas para todos los genes codificantes de proteínas y no-codificantes de GENCODE se descargaron de ENSEMBL Biomart. Para asignar secuencias Alu a los lincRNAs, utilizamos Bedtools (Quinlan & Hall, 2010) `intersect -wao` y se realizaron análisis adicionales en R versión 3.4. Los lincRNAs se clasificaron como superpuestos a Alu, si al menos una secuencia de Alu estaba completamente contenida dentro del no-codificante; o libres de Alu, si ninguna secuencia de Alu traslapa con él. Para relacionar la presencia de secuencias Alu con los niveles de expresión de los lincRNAs, calculamos el nivel medio de expresión de cada lincRNA en los 53 grupos de datos de RNA-seq contenidos en la tabla de datos GTEx y trazamos la distribución de los valores medios de  $\log_2(\text{RPMK}+1)$  para todos los lincRNAs en las categorías sin-Alu o Alu-superpuesto. Finalmente, trazamos los valores medios de expresión para todos los lincRNA clasificados según el tipo de subfamilia Alu que se superpone al no-codificante. Es importante destacar que un solo lincRNA puede albergar múltiples secuencias Alu de diferentes subfamilias.

**Tabla 1.** Secuencias de oligonucleótidos utilizados.

<b>NOMBRE</b>		<b>Secuencia 5'-3'</b>	<b>Técnica en que fue empleado y figura relacionada</b>
<b>Rb</b>	<b>FORWARD</b>	5'-CGGGATCCAGACTCTTTGTATAGCC-3'	Amplificación por PCR para construcción de plásmidos para ensayos reporteros, Figura 8
	<b>REVERSE</b>	5'-CGGGATCCGAGCTGTGGAGGAG-3'	
<b>Alu(S1)</b>	<b>FORWARD</b>	5'-GATGGGGTCTTGCTCTTG-3'	Amplificación por PCR de la secuencia repetida AluSx para ensayos reporteros, Figura 8
	<b>REVERSE</b>	5'-CTAGACCAGGTACGGTGG-3'	
<b>GFP1</b>	<b>FORWARD</b>	5'-GCTACCGGACTCAGATCTC-3'	Amplificación por PCR de la región 5'-promRb-3' o 5'-Alu+promRb-3' sobre las construcciones en ensayos reporteros, Figura 13
	<b>REVERSE</b>	5'-CTCCTCGCCCTTGCTCAC-3'	
<b>GFP2</b>	<b>FORWARD</b>	5'-GTGAGCAAGGGCGAGGAG-3'	Amplificación por PCR de la región codificante para la GFP sobre las construcciones en ensayos reporteros, Figura 13
	<b>REVERSE</b>	5'-ACATTGATGAGTTGGACAAAC-3'	
<b>GFP3</b>	<b>FORWARD</b>	5'-CACTGCATTCTAGTTGTGGT-3'	Amplificación por PCR del extremo 3' sobre las construcciones en ensayos reporteros. Solo en clones multicopias, amplifica por diseñarse en sitios divergentes de los plásmidos, Figura 13
	<b>REVERSE</b>	5'-CTTTAGGGTCCGATTAGTG-3'	
<b>sgRNA_Alu 5'</b>	<b>SENSE</b>	5'-CACCGGTGAATTTCAAGAAGTTCAGG-3'	RNAs guía (sgRNAs) utilizados para la delección por CRISPR/Cas9, Figura 9
	<b>ANTISENSE</b>	5'-AAACCCTGAACCTCTTGAAATTCACC-3'	
<b>sgRNA_Alu 3'</b>	<b>SENSE</b>	5'-CACCGGATTCGGCAGAGCTAGACC-3'	RNAs guía (sgRNAs) utilizados para la delección por CRISPR/Cas9, Figura 9
	<b>ANTISENSE</b>	5'-AAACGGTCTAGACTCTGCCGAAATCC-3'	
<b>Alu(Screen_Big)</b>	<b>FORWARD</b>	5'-GGAAGCAGAGAGAACCAATGG-3'	Amplificación por PCR para evaluar delección por CRISPR/Cas9, Figura 10
	<b>REVERSE</b>	5'-CACACCACCATAGCACACAT-3'	
<b>Alu(Screen_Sm)</b>	<b>FORWARD</b>	5'-AGGAGTCTTACAGCAATCTTCTT-3'	Amplificación por PCR para evaluar delección por CRISPR/Cas9, Figura 10
	<b>REVERSE</b>	5'-CATCCACACAACCTCAGAAGCA-3'	
<b>Rb_exp</b>	<b>FORWARD</b>	5'-GGATCAGATGAAGCAGATGGAAG-3'	Amplificación por RT-PCR del exon 27 del gen Rb para evaluar su expresión, Figura 16
	<b>REVERSE</b>	5'-GCATTCGTGTTTCGAGTAGAAGTC-3'	
<b>linc00441_exp</b>	<b>FORWARD</b>	5'-GGCTGGGACGCTAAGTCATG-3'	Amplificación por RT-PCR del exon 1 del linc00441 para evaluar su expresión, Figura 16
	<b>REVERSE</b>	5'-GGGGTGGTTCTGGGTAGAAG-3'	
<b>a_TUB</b>	<b>FORWARD</b>	5'-TGGAACCCACAGTCATTGATGA-3'	Amplificación por RT-PCR del exon 2 del gen $\alpha$ -Tubulina para evaluar su expresión, Figura 16
	<b>REVERSE</b>	5'-TGATCTCCTTGCCAATGGTGTA-3'	

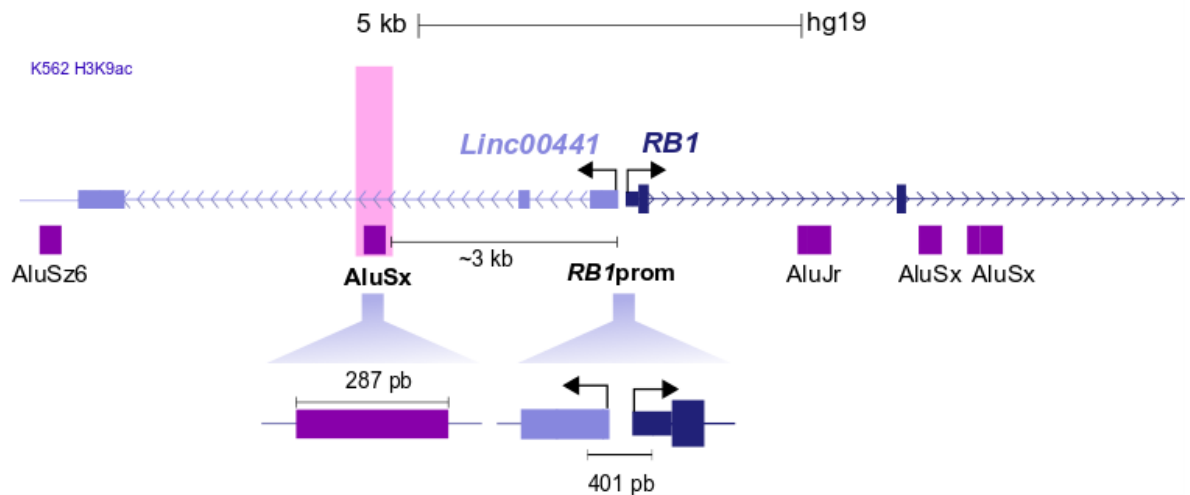


## VII. ESTRATEGIA EXPERIMENTAL

### 1. Determinar la actividad de la secuencia repetida AluSx localizada río arriba del promotor del gen *RB1*.

Para caracterizar el posible efecto regulatorio en *cis* de las secuencias repetidas tipo Alu en este *locus*, fue necesario analizar la distribución de secuencias repetidas y seleccionar candidatas para realizar los ensayos de gen reportero. Por lo que se obtuvo la ubicación de las secuencias repetidas Alu próximas al promotor mínimo del gen *RB1*, que incluye secuencias 5 Kb río arriba y abajo del sitio de inicio de la transcripción (TSS), basados en la anotación de Repeat Masker (Figura 8). Las dos secuencias repetidas tipo Alu más cercanas al promotor del gen *RB1* son una secuencia de la subfamilia AluSx y una de la subfamilia AluJr. La secuencia repetida AluSx se encuentra a 3 Kb río arriba del TSS del gen *RB1* que, a su vez, se localiza dentro del segundo intrón del *LINC00441*; un lincRNA divergente en relación al gen *RB1*. La secuencia repetida AluJr está localizada en el primer intrón del gen *RB1*, aproximadamente 2.5 kb río abajo del TSS. Ambas secuencias repetidas fueron caracterizadas para múltiples TFBs así como su asociación con los distintos estados de la cromatina definidos por ENCODE (Proyecto ENCODE Consorcio, 2012). Este *track* muestra una segmentación del genoma en 15 estados de la cromatina definidos mediante la integración computacional de datos de ChIP-seq para nueve factores y/o marcas de histonas en nueve líneas celulares humanas. Cada segmento está coloreado y asociado a algún elemento funcional (*insulator*, *enhancer*, promotor, etc.) y/o actividad transcripcional. También se evaluó el enriquecimiento de marcas de histonas, principalmente aquellas que se asocian a elementos funcionales bien caracterizados con el fin de relacionar con la posible función de estas repetidas en el *locus* del gen *RB1*.

Posterior a la selección de la secuencia repetida Alu, se generaron las distintas construcciones utilizando al gen reportero *GFP* para evaluar el efecto sobre su promotor más cercano.

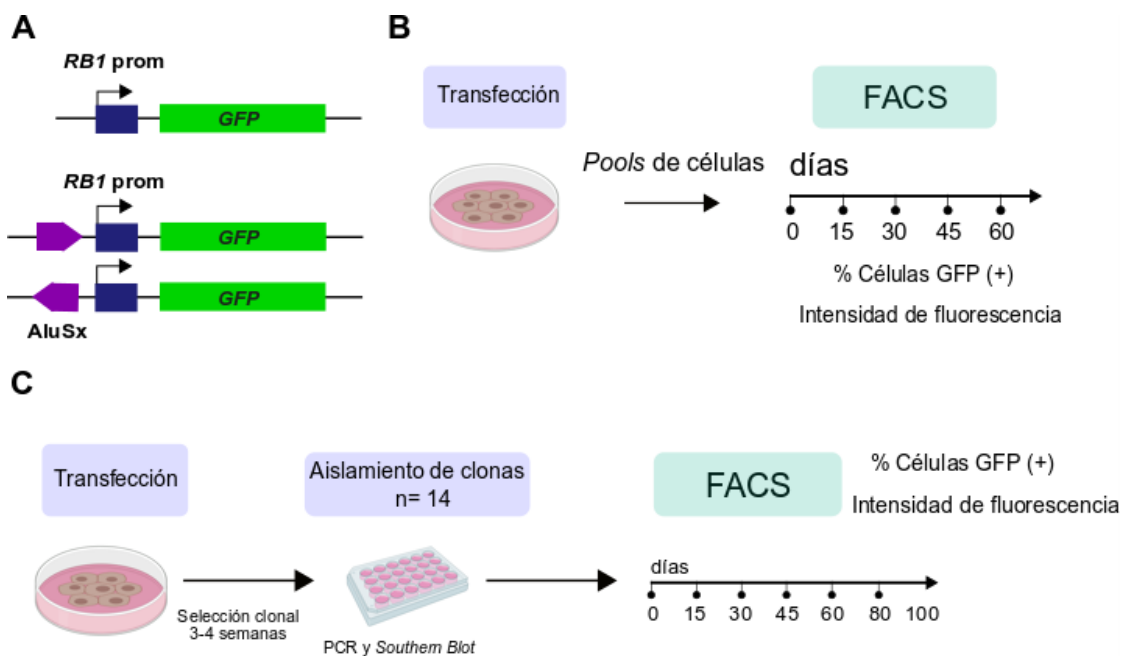


**Figura 8. Región del locus de *RB1* humano en el cromosoma 13.** Se indican con azul las regiones amplificadas por PCR para ser clonadas en el plásmido reportero pEGFP, *RB1prom* que corresponde a la región del promotor del gen *RB1* (401 pb) y la secuencia AluSx (287 pb) resaltada en recuadro rosa. En recuadros morados se representan los sitios donde se localizan las diferentes secuencias Alu en esta región genómica.

Se generaron líneas estables en la línea celular eritroleucémica K562 con tres construcciones que contenían al promotor del gen *RB1* (Figura 9A), donde dos de ellas contenían la repetida Alu (pAlu+*RB1* prom+GFP) en ambas orientaciones (5'-3' y 3'-5') y la otra en ausencia de ella (p*RB1* prom+GFP).

- a) Análisis por *pools*: Una vez generadas las construcciones, se realizaron una serie de transfecciones estables. Las células transfectadas fueron seleccionadas con geneticina y se obtuvieron *pools* para ambas construcciones para posteriormente ser evaluadas por citometría de flujo (FACS) durante 60 días a diferentes tiempos (0, 15, 30, 45 y 60 días), obteniéndose datos tanto de número de células positivas a GFP (GFP(+)), así como la intensidad de fluorescencia para cada una de las construcciones analizadas (Figura 9B).
- b) Análisis por clones independientes: Con el objetivo de evaluar el efecto de la secuencia repetida independiente del contexto cromatínico del sitio de integración del transgén, se obtuvieron treinta clones independientes de cada

grupo después de tres semanas de selección en agar en presencia de geneticina. En todas las clonas se evaluó la integridad del transgén a través de PCRs (tres pares de oligonucleótidos distintos diseñados en las diferentes secuencias de los plásmidos). Algunas clonas también fueron corroboradas por Southern blot. Finalmente se evaluaron 14 clonas por FACS durante 100 días a diferentes puntos (0, 15, 30, 45, 60, 80 y 100 días), obteniendo el número de células GFP(+) así como la intensidad de fluorescencia (Figura 9C).



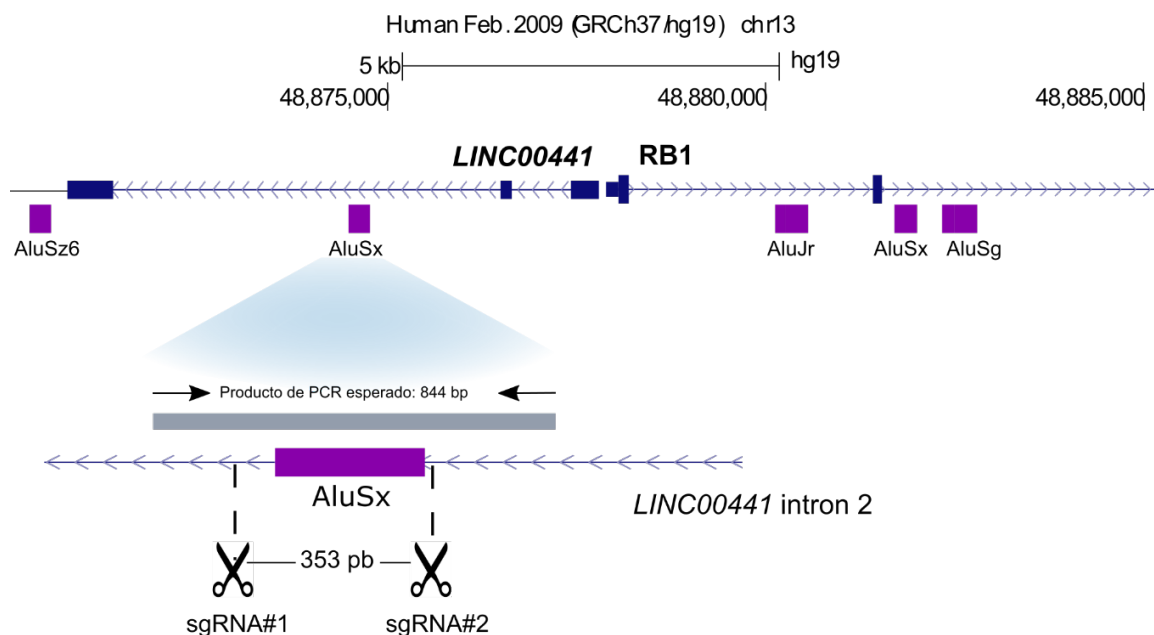
**Figura 9. Estrategia de ensayos de gen reportero para determinar la actividad de la repetida AluSx. A.** Construcciones generadas para evaluar la actividad bioquímica de la secuencia AluSx frente al promotor del gen *RB1*. De arriba hacia abajo: *pRB1*prom+GFP, *pAlu*(5'-3')+*RB1*prom+GFP y *pAlu*(3'-5')+*RB1*prom+GFP. **B.** Transfección de construcciones y análisis por pools. **C.** Transfección de construcciones y análisis por clones independientes.

## 2. Caracterización de la función de la secuencia repetida AluSx en su contexto endógeno.

Con el objetivo de eliminar específicamente la secuencia AluSx río arriba del promotor del gen *RB1* y evaluar su función, se utilizó el sistema de edición génica

CRISPR-Cas9. Se diseñaron dos ARNs guías (sgRNAs) que dirigen un corte de doble cadena (DSB) en la secuencia que delimita a la AluSx de manera específica (chr13:48,874,390-48,874,409; chr13:48,874,757-48,874,776) y que coinciden con el segundo intrón de un ARN largo no-codificante (*LINC00441*) anotado en la última versión disponible del genoma humano (hg19). Los sgRNAs y los plásmidos que contienen el sistema CRISPR-Cas9 fueron cotransfectados en la línea celular eritroleucémica humana K562. Lo anterior permite que los sgRNAs diseñados en los sitios que delimitan a la AluSx, dirijan a la endonucleasa Cas9 para realizar un corte de doble cadena en cada extremo de la secuencia repetida Alu. Posteriormente el sistema de reparación de doble cadena NHEJ, se encarga de unir los extremos libres de ambas hebras de ADN, generando inserciones o deleciones de algunas pares de bases.

Después de 3-4 semanas, se obtuvieron 48 clones independientes por dilución clonal en presencia de puromicina, de las cuales se extrajo ADN genómico para su validación. Se evaluaron por PCR con oligonucleótidos diseñados fuera de la región del DSB para seleccionar aquellas donde se eliminó por completo la Alu o alguna parte de su secuencia (Figura 10).



**Figura 10. Estrategia experimental para la eliminación por la técnica de CRISPR-Cas9 de la secuencia repetida AluSx.** Con las tijeras se representa el sitio donde fueron diseñados los sgRNAs para eliminar una región de ~350 pb donde se encuentra ubicada la AluSx localizada a 3.5 kb río arriba del promotor del gen *RB1*.

Las flechas negras indican la posición de los oligonucleótidos diseñados para confirmar la eliminación, dando un tamaño esperado para la región silvestre de 844 pb, mientras que en clonas donde haya ocurrido la delección, esperamos la amplificación de un fragmento de ADN aproximado de 353 pb.

## **2.1 Evaluación del efecto en la expresión del gen *RB1* y del *LINC00441*.**

Con el fin de evaluar cambios en la expresión de los elementos cercanos a la repetida (*RB1* y *LINC00441*) en ausencia de la secuencia repetida Alu, se extrajo ARN total de las clonas aisladas, únicamente aquellas seleccionadas por presentar lesiones moleculares distintas (eliminación total o parcial de la secuencia repetida AluSx). Se realizó un ensayo de expresión mediante PCR en tiempo real (RT-qPCR) utilizando oligonucleótidos específicos para el gen *RB1*, el *LINC00441* y la  $\alpha$ -tubulina como control endógeno de normalización. Cada ensayo se realizó por triplicado y los niveles relativos de ARN fueron calculados usando el método de  $\Delta\Delta Ct$ . Las diferencias significativas en la expresión génica fueron evaluadas por la prueba t de Student.

## **2.2 Ensayos de proliferación celular.**

Se ha propuesto que al observar alteraciones en el fenotipo celular es una manera de demostrar que una repetida posee una actividad regulatoria *in vivo* (Feschotte, 2008). Por lo anterior se realizaron conteos celulares por triplicado de las distintas clonas obtenidas así como de la línea celular K562 silvestre durante cuatro días consecutivos con el objetivo de definir su tasa de proliferación. Los resultados fueron graficados a lo largo del tiempo para demostrar cambios en la tasa de proliferación celular en las distintas líneas celulares.

## **3. Análisis genómico del efecto de secuencias repetidas tipo Alu en ARN largos no-codificantes intergénicos (lincRNAs).**

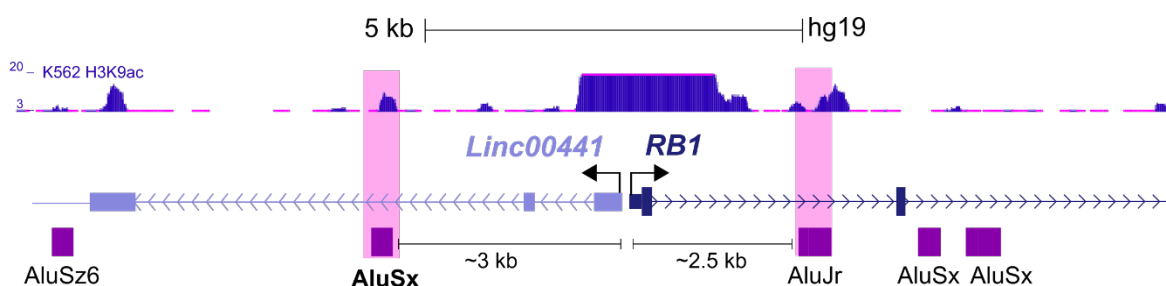
Las secuencias repetidas Alu han sido propuestas como elementos tipo *enhancer* o potenciadores de las secuencias cercanas a ellas (Su *et al.*, 2014; Zhang *et al.*,

2019; Haws *et al.*, 2022; Modzelewski *et al.*, 2022). En particular se ha demostrado que *enhancers* intragénicos pueden atenuar la transcripción de su gen hospedero (Cinghu *et al.*, 2017). Basados en las evidencias obtenidas y en búsqueda de un efecto generalizado acerca de la función de las secuencias repetidas Alu, realizamos un análisis bioinformático a nivel del genoma completo. Por consiguiente, se evaluó la frecuencia de la presencia de secuencias repetidas Alu contenidas en ARNs largos no-codificantes, así como su correlación con los niveles de expresión en diferentes tejidos de los elementos que las contienen.

## VIII. RESULTADOS

### 1. Análisis de la distribución de las secuencias repetidas tipo Alu en el locus del gen *RB1*

Con el fin de evaluar la función regulatoria de las secuencias repetidas tipo Alu en la región que contiene al promotor del gen *RB1*, se analizó la distribución de los diferentes tipos de secuencias repetidas en una región de 10 Kb que comprende al TSS del gen *RB1* (Figura 11).



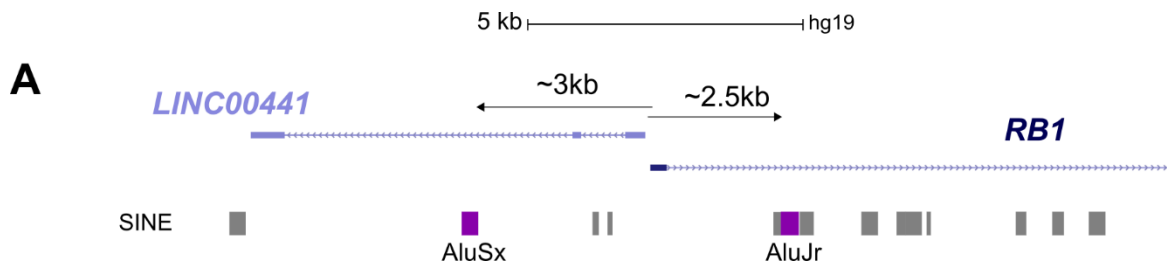
**Figura 11. Esquema del locus *LINC00441-RB1*.** En recuadros morados la distribución de secuencias repetidas tipo Alu y la distancia entre las dos secuencias repetidas más cercanas, la AluSx y la AluJr (enmarcadas en rosa claro).

En un inicio nos enfocamos en las dos secuencias repetidas tipo Alu más cercanas al promotor del gen *RB1*, una secuencia de la subfamilia AluSx y la otra de la subfamilia AluJr. La secuencia repetida AluSx se encuentra 3 Kb río arriba del TSS del gen *RB1* que a su vez, se encuentra dentro del segundo intrón del *LINC00441*; un lincRNA divergente de *RB1*, que se ha demostrado que afecta la transcripción de *RB1* en células cancerosas (Tang *et al.*, 2017; Zhou *et al.*, 2018). La secuencia repetida AluJr está localizada en el primer intrón del gen *RB1*, aproximadamente 2.5 Kb río abajo del TSS. Posteriormente, tomando en cuenta la anotación de estas repetidas, se determinó en estas mismas regiones un enriquecimiento de la marca de eucromatina en la histona H3K9ac; modificación asociada con *enhancers* activos y que se encontró recientemente enriquecida en elementos Alu con una expresión célula-específica (Zhang *et al.*, 2019) (Figura 12A). También fueron asignados a esta región genómica, los diferentes estados de la cromatina definidos por el

proyecto ENCODE (Proyecto ENCODE Consorcio, 2012). Ambas secuencias repetidas coinciden con regiones de apertura cromatínica, actividad transcripcional débil y secuencias tipo *enhancer*. En este sentido, la secuencia repetida AluSx solo se asocia a un estado de la cromatina, mientras que la secuencia repetida AluJr se superpone parcialmente con una región definida también para promotores (Figura 12B).

Además, se obtuvo la distribución de los distintos TFBs que contenían ambas secuencias repetidas. Se observa, que como bien se ha descrito anteriormente (Polak & Domany, 2006; Bouttier *et al.*, 2016), estas regiones coinciden con una alta frecuencia de diversos TFBs. Muchos de ellos están involucrados en la regulación de procesos en el desarrollo y metabolismo (Gfi1, PITX2, SREB, HoxB4, etc.). Cabe resaltar que esta característica es una de las razones para considerarlas candidatas potenciales a presentar una actividad regulatoria.

Por último, uno de los aspectos determinantes para la elección entre estas dos secuencias repetidas fue la localización de la secuencia repetida AluJr. Esta secuencia se localiza inmediatamente al lado de otra secuencia repetida SINE tipo FLAM, lo cual representaba un desafío técnico para la manipulación de esta secuencia en los distintos ensayos para estudiar su función. Por lo que descartamos trabajar con esta última y decidimos centrar nuestro estudio en la actividad reguladora del elemento AluSx río arriba del TSS del gen *RB1* (Figura 12C).







C

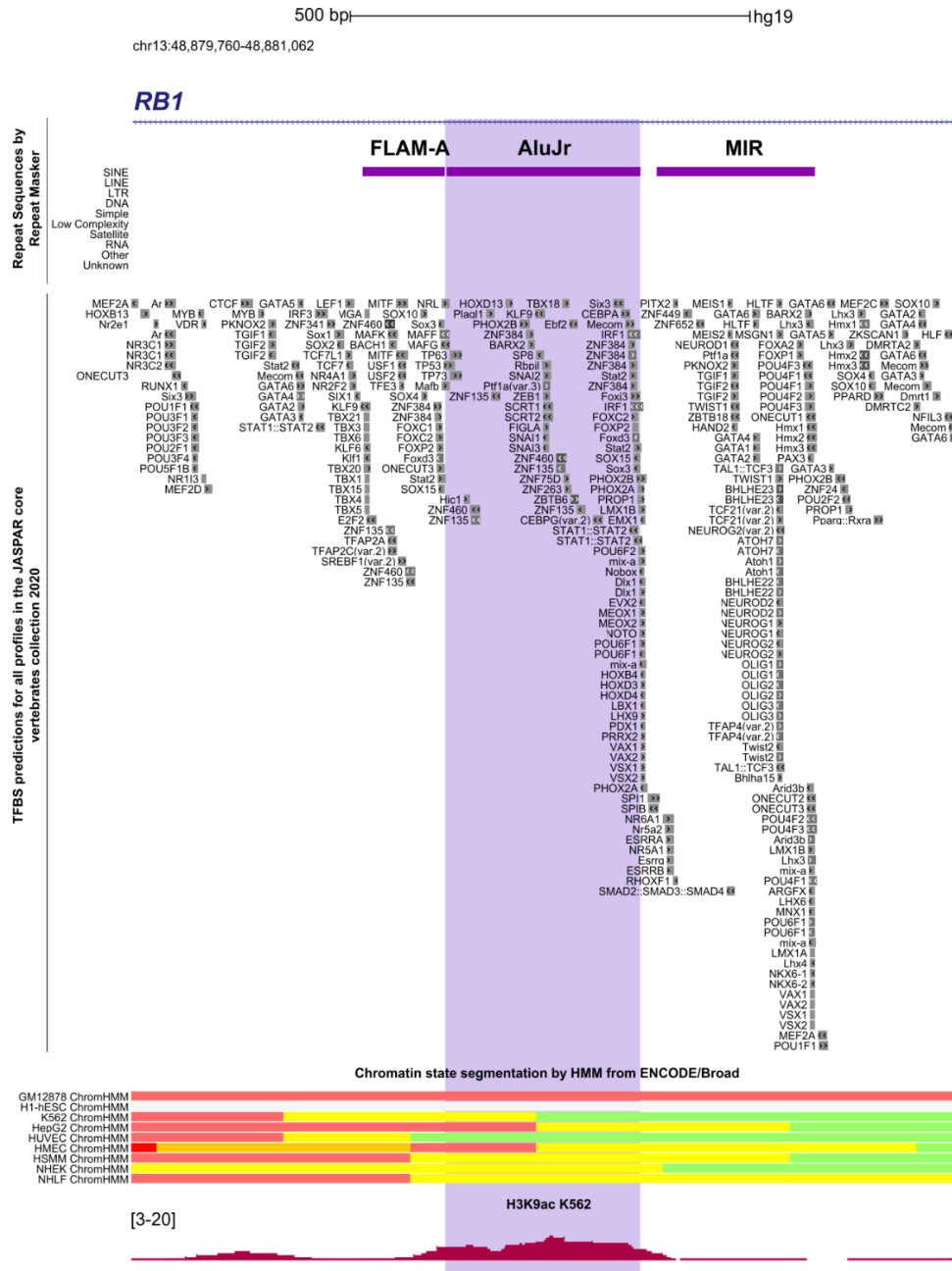


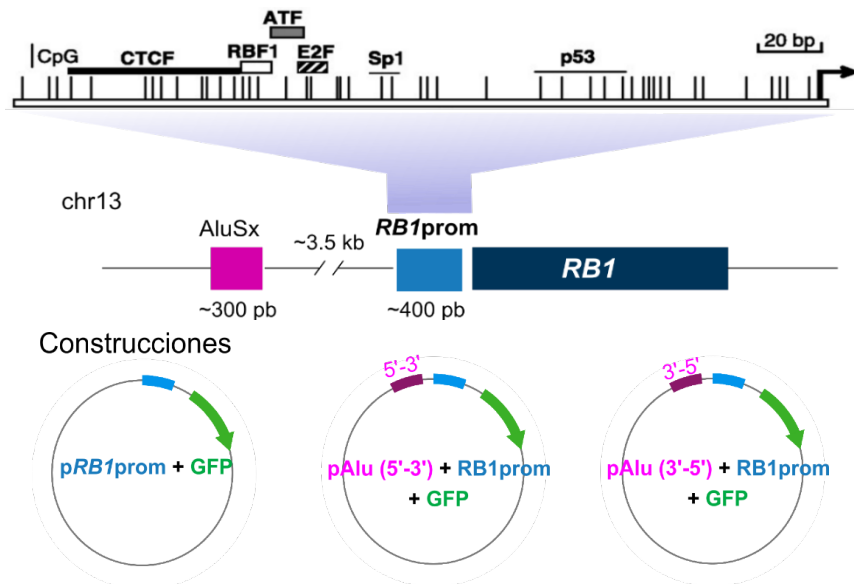
Figura 12. Los elementos Alu que se localizan en el locus *LINC00441-RB1* están enriquecidos en sitios de unión a FTs y muestran enriquecimiento en la marca en la histona H3K9ac. **A.** Distribución de los elementos genómicos que se localizan en el locus *LINC00441-RB1*. Datos derivados del UCSC Genome Browser. En la parte inferior el *track* con la distribución de secuencias repetidas tipo SINE. **B.** Paisaje genómico del elemento AluSx río arriba del promotor del gen *RB1* y localizado en el intrón 2 del locus *LINC00441*. **C.** Paisaje genómico del elemento AluJr localizado en el intrón 1 del gen *RB1*. En B y C, se muestran 4 *tracks* de datos genómicos anotados en distintas bases de datos para las dos regiones donde se localizan las secuencias repetidas Alu de nuestro interés. De arriba hacia abajo; el primero es la distribución de todas las familias de repetidas anotadas en el Repeat Masker, el segundo *track* contiene la distribución de TFBS, el tercer *track* los estados de cromatina definidos por el proyecto ENCODE y el cuarto *track* muestra datos de ChIPseq de la presencia de la marca de histona H3K9ac para cada una de las regiones que contiene a la secuencia repetida Alu.

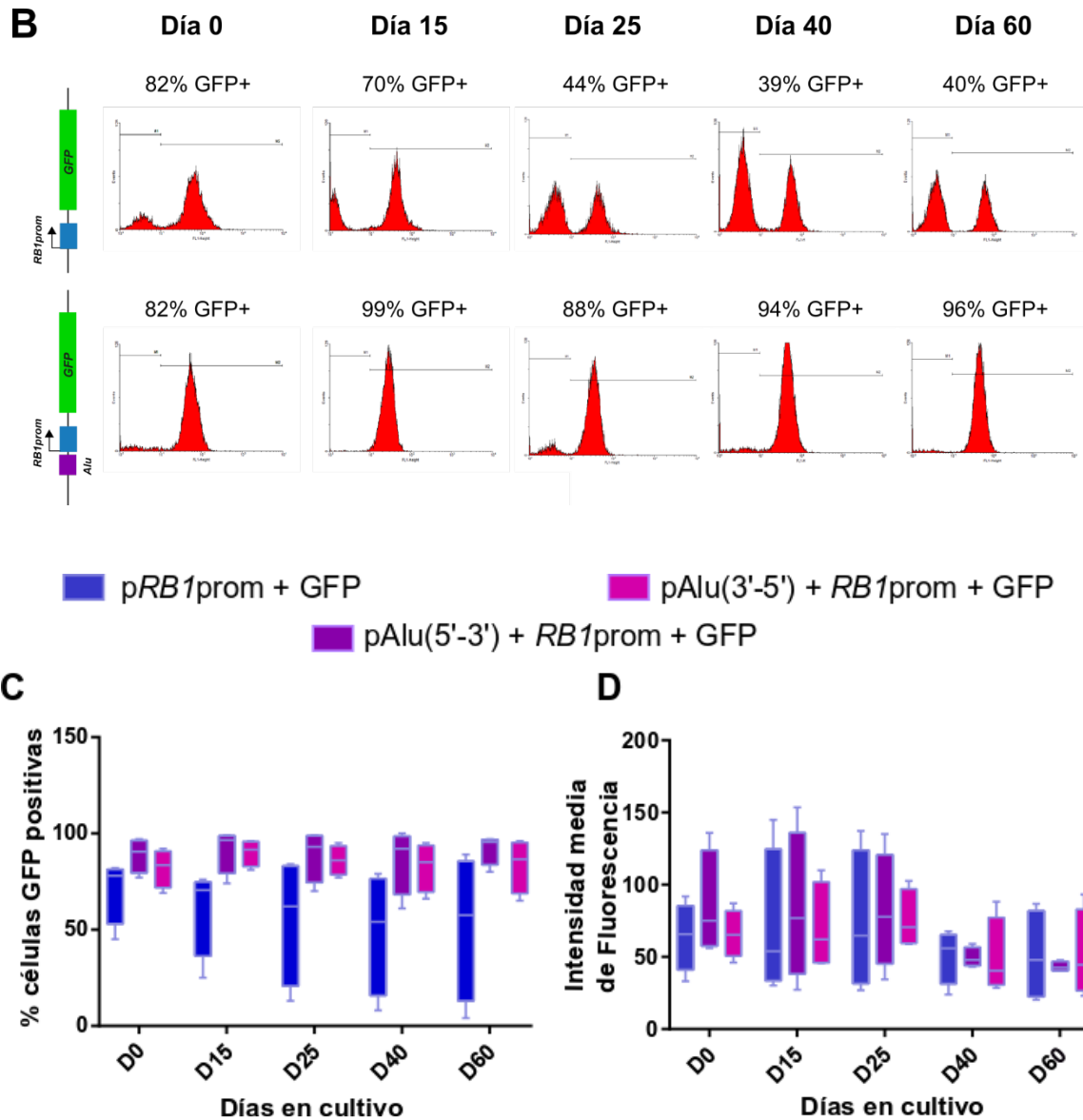
## 2. La secuencia repetida Alu se comporta como un elemento *enhancer* y protege contra el silenciamiento epigenético en construcciones reporteras

Para evaluar la naturaleza bioquímica de la secuencia repetida tipo AluSx seleccionada, se realizaron construcciones donde esta repetida fue clonada río arriba del promotor del gen *RB1* en ambas direcciones, en presencia de la *GFP* como gen reportero. Se obtuvieron tres construcciones las cuales fueron transfectadas y seleccionadas en la línea celular K562 (Figura 13A). Cada *pool* fue evaluado por FACS a diferentes tiempos durante 60 días a partir de cultivos celulares continuos.

De manera particular se muestran los perfiles de FACS de *pools* con las construcciones p*RB1*+*GFP* vs pAlu+*RB1*+*GFP* después de 0, 15, 30, 45 y 60 días de cultivo celular continuo en uno de los ensayos de transfección (Figura 13B). De este tipo de gráficas obtuvimos el porcentaje de células GFP(+) así como la intensidad media de fluorescencia para cada construcción. Se muestran a continuación los gráficos de los datos obtenidos de 4 ensayos independientes. El análisis de datos se realizó con Graphpad Prisma Software 7.0. Las diferencias estadísticamente significativas en los valores de la media de intensidad de fluorescencia entre las construcciones que contenían Alu y las que no, se calcularon utilizando una prueba de Mann-Whitney de dos colas (valor de  $p < 0,05$ ).

**A**





**Figura 13. Evaluación por FACS de los pools obtenidos para las diferentes construcciones. A.** Esquema de los elementos utilizados para las tres construcciones utilizadas para evaluar la naturaleza bioquímica de la secuencia repetida AluSx. Representación del fragmento utilizado como promotor de *RB1* donde se muestran en rectángulos y líneas horizontales, los distintos sitios de unión a sus FTs y en líneas verticales se representan los dinucleótidos CpGs presentes en esta región (De La Rosa-Velázquez *et al.*, 2007). En las construcciones se señala en azul el *RB1*prom, en rosa la secuencia repetida AluSx y en verde la GFP. **B.** Histogramas representativos de FACS obtenidos para ambas construcciones, pRB1+GFP vs pAlu+RB1+GFP en un ensayo de transfección. **C.** Gráfica del porcentaje de células GFP(+) **D.** Gráfica de la intensidad media de fluorescencia. Las gráficas C y D representan los datos obtenidos en cuatro ensayos independientes realizados durante 60 días (D60), n= 4.

Los resultados mostraron que en un inicio (D0), más del 50% de las células transfectadas con la construcción pRB1prom+GFP son GFP(+) (Figura 13B y C).

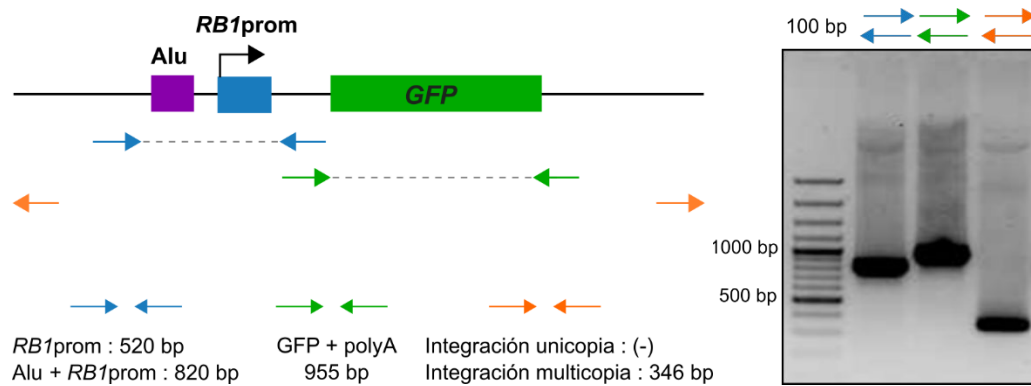
Esto es consistente con la naturaleza promotora de esta secuencia la cual ha sido evaluada y definida por ensayos reporteros, como la región mínima que contiene a los TFBs indispensables para la activación de la transcripción del gen *RB1* (Hong *et al.*, 1989; De La Rosa-Velázquez *et al.*, 2007) (Figura 13A). De manera inesperada, la media de células GFP(+) de las construcciones que contienen a la secuencia repetida Alu (pAlu(5'-3')+*RB1*prom+GFP y pAlu(3'-5')+*RB1*prom+GFP) fue mayor (89% y 82%, pAlu+*RB1*prom) en comparación a 71% de p*RB1*prom de GFP(+) así como la intensidad media de fluorescencia siendo también mayor (Figura 13C y D). Estos datos sugieren que la secuencia repetida Alu se comporta como un elemento *enhancer*.

Nuestro grupo y otros hemos caracterizado el silenciamiento epigenético progresivo de transgenes integrados de manera estable a lo largo del tiempo en cultivos celulares (De La Rosa-Velázquez *et al.*, 2007; Dávalos-Salas *et al.*, 2011). Dado que la adquisición del silenciamiento epigenético puede ser un proceso dependiente del tiempo, entonces el elemento repetido tipo Alu aún podría promover dicho silenciamiento epigenético. Por lo tanto, mantuvimos los cultivos celulares con las construcciones integradas de manera estable durante 60 días y cuantificamos el número de células GFP(+). Descubrimos que, en presencia de la secuencia repetida Alu, el transgén se protege contra el silenciamiento epigenético en contraste con los transgenes sin la secuencia repetida, las cuales exhibían el silenciamiento esperado con el paso del tiempo (Figura 13C y D). Por ejemplo, en el día 60, el 92% de las células Alu+*RB1*prom+GFP en ambas orientaciones fueron GFP(+), por el contrario, solo el 52% de las células *RB1*prom+GFP eran GFP(+). Cabe resaltar que observamos un aumento en el número de células GFP(+) para las construcciones en ambas orientaciones de la secuencia repetida Alu, un efecto bien caracterizado para elementos *enhancer*.

Como se puede ver en las gráficas, existe una gran dispersión de los datos lo cual nos indica una variabilidad entre las diferentes transfecciones que podría estar reflejando la falta de control (por el tipo de ensayo) de las siguientes variables: la eficiencia de transfección entre un ensayo y otro, el sitio de inserción en el genoma

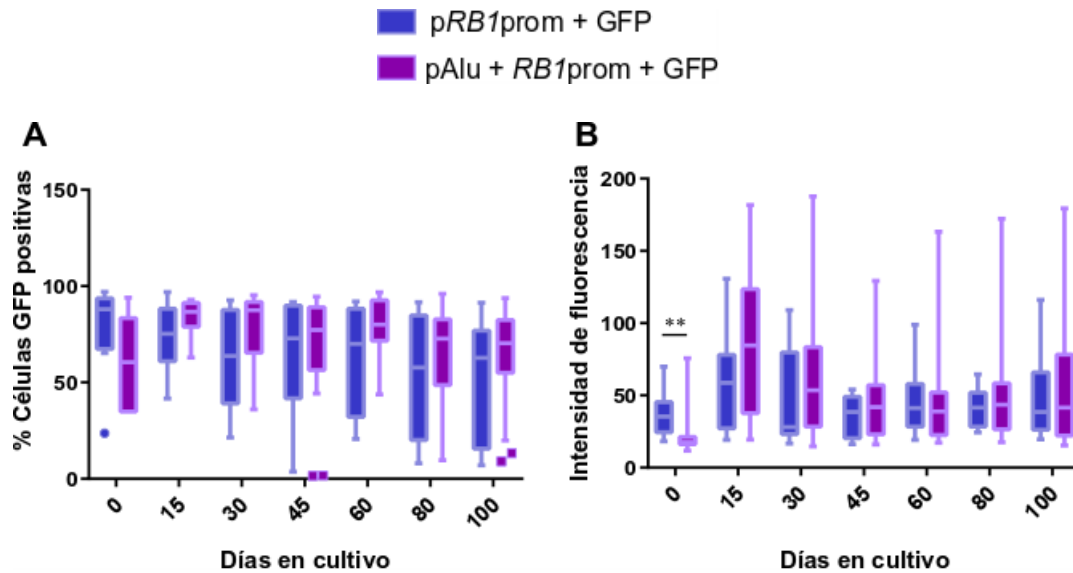
de cada una de las construcciones y/o la integridad del transgén. Por lo tanto, tratando de minimizar estas variables, repetimos el mismo ensayo, pero aislando clones independientes.

Después de aislar en agar en medio selectivo después de cuatro semanas, obtuvimos 30 clones independientes para cada construcción. De las cuales se extrajo ADN genómico para ser evaluadas por PCR y así corroborar la integridad del transgén (Figura 14).



**Figura 14. Estrategia experimental para evaluar por PCRs la integridad de los transgenes en clones independientes.** Se indican con flechas en azul, verde y naranja los tres pares de oligonucleótidos utilizados para amplificar cada una de las regiones que comprenden los transgenes en las distintas construcciones y sus tamaños esperados en pb. Así como un gel de agarosa al 1% que muestra el tamaño molecular de cada uno de los fragmentos de ADN de amplificación esperados en caso de estar íntegra la construcción para cada transgén.

Finalmente aislamos 14 clones para cada grupo, las cuales fueron evaluadas por FACS durante un periodo de 100 días de cultivo celular continuo. Como resultado, la tendencia observada previamente en *pools* también fue evidente en construcciones integradas de forma estable. Cabe mencionar que nuestros datos aún muestran una gran dispersión probablemente debido a la integración en diferentes entornos de cromatina (Figura 15).

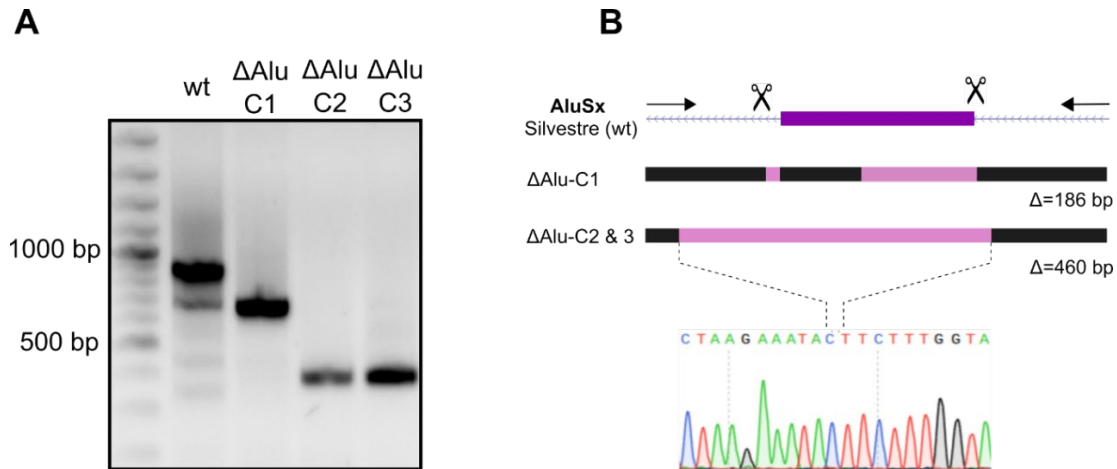


**Figura 15. Evaluación de las clonas independientes obtenidas para ambas construcciones. A.** Gráfica del porcentaje de células K562 GFP(+) **B.** Gráfica de la intensidad de la fluorescencia. Ensayo realizado durante 100 días de cultivo celular continuo (D100), n= 14.

En resumen, con los resultados obtenidos en los diferentes ensayos donde se utilizaron plásmidos integrados establemente, se observa que la secuencia Alu posee una actividad de acuerdo al modelo binario de acción de los *enhancers*. Este modelo plantea la forma en que un *enhancer* incrementa la probabilidad de transcripción (“on/off”) y no, la cantidad de transcrito presente en una célula al no observarse diferencias significativas en la intensidad de fluorescencia.

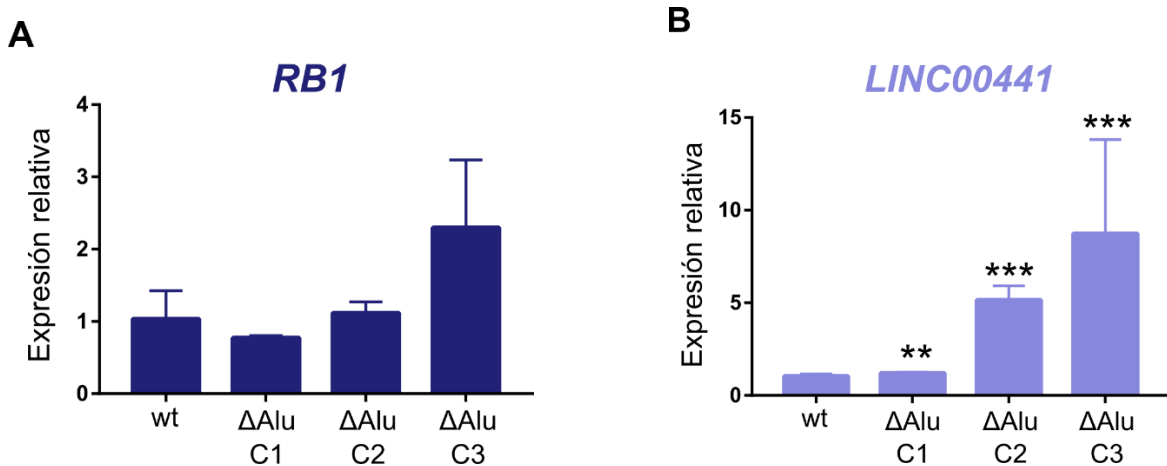
### 3. La eliminación de la secuencia de la repetida Alu promueve la regulación positiva del *LINC00441* ocasionando alteraciones en la proliferación celular

Para evaluar la función *in situ* del elemento repetido Alu, diseñamos sgRNAs para generar una eliminación específica de la secuencia con el sistema CRISPR-Cas9 en las células K562 (Figura 10). Aislamos tres clonas homocigotas que presentan dos eliminaciones diferentes (Figura 16).



**Figura 16. Validación por PCR y secuenciación de las regiones adyacentes a la eliminación esperada de la secuencia repetida AluSx para las clonas obtenidas por CRISPR-Cas9. A.** Gel de agarosa al 1% que muestra los productos de PCR para determinar el tamaño de los alelos con eliminaciones de la secuencia repetida en las tres clonas aisladas. El alelo silvestre amplificó las 844 pb esperadas mientras que los tamaños obtenidos de las mutantes fueron más pequeños (~650 pb y ~350 pb). **B.** Esquema de la región amplificada para evaluar la eliminación de la AluSx en las clonas obtenidas. Se representan en lila, las eliminaciones definidas por secuenciación de estos amplicones, presentes en las tres clonas obtenidas ( $\Delta$ Alu-C1, la  $\Delta$ Alu-C2 y  $\Delta$ Alu-C3).

La línea  $\Delta$ Alu-C1 presentó una eliminación de 186 pb que abarca parte de la región 5' de la Alu. En contraste, las líneas  $\Delta$ Alu-C2 y  $\Delta$ Alu-C3 tienen una eliminación de 460 pb que elimina completamente la secuencia repetida Alu, así como 150 pb adicionales río arriba de la secuencia repetida (Figura 16B). Para posteriormente evaluar el efecto de la eliminación en la transcripción de *LINC00441*, la región huésped de la secuencia repetida Alu eliminada, así como del gen *RB1* en las tres líneas celulares mutantes (Figura 17). Estas gráficas representan los datos obtenidos de tres réplicas biológicas y tres réplicas técnicas.



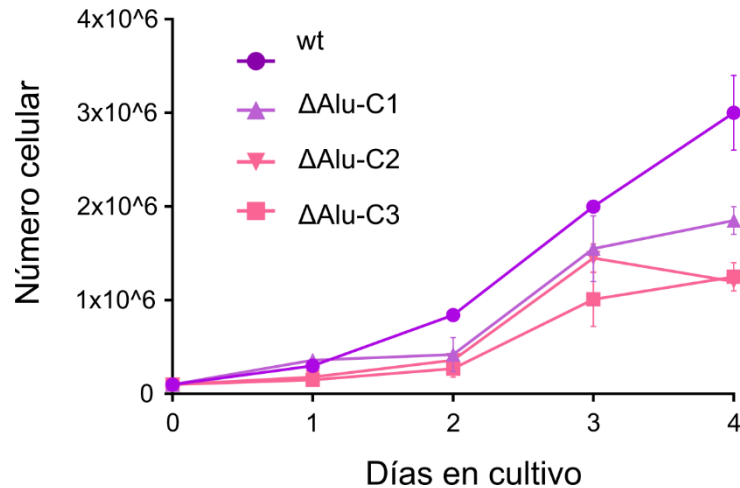


**Figura 17. Análisis de la expresión del gen *RB1* y el *LINC00441* por RTqPCR. A.** Amplificación del exón 27 del gen *RB1*. **B.** Amplificación del exón 1 del *LINC00441*. Ambos ensayos se realizaron para evaluar el efecto de la eliminación de la secuencia repetida Alu en las tres clonas obtenidas por CRISPR-Cas9 versus la línea K562 silvestre, n=3.

Inesperadamente, encontramos que para todas las clonas, la transcripción del *LINC00441* se encuentra regulado al alza. Este aumento en la transcripción fue mayor en las clonas que carecen del elemento Alu por completo. Por el contrario, al evaluar la transcripción del gen *RB1*; solo en la clona  $\Delta$ Alu-C3 observamos una tendencia hacia el aumento en la transcripción del gen *RB1*, sin embargo, este cambio no fue estadísticamente significativo.

Estos datos sugieren que el elemento Alu podría estar actuando como un *enhancer* intragénico localizado en el *LINC00441* atenuando su expresión. Aunado a lo que se ha descrito acerca de los *enhancers* y el efecto de potenciar la transcripción de su gen blanco, recientemente fue descrito el concepto de *enhancer* atenuador. Este tipo de elementos fuera de incrementar la tasa de transcripción de su gen blanco, funcionan más como reguladores de su transcripción, efecto que se ha descrito recientemente para *enhancers* intragénicos de genes codificantes en humanos que presentan bajos niveles de expresión (Cinghu *et al.*, 2017).

Con el objetivo de evaluar si las clonas que carecen de la secuencia repetida AluSx presentan alguna alteración en su fenotipo, se realizaron conteos celulares de un cultivo cada 24 hrs durante 4 días. Con estos datos se elaboró una curva de proliferación por triplicado en las tres clonas mutantes tomando como control una línea silvestre (Figura 18).



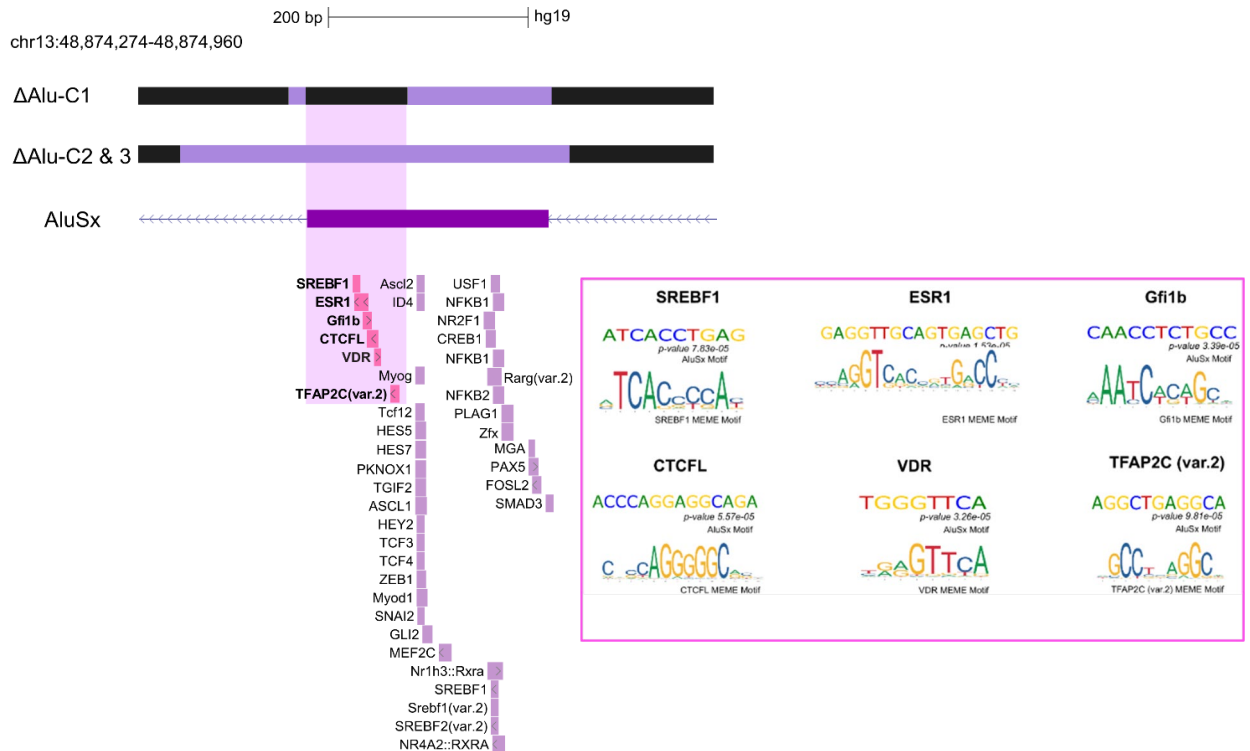
**Figura 18. Ensayo de proliferación.** Gráfica del conteo celular de los cultivos de las tres clonas obtenidas por CRISPR-Cas9 ( $\Delta$ Alu-C1, la  $\Delta$ Alu-C2 y  $\Delta$ Alu-C3) versus la línea celular K562 silvestre durante 4 días consecutivos.

Como se muestra en la gráfica de la Figura 18, en todas las clonas mutantes se observó una disminución en la proliferación celular siendo esta reducción más pronunciada en las dos clonas con la eliminación total de la secuencia repetida Alu. Por los datos anteriores, podemos sugerir que la secuencia repetida Alu posee una función regulatoria sobre elementos involucrados en la proliferación celular ya sea de manera directa a través del *LINC00441* o indirecta a través de las funciones de pRb, actuando como un inhibidor parcial de la progresión del ciclo celular.

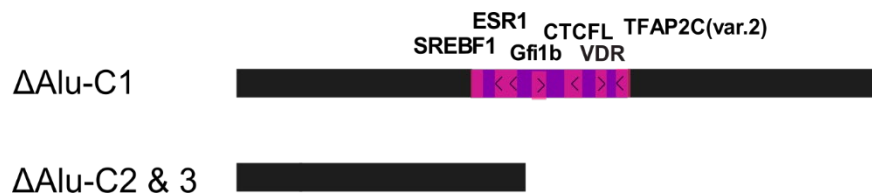
De manera interesante, observamos diferencias tanto en los ensayos de expresión como de proliferación no sólo al comparar las líneas mutantes versus la silvestre, sino también entre las mismas líneas con eliminaciones diferenciales en la AluSx. Por lo que consideramos relevante evaluar si estas diferencias se relacionaban con la distribución de los distintos tipos de TFBs que contiene esta secuencia repetida. Por lo tanto, basándonos en el análisis previo de los motivos de unión a TFs en AluSx y la secuenciación de las regiones eliminadas entre las mutantes CRISPR-Cas9, pudimos determinar que solo 6 TFBs son los que conserva la clona  $\Delta$ Alu-C1 a diferencia de las otras dos donde fue eliminada en su totalidad la secuencia repetida Alu (Figura 19). Estudios más detallados permitirán evaluar la contribución de estos FTs para el entendimiento de la forma en que los TFBs presentes en las

secuencias repetidas Alu se relacionan con la actividad tipo *enhancer* que se propone para la secuencia repetida AluSx en esta región.

**A**



**B**



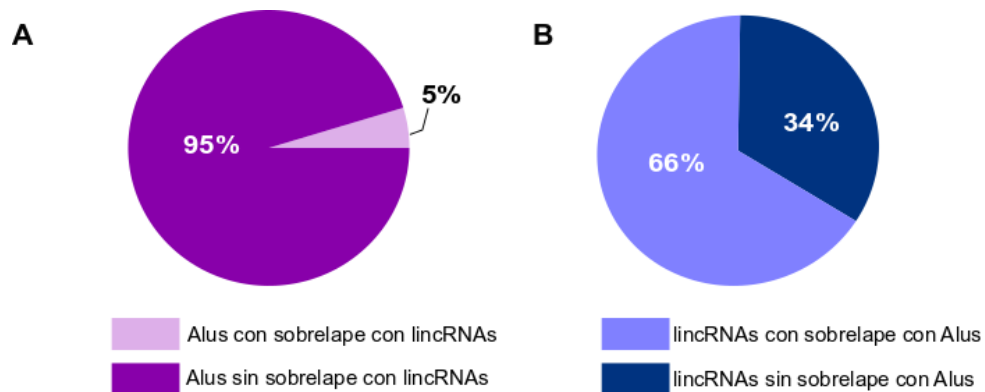
**Figura 19. Análisis de motivos de unión a FTs y deleciones en las distintas mutantes. A.** Representación esquemática de la región (chr13:48,874,274-48,874,960) que contiene el elemento AluSx río arriba del promotor *RB1*. Los rectángulos de color lila representan la región eliminada en cada mutante (arriba). Los sitios de unión del motivo para AluSx (chr13:48,874,474-48,874,760) por MEME (valor  $p < 0,0001$ ) se muestran como recuadros en violeta claro. La región resaltada corresponde a la secuencia no eliminada en el alelo mutante  $\Delta$ Alu-C1 que presenta una eliminación parcial de AluSx y contiene seis sitios de unión a la transcripción indicados en rosa (centro). A su derecha, se muestra un recuadro con los motivos del logo de cada TFBS con su valor  $p$  correspondiente. **B.** Esquema que resume la secuencia que posee cada mutante. En negro se representa la secuencia que rodea al elemento tipo AluSx, en morado se enmarca la secuencia que corresponde a la AluSx donde se señalan en recuadros rosas los TFBS remanentes en la clona  $\Delta$ Alu-C1.

#### **4. Las secuencias repetidas Alu intragénicas correlacionan con niveles más bajos de transcripción de lincRNAs en diferentes tejidos humanos.**

Se ha propuesto que las secuencias repetidas Alu en humanos actúan como elementos tipo *enhancers* (Su *et al.*, 2014). Esta actividad reguladora parece estar relacionada con un proceso de domesticación el cual propone, que los eucariontes han desarrollado mecanismos que ayudan a convertir su inserción y sus efectos en la estabilidad genómica, en algo benéfico para el genoma huésped. Por ejemplo, la expresión y movilidad que caracteriza a estos elementos repetidos, se encuentra controlada por procesos de silenciamiento epigenético en diferentes vías como los pequeños ARNs, proteínas con dominio KRAB, metilación del ADN, modificaciones de histonas e inhibición del procesamiento y modificaciones al ARN (Ecco *et al.*, 2017; Ernst *et al.*, 2017). Características de silenciamiento que han sido descritas en familias jóvenes de secuencias repetidas Alu y se propone, podrían estar funcionando como un mecanismo indirecto de silenciamiento en regiones cercanas a estas secuencias repetidas. Y a su vez, las familias de secuencias repetidas Alu más antiguas, han evolucionado hacia secuencias con características propias de reguladores de la transcripción como son los *enhancers* (Almeida *et al.*, 2022). Adicional a estas evidencias, se ha demostrado que un *enhancer* intragénico puede regular negativamente la transcripción de sus genes huésped mediante un mecanismo de interferencia de su promotor blanco (Cinghu *et al.*, 2017).

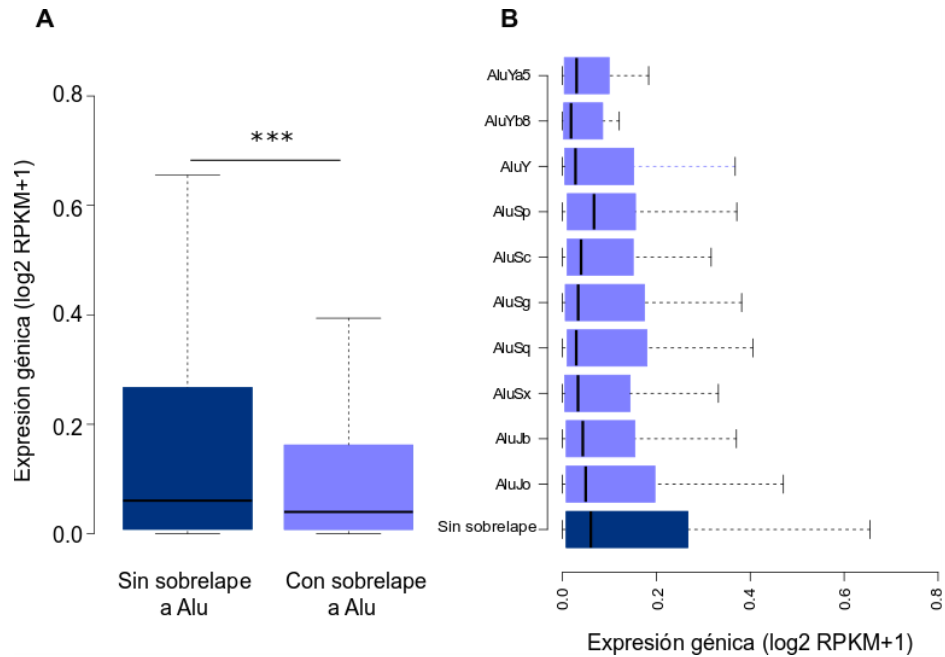
Dado que la secuencia repetida Alu inmersa en el *LINC00441* se comporta como un elemento potenciador y la eliminación del elemento da como resultado un aumento en la expresión del *LINC00441*, nos preguntamos si los elementos tipo Alu localizados en los lincRNAs podrían correlacionarse con los niveles de expresión de estos elementos. Para averiguarlo, descargamos el conjunto de datos de GENTEX que contiene la anotación más confiable de los lincRNAs en el genoma humano, así como sus niveles de transcripción en más de 50 tejidos recolectados de donantes humanos provenientes de la base de datos GENTEX. Posteriormente, clasificamos los lincRNAs en dos grupos: 1) aquellos que contienen por lo menos una secuencia repetida tipo Alu traslapando su secuencia o 2) en aquellos donde no hay traslape

con secuencias repetidas Alu. En primer lugar, encontramos de manera significativa que el 66% de todos los lincRNAs contienen al menos una secuencia repetida Alu intragénica, frecuencia que no se observa en sentido opuesto donde solo una baja proporción (5%) del conjunto de secuencias repetidas tipo Alu traslapa con lincRNAs (Figura 20).



**Figura 20. Análisis bioinformático de la distribución en el genoma completo de lincRNAs y secuencias repetidas tipo Alu.** **A.** Gráfico circular del porcentaje de secuencias Alu en el genoma humano que se superponen con lincRNA. **B.** Gráfico circular del porcentaje de lincRNA con al menos un elemento tipo Alu totalmente superpuesto.

Finalmente graficamos la distribución de los valores de expresión de los lincRNAs con y sin secuencias repetidas Alu intragénicas derivados del RNA-seq de 50 tejidos humanos distintos. De manera interesante, los lincRNAs con Alus intragénicas se expresan significativamente menos que los lincRNAs sin este tipo de repetidas (Figura 21A). Este mismo análisis lo realizamos tomando en cuenta el tipo de subfamilia de las secuencias repetidas tipo Alu que se traslapan con lincRNAs. A excepción de los lincRNAs que contienen elementos Alu de la subfamilia Sp, los lincRNAs con secuencias Alu intragénicas de cualquier otra familia exhibieron niveles más bajos de expresión en comparación con los lincRNAs que carecen de elementos Alu intragénicos (Figura 21B).



**Figura 21. Los lincRNA con secuencias Alu totalmente superpuestas muestran niveles de expresión reducidos. A.** Diagrama de caja de los niveles medios de expresión para lincRNAs que contienen secuencias repetidas Alu intragénicas y lincRNAs sin traslape a secuencias repetidas Alu. **B.** Diagrama de caja de los niveles medios de expresión para lincRNA que contienen secuencias repetidas Alu intragénicas y lincRNAs sin traslape a secuencias repetidas Alu; clasificados por subfamilias de Alu.

Estos resultados sugieren que las secuencias Alu intragénicas podrían estar promoviendo bajos niveles de expresión de sus lincRNAs hospederos proponiendo así una función más generalizada para este tipo de secuencias repetidas para lo cual se requiere de evidencias directas que lo demuestren.

## IX. DISCUSIÓN

En la actualidad, existen varios estudios recientes que revelan un vínculo entre los elementos repetidos tipo Alu y el control de la expresión génica (Hanke *et al.*, 1995; Mallona *et al.*, 2016; Chen *et al.*, 2018). Muchos de estos hallazgos provienen de extensos análisis computacionales apoyados en bases de datos epigenómicos y transcriptómicos de genomas completos (Goerner-Potvin & Bourque, 2018; Zhang *et al.*, 2019). A pesar de todos estos análisis, evidencias experimentales directas de secuencias repetidas Alu y su contribución en la red de regulación génica ha sido poco explorada.

En este trabajo, obtuvimos evidencias que sugieren que una secuencia repetida de tipo Alu puede comportarse como un *enhancer* y proteger contra el silenciamiento epigenético en ensayos reporteros en un contexto cromatínico. Además, demostramos que la eliminación parcial o completa de la secuencia repetida AluSx en su locus endógeno, aumenta la transcripción de su lincRNA huésped, afectando también la proliferación celular. Estos hallazgos contribuyen a una mejor comprensión del potencial regulatorio que estas secuencias específicas de primates tienen sobre la regulación de la expresión génica en el genoma humano.

En lo particular, en este trabajo nos enfocamos en la secuencia repetida AluSx ubicada río arriba del promotor del gen *RB1*, localizada a su vez en el intrón 2 del *LINC00441*. Mediante ensayos de gen reportero, observamos que la presencia de esta secuencia repetida Alu aumenta tanto el número de células positivas para *GFP* como la intensidad de fluorescencia media, lo que sugiere que actúa como un elemento *enhancer* o potenciador. En apoyo a lo anterior, un estudio reciente analizó la ocupación de nucleosomas en todo el genoma, la modificación de histonas y las características de los motivos de unión a factores de transcripción en los elementos repetidos tipo Alu. En este estudio los autores concluyeron que los elementos repetidos Alu mostraron características de *enhancers* sugiriendo que este tipo de repetidas son el antecesor evolutivo de secuencias regulatorias en el genoma humano (Su *et al.*, 2014; Modzelewski *et al.*, 2022). Particularmente, el

efecto de la secuencia repetida AluSx fue más evidente en el aumento de células positivas para *GFP* en lugar de la intensidad media de fluorescencia. Lo anterior sugiere que la repetida AluSx incrementa la frecuencia de eventos de inicio de la transcripción en las células según el modelo binario, donde plantea la forma en que los *enhancers* aumentan los niveles transcripcionales de los promotores asociados (Blackwood & Kadonaga, 1998; García-González & Recillas-Targa, 2014). Cabe mencionar que, si bien nuestros datos provenientes de ensayos utilizando plásmidos reporteros para probar el potencial regulador de la secuencia repetida AluSx y que nuestros resultados sugieren un efecto en la transcripción y protección contra el silenciamiento epigenético; no podemos descartar que la secuencia de AluSx, fuera de su contexto genómico, pueda actuar como un espaciador inespecífico de ADN que afecta el silenciamiento epigenético del promotor del gen *RB1*. Ensayos utilizando como control una secuencia arbitraria con un tamaño similar a la secuencia repetida Alu (~300 pb), nos podrían servir como control para descartar la opción de que esta secuencia repetida pueda actuar como un espaciador inespecífico en otras regiones genómicas.

Aunque demostramos el efecto regulador en *cis* de una secuencia repetida Alu en ensayos reporteros, estos experimentos están naturalmente limitados por el hecho de que las secuencias a evaluar se investigan independientemente de su contexto cromosómico nativo. Por lo tanto, elegimos determinar su papel endógeno en la regulación de su lincRNA huésped y genes vecinos; así como sus implicaciones en la proliferación celular, a través de una combinación de eliminaciones realizadas mediante el sistema CRISPR-Cas9 y análisis de expresión génica. Sorprendentemente, contrario a la función clásica de un elemento potenciador, observamos en mutantes que carecían de la secuencia repetida AluSx, un aumento significativo en la expresión de *LINC00441*, pero no del gen *RB1*. Con base a lo anterior, concluimos que la secuencia repetida AluSx atenúa la expresión de su secuencia huésped, el *LINC00441*, como se ha reportado que actúan de manera específica *enhancers* intragénicos de genes codificantes de proteínas en humanos (Cinghu *et al.*, 2017). En particular, este efecto atenuador se ha descrito



concretamente para genes que contienen *enhancers* intragénicos con niveles de expresión bajos a moderados en células madre embrionarias, donde la función dominante del *enhancer* es probablemente la de un atenuador en genes con este perfil de expresión (Cinghu *et al.*, 2017). Por este motivo, realizamos un análisis de todo el genoma, observando que los lincRNAs con secuencias repetidas Alu intragénicas se transcriben significativamente menos que los lincRNA en ausencia de secuencias repetidas Alu intragénicas. Esta correlación podría deberse a que las secuencias repetidas Alu intragénicas podrían estar relacionadas con la reducción en la transcripción de los lincRNAs, que en general están presentes con una abundancia ~10 veces menor que los mRNA en una población celular (Cabili *et al.*, 2011). Lo anterior sustenta el considerar a las secuencias repetidas tipo Alu, como buenos candidatos para ser considerados como elementos en el genoma con potencial regulatorio, responsables específicamente del ajuste fino de la transcripción de sus genes huésped.

Un reporte reciente plantea otro aspecto importante acerca de la actividad tipo *enhancer* que se relaciona con la naturaleza de los ETs. Ellos proponen que las inserciones ancestrales de ETs podrían ser el origen de módulos de secuencias reguladoras en *cis* a los genomas, lo cual coincide con la localización preferencial observada de múltiples factores de transcripción en este tipo de secuencias (Sundaram *et al.*, 2017). Por lo tanto, realizamos un análisis de los motivos de unión a factores de transcripción presentes en la secuencia repetida AluSx. Encontramos que muchos de los sitios identificados son sitios de unión para factores nucleares, ligandos de hormonas y otros FTs relacionados con procesos de diferenciación y desarrollo. Lo anterior es consistente con la presencia de los motivos que se han descrito como enriquecidos en secuencias repetidas tipo Alu (Polak & Domany, 2006). De los motivos de unión hallados en la secuencia repetida AluSx, encontramos que seis sitios de unión a FTs permanecen intactos en la mutante que presenta una eliminación parcial de la secuencia repetida. Esta diferencia en la presencia de sitios de unión a FTs probablemente esté relacionada con el efecto más sutil observado en la expresión y proliferación, en comparación con las

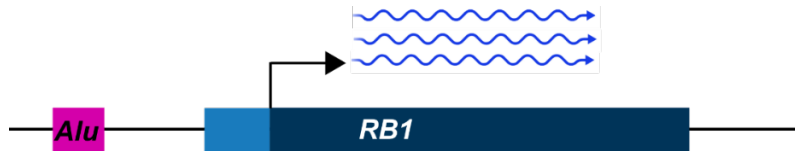
mutantes que tienen una eliminación completa de la secuencia repetida AluSx. Entre estos sitios, identificamos motivos de unión de factores relacionados con vías metabólicas (SREBF1; biosíntesis de esteroides), elementos de respuesta hormonal (ESR1 y VDR) y procesos de diferenciación (Gfi1b en linaje hematopoyético y TFAP2C en morfogénesis temprana). De manera notoria, también identificamos un sitio de unión parálogo de CTCF (CTCFL) que puede estar relacionado con el efecto protector contra el silenciamiento epigenético observado para la secuencia repetida AluSx. Sin embargo, actualmente se desconoce cómo estos FTs están involucrados en la actividad potenciadora de secuencias repetidas tipo Alu. Por lo tanto, identificar la contribución diferencial de cada sitio de unión a FTs relacionado con la actividad atenuadora de estas secuencias, podría ayudar a comprender mejor esta nueva función de los elementos repetidos y poder distinguirlos de otras clases de secuencias repetidas.

Derivado de las recientes observaciones donde identificaron el enriquecimiento de modificaciones postraduccionales de histonas asociadas a elementos *enhancer* así como el reclutamiento célula específica de RNA Pol II/III a las distintas subfamilias de secuencias repetidas Alu (Su *et al.*, 2014; Zhang *et al.*, 2019), resulta de gran interés investigar el papel de AluSx en otros contextos celulares. Estos ensayos permitirían conocer el papel de factores de transcripción específicos para diferentes tipos celulares en relación con la actividad reguladora de una secuencia repetida tipo Alu.

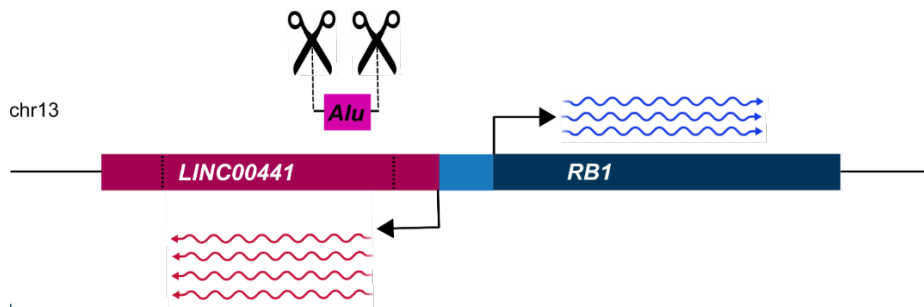
Se requieren más estudios para abordar a escala genómica, el impacto que tienen las secuencias repetidas Alu sobre la expresión de los lincRNAs. Sin embargo, nuestros hallazgos sugieren que este tipo de secuencias repetidas forman parte de los mecanismos involucrados en la regulación transcripcional de genes que requieren de un ajuste más preciso, lo que destaca la necesidad de realizar más ensayos funcionales para un mejor entendimiento de la forma en que operan estos elementos enigmáticos en el genoma humano.

## X. CONCLUSIONES

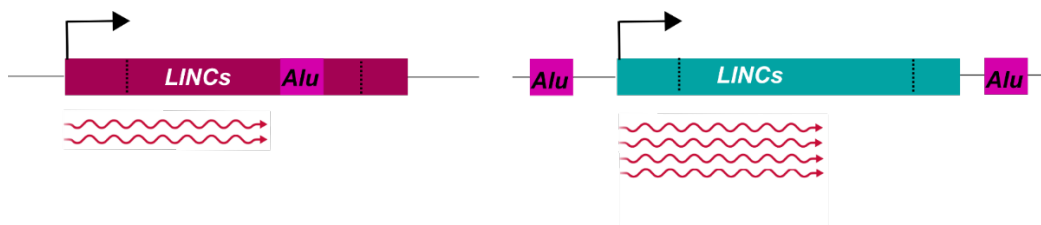
1. La secuencia repetida Alu río arriba del promotor del gen *RB1* tiene actividad de *enhancer* en plásmidos integrados establemente.



2. La remoción de la secuencia repetida Alu resultó en la sobreexpresión del *LINC00441* sugiriendo funcionar como un elemento *enhancer* con propiedades de atenuador.



3. La ausencia de la secuencia repetida Alu en el *LINC00441* afecta la proliferación celular.
4. A nivel genómico, al evaluar la presencia de secuencias repetidas tipo Alu dentro de lincRNAs, se observó una tendencia donde el traslape de estas secuencias repetidas correlaciona con una baja expresión de los lincRNAs en diferentes contextos celulares.



## XI. PERSPECTIVAS

- Evaluar otras *repetidas* de la misma región para conocer si tiene el mismo efecto sobre el *LINC00441* y el gen *RB1*.
- Identificar el mecanismo por el cual las secuencias repetidas participan en la atenuación de la transcripción de sus blancos.
- En base a los análisis bioinformáticos corroborar la misma función de *enhancer* en otros lincRNAs que contengan secuencias repetidas Alu intragénicas.
- Evaluar si la regulación como *enhancer* de secuencias repetidas Alu intragénicas es el mecanismo por el cual las repetidas actúan sobre lincRNAs en procesos patológicos como el cáncer.

## XII. BIBLIOGRAFÍA

- Ade, C., Roy-Engel, A. M., & Deininger, P. L. (2013). Alu elements: an intrinsic source of human genome instability. *Current Opinion in Virology*, 3(6), 639–645.
- Almeida, M. V., Vernaz, G., Putman, A. L. K., & Miska, E. A. (2022). Taming transposable elements in vertebrates: from epigenetic silencing to domestication. *Trends in Genetics : TIG*, 38(6), 529–553.
- Amir, R. E., Van Den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics*, 23(2), 185–188.
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K., Schwarzfischer, L., Glatz, D., Raithel, J., Lilje, B., Rapin, N., ... Sandelin, A. (2014). An atlas of active enhancers across human cell types and tissues. *Nature*, 507(7493), 455–461.
- Arzate-Mejía, R. G., Josué Cerecedo-Castillo, A., Guerrero, G., Furlan-Magaril, M., & Recillas-Targa, F. (2020). In situ dissection of domain boundaries affect genome topology and gene transcription in *Drosophila*. *Nature Communications*, 11(1).
- Arzate-Mejía, R. G., Recillas-Targa, F., & Corces, V. G. (2018). Developing in 3D: the role of CTCF in cell differentiation. *Development (Cambridge, England)*, 145(6).
- Bandara, T. A. M. K., Otsuka, K., Matsubara, S., Shiraishi, A., Satake, H., & Kimura, A. P. (2021). A dual enhancer-silencer element, DES-K16, in mouse spermatocyte-derived GC-2spd(ts) cells. *Biochemical and Biophysical Research Communications*, 534, 1007–1012.
- Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., & Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4), 823–837.
- Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell*, 136(2), 215–233.
- Batzer, M. A., & Deininger, P. L. (2002). Alu repeats and human genomic diversity. *Nature Reviews. Genetics*, 3(5), 370–379.
- Bernstein, E., & Allis, C. D. (2005). RNA meets chromatin. *Genes & Development*, 19(14), 1635–1655.
- Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E., & Allis, C. D. (2006). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Molecular and Cellular Biology*,

26(7), 2560–2569.

- Blackwood, E. M., & Kadonaga, J. T. (1998). Going the distance: A current view of enhancer action. *Science*, *281*(5373), 60–63.
- Bose, D. A., Donahue, G., Reinberg, D., Shiekhhattar, R., Bonasio, R., & Berger, S. L. (2017). RNA Binding to CBP Stimulates Histone Acetylation and Transcription. *Cell*, *168*(1–2), 135–149.e22.
- Bouttier, M., Laperriere, D., Memari, B., Mangiapane, J., Fiore, A., Mitchell, E., Verway, M., Behr, M. A., Sladek, R., Barreiro, L. B., Mader, S., & White, J. H. (2016). Alu repeats as transcriptional regulatory platforms in macrophage responses to *M. tuberculosis* infection. *Nucleic Acids Research*, *44*(22), 10571.
- Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S., & Crawford, G. E. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell*, *132*(2), 311–322.
- Bulger, M., & Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell*, *144*(3), 327–339.
- Cabili, M., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., & Rinn, J. L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development*, *25*(18), 1915–1927.
- Chen, K., Wang, Y., & Sun, J. (2018). A statistical analysis on transcriptome sequences: The enrichment of Alu-element is associated with subcellular location. *Biochemical and Biophysical Research Communications*, *499*(3), 397–402.
- Cinghu, S., Yang, P., Kosak, J. P., Conway, A. E., Kumar, D., Oldfield, A. J., Adelman, K., & Jothi, R. (2017). Intragenic Enhancers Attenuate Host Gene Expression. *Molecular Cell*, *68*(1), 104–117.e6.
- Cook, P. R. (2010). A model for all genomes: the role of transcription factories. *Journal of Molecular Biology*, *395*(1), 1–10.
- Cooper, S. J., Trinklein, N. D., Anton, E. D., Nguyen, L., & Myers, R. M. (2006). Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Research*, *16*(1), 1–10.
- Dávalos-Salas, M., Furlan-Magaril, M., González-Buendía, E., Valdes-Quezada, C., Ayala-Ortega, E., & Recillas-Targa, F. (2011). Gain of DNA methylation is enhanced in the absence of CTCF at the human retinoblastoma gene promoter. *BMC Cancer*, *11*, 232.
- De La Rosa-Velázquez, I. A., Rincón-Arano, H., Benítez-Bribiesca, L., & Recillas-Targa, F. (2007). Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Research*, *67*(6), 2577–2585.
- de Lara, J. C. F., Arzate-Mejía, R. G., & Recillas-Targa, F. (2019). Enhancer RNAs:

Insights Into Their Biological Role. *Epigenetics Insights*, 12, 2516865719846093.

- De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B. K., Muller, H., Ragoussis, J., Wei, C. L., & Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biology*, 8(5), e1000384.
- Deininger, P. (2011). Alu elements: know the SINEs. *Genome Biology*, 12(12), 236.
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J. B., Lipovich, L., Gonzalez, J. M., ... Guigó, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Research*, 22(9), 1775–1789.
- Dieci, G., Preti, M., & Montanini, B. (2009). Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics*, 94(2), 83–88.
- Dinger, M. E., Amara, P. P., Mercer, T. R., Pang, K. C., Bruce, S. J., Gardiner, B. B., Askarian-Amiri, M. E., Ru, K., Soldà, G., Simons, C., Sunkin, S. M., Crowe, M. L., Grimmond, S. M., Perkins, A. C., & Mattick, J. S. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Research*, 18(9), 1433–1445.
- Dinger, M. E., Amaral, P. P., Mercer, T. R., & Mattick, J. S. (2009). Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications. *Briefings in Functional Genomics & Proteomics*, 8(6), 407–423.
- Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R. F., ... Gingeras, T. R. (2012). Landscape of transcription in human cells. *Nature*, 489(7414), 101–108.
- Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Fietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B. K., Pauli, F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., ... Lochovsky, L. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414), 57–74.
- Ecco, G., Imbeault, M., & Trono, D. (2017). KRAB zinc finger proteins. *Development (Cambridge, England)*, 144(15), 2719–2729.
- Elkon, R., & Agami, R. (2017). Characterization of noncoding regulatory DNA in the human genome. *Nature Biotechnology*, 35(8), 732–746.
- Ernst, C., Odom, D. T., & Kutter, C. (2017). The emergence of piRNAs against transposon invasion to preserve mammalian genome integrity. *Nature Communications*, 8(1), 1411.
- Feschotte, C. (2008). Transposable elements and the evolution of regulatory

- networks. *Nature Reviews. Genetics*, 9(5), 397–405.
- Fornes, O., Castro-Mondragon, J. A., Khan, A., Van Der Lee, R., Zhang, X., Richmond, P. A., Modi, B. P., Correard, S., Gheorghe, M., Baranašić, D., Santana-Garcia, W., Tan, G., Chèneby, J., Ballester, B., Parcy, F., Sandelin, A., Lenhard, B., Wasserman, W. W., & Mathelier, A. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*, 48(D1), D87–D92.
- Friedli, M., & Trono, D. (2015). The developmental control of transposable elements and the evolution of higher species. *Annual Review of Cell and Developmental Biology*, 31, 429–451.
- García-González, E., Escamilla-Del-Arenal, M., Arzate-Mejía, R., & Recillas-Targa, F. (2016). Chromatin remodeling effects on enhancer activity. *Cellular and Molecular Life Sciences*, 73(15), 2897–2910.
- García-González, E., & Recillas-Targa, F. (2014). A regulatory element affects the activity and chromatin structure of the chicken  $\alpha$ -globin 3' enhancer. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1839(11), 1233–1241.
- Gavazzo, P., Vassalli, M., Costa, D., & Pagano, A. (2013). Novel ncRNAs transcribed by Pol III and elucidation of their functional relevance by biophysical approaches. *Frontiers in Cellular Neuroscience*, 7, 203.
- Gershenzon, N. I., Trifonov, E. N., & Ioshikhes, I. P. (2006). The features of *Drosophila* core promoters revealed by statistical analysis. *BMC Genomics*, 7, 161.
- Giacinti, C., & Giordano, A. (2006). RB and cell cycle progression. *Oncogene*, 25(38), 5220–5227.
- Giordano, J., Ge, Y., Gelfand, Y., Abrusán, G., Benson, G., & Warburton, P. E. (2007). Evolutionary History of Mammalian Transposons Determined by Genome-Wide Defragmentation. *PLoS Computational Biology*, 3(7), 1321–1334.
- Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R., & Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Research*, 17(6), 877–885.
- Gisselbrecht, S. S., Palagi, A., Kurland, J. V., Rogers, J. M., Ozadam, H., Zhan, Y., Dekker, J., & Bulyk, M. L. (2020). Transcriptional Silencers in *Drosophila* Serve a Dual Role as Transcriptional Enhancers in Alternate Cellular Contexts. *Molecular Cell*, 77(2), 324–337.e8.
- Goerner-Potvin, P., & Bourque, G. (2018). Computational tools to unmask transposable elements. *Nature Reviews. Genetics*, 19(11), 688–704.
- Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., Huarte, M., Zuk, O., Carey, B. W., Cassady, J. P., Cabili, M. N., Jaenisch, R., Mikkelsen, T.



- S., Jacks, T., Hacohen, N., Bernstein, B. E., Kellis, M., Regev, A., Rinn, J. L., & Lander, E. S. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, *458*(7235), 223–227.
- Guttman, M., Donaghey, J., Carey, B. W., Garber, M., Grenier, J. K., Munson, G., Young, G., Lucas, A. B., Ach, R., Bruhn, L., Yang, X., Amit, I., Meissner, A., Regev, A., Rinn, J. L., Root, D. E., & Lander, E. S. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, *477*(7364), 295–300.
- Guttman, M., & Rinn, J. L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature*, *482*(7385), 339–346.
- Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J. B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., Joly, J. S., & Concordet, J. P. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology*, *17*(1), 148.
- Hanke, J. H., Hambor, J. E., & Kavathas, P. (1995). Repetitive Alu elements form a cruciform structure that regulates the function of the human CD8 alpha T cell-specific enhancer. *Journal of Molecular Biology*, *246*(1), 63–73.
- Hare, E. E., Peterson, B. K., Iyer, V. N., Meier, R., & Eisen, M. B. (2008). Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genetics*, *4*(6), e1000106.
- Häsler, J., & Strub, K. (2006). Alu elements as regulators of gene expression. *Nucleic Acids Research*, *34*(19), 5491–5497.
- Hatakeyama, M., & Weinberg, R. A. (1995). The role of RB in cell cycle control. *Progress in Cell Cycle Research*, *1*, 9–19.
- Haws, S. A., Simandi, Z., Barnett, R. J., & Phillips-Cremins, J. E. (2022). 3D genome, on repeat: Higher-order folding principles of the heterochromatinized repetitive genome. *Cell*, *185*(15), 2690–2707.
- He, H. H., Meyer, C. A., Shin, H., Bailey, S. T., Wei, G., Wang, Q., Zhang, Y., Xu, K., Ni, M., Lupien, M., Mieczkowski, P., Lieb, J. D., Zhao, K., Brown, M., & Liu, X. S. (2010). Nucleosome dynamics define transcriptional enhancers. *Nature Genetics* *2010* *42*:4, *42*(4), 343–347.
- Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E., ... Ren, B. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, *459*(7243), 108–112.
- Hewitt, S. M., Fraizer, G. C., & Saunders, G. F. (1995). Transcriptional silencer of the Wilms' tumor gene WT1 contains an Alu repeat. *The Journal of Biological*

*Chemistry*, 270(30), 17908–17912.

Hnisz, D., Abraham, B. J., Lee, T. I., Lau, A., Saint-André, V., Sigova, A. A., Hoke, H. A., & Young, R. A. (2013). Super-enhancers in the control of cell identity and disease. *Cell*, 155(4), 934.

Hong, F. D., Huang, H. J. S., To, H., Young, L. J. S., Oro, A., Bookstein, R., Lee, E. Y. H. P., & Lee, W. H. (1989). Structure of the human retinoblastoma gene. *Proceedings of the National Academy of Sciences of the United States of America*, 86(14), 5502–5506.

Hoyt, S. J., Storer, J. M., Hartley, G. A., Grady, P. G. S., Gershman, A., de Lima, L. G., Limouse, C., Halabian, R., Wojenski, L., Rodriguez, M., Altemose, N., Rhie, A., Core, L. J., Gerton, J. L., Makalowski, W., Olson, D., Rosen, J., Smit, A. F. A., Straight, A. F., ... O'Neill, R. J. (2022). From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. *Science (New York, N.Y.)*, 376(6588), eabk3112.

Huang, Q. (2015). Genetic study of complex diseases in the post-GWAS era. *Journal of Genetics and Genomics = Yi Chuan Xue Bao*, 42(3), 87–98.

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M. J., Kenzelmann-Broz, D., Khalil, A. M., Zuk, O., Amit, I., Rabani, M., Attardi, L. D., Regev, A., Lander, E. S., Jacks, T., & Rinn, J. L. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell*, 142(3), 409–419.

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3(3), 318–356.

Jangam, D., Feschotte, C., & Betrán, E. (2017). Transposable Element Domestication As an Adaptation to Evolutionary Conflicts. *Trends in Genetics : TIG*, 33(11), 817–831.

Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. *Trends in Genetics : TIG*, 16(9), 418–420.

Karolchik, D., Hinricks, A. S., Furey, T. S., Roskin, K. M., Sugnet, C. W., Haussler, D., & Kent, W. J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Research*, 32(Database issue), D493–D496.

Kim, T. K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., Harmin, D. A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P. F., Kreiman, G., & Greenberg, M. E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, 465(7295), 182–187.

Kitajima, S., Li, F., & Takahashi, C. (2020). Tumor Milieu Controlled by RB Tumor Suppressor. *International Journal of Molecular Sciences*, 21(7), 2450.

Knudsen, E. S., & Knudsen, K. E. (2008). Tailoring to RB: tumour suppressor status and therapeutic response. *Nature Reviews. Cancer*, 8(9), 714–724.

- Ko, M. S. H., Nakauchi, H., & Takahashi, N. (1990). The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *The EMBO Journal*, 9(9), 2835–2842.
- Kolesky, S. E., Ouhammouch, M., & Peter Geiduschek, E. (2002). The mechanism of transcriptional activation by the topologically DNA-linked sliding clamp of bacteriophage T4. *Journal of Molecular Biology*, 321(5), 767–784.
- Kolovos, P., Knoch, T. A., Grosveld, F. G., Cook, P. R., & Papantonis, A. (2012). Enhancers and silencers: an integrated and simple model for their function. *Epigenetics & Chromatin*, 5(1), 1.
- Lanctôt, C., Cheutin, T., Cremer, M., Cavalli, G., & Cremer, T. (2007). Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nature Reviews. Genetics*, 8(2), 104–115.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., ... Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860–921.
- Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A. Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., Oh, S., Kim, H. S., Glass, C. K., & Rosenfeld, M. G. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature*, 498(7455), 516–520.
- Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L. A., Lander, E. S., & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, 326(5950), 289–293.
- Ling, J. Q., Li, T., Hu, J. F., Vu, T. H., Chen, H. L., Qiu, X. W., Cherry, A. M., & Hoffman, A. R. (2006). CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science (New York, N.Y.)*, 312(5771), 269–272.
- Liu, W. man, & Schmid, C. W. (1993). Proposed roles for DNA methylation in Alu transcriptional repression and mutational inactivation. *Nucleic Acids Research*, 21(6), 1351–1359.
- Loewer, S., Cabili, M. N., Guttman, M., Loh, Y. H., Thomas, K., Park, I. H., Garber, M., Curran, M., Onder, T., Agarwal, S., Manos, P. D., Datta, S., Lander, E. S., Schlaeger, T. M., Daley, G. Q., & Rinn, J. L. (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nature Genetics*, 42(12), 1113–1117.
- Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J., & Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell*,

126(2), 403–413.

- Luo, Y., Lu, X., & Xie, H. (2014). Dynamic Alu methylation during normal development, aging, and tumorigenesis. *BioMed Research International*, 2014, 784706.
- Lupiáñez, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J. M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S. A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., ... Mundlos, S. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161(5), 1012–1025.
- Maison, C., Bailly, D., Peters, A. H. F. M., Quivy, J. P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T., & Almouzni, G. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nature Genetics*, 30(3), 329–334.
- Mallona, I., Jordà, M., & Peinado, M. A. (2016). A knowledgebase of the human Alu repetitive elements. *Journal of Biomedical Informatics*, 60, 77–83.
- Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., & Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature*, 445(7128), 666–670.
- McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America*, 36(6), 344–355.
- McPherson, J. D., Marra, M., Hillier, L. D., Waterston, R. H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E. R., Wilson, R. K., Fulton, R., Kucaba, T. A., Wagner-McPherson, C., Barbazuk, W. B., Gregory, S. G., Humphray, S. J., French, L., Evans, R. S., Bethel, G., ... Lehrach, H. (2001). A physical map of the human genome. *Nature*, 409(6822), 934–941.
- Mercer, T. R., & Mattick, J. S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nature Structural & Molecular Biology*, 20(3), 300–307.
- Miele, A., & Dekker, J. (2008). Long-range chromosomal interactions and gene regulation. *Molecular BioSystems*, 4(11), 1046–1058.
- Modzelewski, A. J., Gan Chong, J., Wang, T., & He, L. (2022). Mammalian genome innovation through transposon domestication. *Nature Cell Biology*, 24(9), 1332–1340.
- Nagano, T., & Fraser, P. (2011). No-nonsense functions for long noncoding RNAs. *Cell*, 145(2), 178–181.
- Ngan, C. Y., Wong, C. H., Tjong, H., Wang, W., Goldfeder, R. L., Choi, C., He, H., Gong, L., Lin, J., Urban, B., Chow, J., Li, M., Lim, J., Philip, V., Murray, S. A., Wang, H., & Wei, C. L. (2020). Chromatin interaction analyses elucidate the roles of PRC2-bound silencers in mouse development. *Nature Genetics*, 52(3),

264–272.

- Núñez-Martínez, H. N., & Recillas-Targa, F. (2022). Emerging Functions of lncRNA Loci beyond the Transcript Itself. *International Journal of Molecular Sciences*, 23(11), 6258.
- Ogbourne, S., & Antalis, T. M. (1998). Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *The Biochemical Journal*, 331 ( Pt 1(Pt 1), 1–14.
- Ohlsson, R., Renkawitz, R., & Lobanenkov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends in Genetics : TIG*, 17(9), 520–527.
- Pandey, R. R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-DiNardo, D., & Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Molecular Cell*, 32(2), 232–246.
- Pang, B., & Snyder, M. P. (2020). Systematic identification of silencers in human cells. *Nature Genetics*, 52(3), 254–263.
- Pang, K. C., Frith, M. C., & Mattick, J. S. (2006). Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in Genetics : TIG*, 22(1), 1–5.
- Paul, I. J., & Duerksen, J. D. (1975). Chromatin-associated RNA content of heterochromatin and euchromatin. *Molecular and Cellular Biochemistry*, 9(1), 9–16.
- Pefanis, E., Wang, J., Rothschild, G., Lim, J., Kazadi, D., Sun, J., Federation, A., Chao, J., Elliott, O., Liu, Z. P., Economides, A. N., Bradner, J. E., Rabadan, R., & Basu, U. (2015). RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell*, 161(4), 774–789.
- Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S., & Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature*, 379(6561), 131–137.
- Polak, P., & Domany, E. (2006). Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics*, 7, 133.
- Ponting, C. P., Oliver, P. L., & Reik, W. (2009). Evolution and functions of long noncoding RNAs. *Cell*, 136(4), 629–641.
- Qi, H., Liu, M., Emery, D. W., & Stamatoyannopoulos, G. (2015). Functional validation of a constitutive autonomous silencer element. *PLoS One*, 10(4), e0124588.
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6), 841–842.

- Ravasi, T., Suzuki, H., Pang, K. C., Katayama, S., Furuno, M., Okunishi, R., Fukuda, S., Ru, K., Frith, M. C., Gongora, M. M., Grimmond, S. M., Hume, D. A., Hayashizaki, Y., & Mattick, J. S. (2006). Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Research*, *16*(1), 11–19.
- Recillas-Targa, F. (2002). DNA methylation, chromatin boundaries, and mechanisms of genomic imprinting. *Archives of Medical Research*, *33*(5), 428–438.
- Recillas-Targa, F., De La Rosa-Velázquez, I. A., & Soto-Reyes, E. (2011). Insulation of tumor suppressor genes by the nuclear factor CTCF. *Biochemistry and Cell Biology = Biochimie et Biologie Cellulaire*, *89*(5), 479–488.
- Riethoven, J. J. M. (2010). Regulatory regions in DNA: promoters, enhancers, silencers, and insulators. *Methods in Molecular Biology (Clifton, N.J.)*, *674*, 33–42.
- Rinn, J. L., & Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annual Review of Biochemistry*, *81*, 145–166.
- Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., & Chang, H. Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, *129*(7), 1311–1323.
- Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., ... Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, *518*(7539), 317–329.
- Salditt-Georgieff, M., Harpold, M. M., Wilson, M. C., & Darnell, J. E. (1981). Large heterogeneous nuclear ribonucleic acid has three times as many 5' caps as polyadenylic acid segments, and most caps do not enter polyribosomes. *Molecular and Cellular Biology*, *1*(2), 179–187.
- Sambrook, J., & Russell, D. W. (2006). Purification of PCR products in preparation for cloning. *CSH Protocols*, *2006*(1), pdb.prot3825.
- Sanidas, I., Lee, H., Rumde, P. H., Boulay, G., Morris, R., Golczer, G., Stanzione, M., Hajizadeh, S., Zhong, J., Ryan, M. B., Corcoran, R. B., Drapkin, B. J., Rivera, M. N., Dyson, N. J., & Lawrence, M. S. (2022). Chromatin-bound RB targets promoters, enhancers, and CTCF-bound loci and is redistributed by cell-cycle progression. *Molecular Cell*, *82*(18), 3333-3349.e9.
- Schaukowitch, K., Joo, J. Y., Liu, X., Watts, J. K., Martinez, C., & Kim, T. K. (2014). Enhancer RNA facilitates NELF release from immediate early genes. *Molecular Cell*, *56*(1), 29–42.
- Schmidt, D., Schwalie, P. C., Wilson, M. D., Ballester, B., Goncalves, Â., Kutter, C.,

- Brown, G. D., Marshall, A., Flicek, P., & Odom, D. T. (2012). Waves of Retrotransposon Expansion Remodel Genome Organization and CTCF Binding in Multiple Mammalian Lineages. *Cell*, *148*(1–2), 335.
- Schmittgen, T. D., & Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nature Protocols*, *3*(6), 1101–1108.
- Segert, J. A., Gisselbrecht, S. S., & Bulyk, M. L. (2021). Transcriptional Silencers: Driving Gene Expression with the Brakes On. *Trends in Genetics: TIG*, *37*(6), 514–527.
- Seidl, C. I. M., Stricker, S. H., & Barlow, D. P. (2006). The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *The EMBO Journal*, *25*(15), 3565–3575.
- Sharan, C., Hamilton, N. M., Parl, A. K., Singh, P. K., & Chaudhuri, G. (1999). Identification and characterization of a transcriptional silencer upstream of the human BRCA2 gene. *Biochemical and Biophysical Research Communications*, *265*(2), 285–290.
- Simonis, M., Kooren, J., & de Laat, W. (2007). An evaluation of 3C-based methods to capture DNA interactions. *Nature Methods*, *4*(11), 895–901.
- Smale, S. T., & Kadonaga, J. T. (2003). The RNA polymerase II core promoter. *Annual Review of Biochemistry*, *72*, 449–479.
- Soufi, A., Donahue, G., & Zaret, K. S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell*, *151*(5), 994–1004.
- Soufi, A., Garcia, M. F., Jaroszewicz, A., Osman, N., Pellegrini, M., & Zaret, K. S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell*, *161*(3), 555–568.
- Spitz, F., & Furlong, E. E. M. (2012). Transcription factors: from enhancer binding to developmental control. *Nature Reviews. Genetics*, *13*(9), 613–626.
- Strober, W. (2001). Trypan blue exclusion test of cell viability. *Current Protocols in Immunology, Appendix 3*.
- Su, M., Han, D., Boyd-Kirkup, J., Yu, X., & Han, J. D. J. (2014). Evolution of Alu Elements toward Enhancers. *Cell Reports*, *7*(2), 376–385.
- Sun, Q., Hao, Q., & Prasanth, K. V. (2018). Nuclear Long Noncoding RNAs: Key Regulators of Gene Expression. *Trends in Genetics*, *34*(2), 142–157.
- Sun, X., Zhang, J., & Cao, C. (2022). CTCF and Its Partners: Shaper of 3D Genome during Development. *Genes*, *13*(8), 1383.
- Sundaram, V., Choudhary, M. N. K., Pehrsson, E., Xing, X., Fiore, C., Pandey, M., Maricque, B., Udawatta, M., Ngo, D., Chen, Y., Paguntalan, A., Ray, T., Hughes, A., Cohen, B. A., & Wang, T. (2017). Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. *Nature Communications*, *8*, 14550.

- Tang, J., Xie, Y. U., Xu, X., Yin, Y., Jiang, R., Deng, L., Tan, Z., Gangarapu, V., Tang, J., & Sun, B. (2017). Bidirectional transcription of Linc00441 and RB1 via H3K27 modification-dependent way promotes hepatocellular carcinoma. *Cell Death & Disease*, 8(3), e2675.
- Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., & De Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular Cell*, 10(6), 1453–1465.
- Tomilin, N. V. (1999). Control of genes by mammalian retroposons. *International Review of Cytology*, 186, 1–48.
- Trujillo, M. A., Sakagashira, M., & Eberhardt, N. L. (2006). The human growth hormone gene contains a silencer embedded within an Alu repeat in the 3'-flanking region. *Molecular Endocrinology (Baltimore, Md.)*, 20(10), 2559–2575.
- Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., & Luscombe, N. M. (2009). A census of human transcription factors: function, expression and evolution. *Nature Reviews. Genetics*, 10(4), 252–263.
- Walters, M. C., Fiering, S., Eidemiller, J., Magis, W., Groudine, M., & Martin, D. I. K. (1995). Enhancers increase the probability but not the level of gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 92(15), 7125–7129.
- Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M. U., Ohgi, K. A., Glass, C. K., Rosenfeld, M. G., & Fu, X. D. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature*, 474(7351), 390–397.
- Wang, K. C., Yang, Y. W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B. R., Protacio, A., Flynn, R. A., Gupta, R. A., Wysocka, J., Lei, M., Dekker, J., Helms, J. A., & Chang, H. Y. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*, 472(7341), 120–126.
- Weinberg, R. A., & Penman, S. (1968). Small molecular weight monodisperse nuclear RNA. *Journal of Molecular Biology*, 38(3), 289–304.
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., Klemm, A., Flicek, P., Manolio, T., Hindorff, L., & Parkinson, H. (2014). The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research*, 42(Database issue), D1001–D1006.
- Westbrook, T. F., Hu, G., Ang, X. L., Mulligan, P., Pavlova, N. N., Liang, A., Leng, Y., Maehr, R., Shi, Y., Harper, J. W., & Elledge, S. J. (2008). SCFbeta-TRCP controls oncogenic transformation and neural differentiation through REST degradation. *Nature*, 452(7185), 370–374.
- Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., & Young, R. A. (2013). Master transcription factors and



- mediator establish super-enhancers at key cell identity genes. *Cell*, 153(2), 307–319.
- Willingham, A. T., Orth, A. P., Batalov, S., Peters, E. C., Wen, B. G., Aza-Blanc, P., Hogenesch, J. B., & Schultz, P. G. (2005). A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science (New York, N.Y.)*, 309(5740), 1570–1573.
- Wu, J., Grindlay, G. J., Bushel, P., Mendelsohn, L., & Allan, M. (1990). Negative regulation of the human epsilon-globin gene by transcriptional interference: role of an Alu repetitive element. *Molecular and Cellular Biology*, 10(3), 1209–1216.
- Wutz, A., Rasmussen, T. P., & Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nature Genetics*, 30(2), 167–174.
- Yang, C., Bolotin, E., Jiang, T., Sladek, F. M., & Martinez, E. (2007). Prevalence of the initiator over the TATA box in human and yeast genes and identification of DNA motifs enriched in human TATA-less core promoters. *Gene*, 389(1), 52–65.
- Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R. D., & Felsenfeld, G. (2010). Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes & Development*, 24(22), 2543–2555.
- You, J. S., Kelly, T. K., De Carvalho, D. D., Taberlay, P. C., Liang, G., & Jones, P. A. (2011). OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. *Proceedings of the National Academy of Sciences of the United States of America*, 108(35), 14497–14502.
- Zaret, K. S. (2020). Pioneer Transcription Factors Initiating Gene Network Changes. *Annual Review of Genetics*, 54, 367–385.
- Zentner, G. E., Tesar, P. J., & Scacheri, P. C. (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Research*, 21(8), 1273–1283.
- Zhang, X. O., Gingeras, T. R., & Weng, Z. (2019). Genome-wide analysis of polymerase III–transcribed Alu elements suggests cell-type–specific enhancer function. *Genome Research*, 29(9), 1402–1414.
- Zhao, Y., Wang, L., Ren, S., Wang, L., Blackburn, P. R., McNulty, M. S., Gao, X., Qiao, M., Vessella, R. L., Kohli, M., Zhang, J., Karnes, R. J., Tindall, D. J., Kim, Y., MacLeod, R., Ekker, S. C., Kang, T., Sun, Y., & Huang, H. (2016). Activation of P-TEFb by Androgen Receptor-Regulated Enhancer RNAs in Castration-Resistant Prostate Cancer. *Cell Reports*, 15(3), 599–610.
- Zheng, P., Pennacchio, L. A., Le Goff, W., Rubin, E. M., & Smith, J. D. (2004). Identification of a novel enhancer of brain expression near the apoE gene

cluster by comparative genomics. *Biochimica et Biophysica Acta - Gene Structure and Expression*, 1676(1), 41–50.

Zhou, J., Shi, J., Fu, X., Mao, B., Wang, W., Li, W., Li, G., & Zhou, S. (2018). Linc00441 interacts with DNMT1 to regulate RB1 gene methylation and expression in gastric cancer. *Oncotarget*, 9(101), 37471–37479.

## XII. ANEXOS

### Publicaciones generadas durante el doctorado

**Pérez-Molina R.**, Arzate-Mejía RG, Ayala-Ortega E, Guerrero G, Meier K, Suaste-Olmos F and Recillas-Targa F (2020) An Intronic Alu Element Attenuates the Transcription of a Long Non-coding RNA in Human Cell Lines. *Front. Genet.* 11:928.

Rodríguez-Aguilera, J. R., Guerrero-Hernández, C., **Pérez-Molina, R.**, Cadena-del-Castillo, C. E., Pérez-Cabeza de Vaca, R., Guerrero-Celis, N., Domínguez-López, M., Murillo-de-Ozores, A. R., Arzate-Mejía, R., Recillas-Targa, F. and Chagoya de Sánchez, V. J. *Cell. Biochem.* 2018, 119: 401-413. Epigenetic Effects of an Adenosine Derivative in a Wistar Rat Model of Liver Cirrhosis.

Ayala-Ortega E, Arzate-Mejía R, **Pérez-Molina R**, González-Buendía E, Meier K, Guerrero G, Recillas-Targa F. *BMC Cancer.* 2016 Mar 16;16:226. Epigenetic silencing of miR-181c by DNA methylation in glioblastoma cell lines.

Martínez-Levy GA, Rocha L, Lubin FD, Alonso-Vanegas MA, Nani A, Buentello-García RM, **Pérez-Molina R**, Briones-Velasco M, Recillas-Targa F, Pérez-Molina A, San-Juan D, Cienfuegos J, Cruz-Fuentes CS. *Neuroscience.* 2016 Feb 9;314:12-21. Increased expression of BDNF transcript with exon VI in hippocampi of patients with pharmaco-resistant temporal lobe epilepsy.

González-Buendía E, Escamilla-Del-Arenal M, **Pérez-Molina R**, Tena JJ, Guerrero G, Suaste-Olmos F, Ayala-Ortega E, Gómez-Skarmeta JL, Recillas-Targa F. *Biochim Biophys Acta.* 2015 Aug;1849(8):955-65. A novel chromatin insulator regulates the chicken folate receptor gene from the influence of nearby constitutive heterochromatin and the  $\beta$ -globin locus.

González-Buendía E, **Pérez-Molina R**, Ayala-Ortega E, Guerrero G, Recillas-Targa F. *Methods Mol Biol.* 2014;1165:53-69. Experimental strategies to manipulate the cellular levels of the multifunctional factor CTCF.



# An Intronic *Alu* Element Attenuates the Transcription of a Long Non-coding RNA in Human Cell Lines

Rosario Pérez-Molina, Rodrigo G. Arzate-Mejía, Erandi Ayala-Ortega, Georgina Guerrero, Karin Meier, Fernando Suaste-Olmos<sup>†</sup> and Félix Recillas-Targa\*

Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Mexico City, Mexico

## OPEN ACCESS

### Edited by:

Katarzyna Oktaba,  
Unidad Irapuato (CINVESTAV),  
Mexico

### Reviewed by:

Lluís Montoliu,  
National Center for Biotechnology  
(CNB), Spain  
Dixie Mager,  
British Columbia Cancer Agency,  
Canada

### \*Correspondence:

Félix Recillas-Targa  
frecilla@ifc.unam.mx

### <sup>†</sup>Present address:

Fernando Suaste-Olmos,  
Departamento de Bioquímica y  
Biología Estructural, Instituto de  
Fisiología Celular, Universidad  
Nacional Autónoma de México,  
Mexico City, Mexico

### Specialty section:

This article was submitted to  
Epigenomics and Epigenetics,  
a section of the journal  
Frontiers in Genetics

**Received:** 06 June 2020

**Accepted:** 27 July 2020

**Published:** 31 August 2020

### Citation:

Pérez-Molina R, Arzate-Mejía RG,  
Ayala-Ortega E, Guerrero G, Meier K,  
Suaste-Olmos F and Recillas-Targa F  
(2020) An Intronic *Alu* Element  
Attenuates the Transcription of a  
Long Non-coding RNA in Human  
Cell Lines.  
Front. Genet. 11:928.  
doi: 10.3389/fgene.2020.00928

*Alu* elements are primate-specific repeats and represent the most abundant type of transposable elements (TE) in the human genome. Genome-wide analysis of the enrichment of histone post-translational modifications suggests that human *Alu* sequences could function as transcriptional enhancers; however, no functional experiments have evaluated the role of *Alu* sequences in the control of transcription *in situ*. The present study analyses the regulatory activity of a human *Alu* sequence from the *AluSx* family located in the second intron of the long intergenic non-coding RNA *Linc00441*, found in divergent orientation to the *RB1* gene. We observed that the *Alu* sequence acts as an enhancer element based on reporter gene assays while CRISPR-Cas9 deletions of the *Alu* sequence in K562 cells resulted in a marked transcriptional upregulation of *Linc00441* and a decrease in proliferation. Our results suggest that an intragenic *Alu* sequence with enhancer activity can act as a transcriptional attenuator of its host lincRNA.

**Keywords:** repeat sequences, *Alu* elements, intragenic enhancer, long intergenic non-coding RNAs, transcription attenuation, transposable elements

## INTRODUCTION

Repetitive elements constitute ~50% of the human genome (Lander et al., 2001; Bannert and Kurth, 2004). Different lines of research suggest that they can affect transcriptional regulation, however, most evidence supporting a direct regulatory activity of repetitive elements has been correlative at best (Mallona et al., 2016). The *Alu* subfamily of repetitive elements is a class of primate-specific Short Interspaced Nuclear Elements (SINEs) of ~300 base pair (bp) length that is present in more than 1 million copies in the human genome and hence constitutes the most abundant class of transposable element in humans (Lander et al., 2001). Nevertheless, their role in regulating gene expression and chromatin structure remains poorly characterized.

*Alu* elements are located preferentially in the proximity of gene-rich regions (Batzer and Deininger, 2002; Kim et al., 2016) and are rich in Transcription Factor Binding Sites (TFBS), which suggests a possible function as regulatory platforms for the transcriptional control of host or neighboring genes (Polak and Domany, 2006). In this regard, *Alu* elements have been suggested to nucleate epigenetic silencing *via* the acquisition of DNA methylation and histone post-translational modification H3K9me3, resulting in transcriptional silencing of neighboring genes (Graff et al., 1997; Baylin et al., 1998; Estecio et al., 2012). Recent reports have put

forward the idea that *Alu* elements have evolved toward enhancer elements in the human genome. This concerns particularly old *Alu* families like *AluSx*, *AluJo*, and *AluJb*, as they are enriched for the histone post-translational modifications H3K4me1, H3K27ac and have gained transcription factor binding motifs over time (Su et al., 2014). These enhancer-like characteristics are present in a tissue-specific manner and preferentially engage in long-range interactions with gene promoters and with *Alu* sequences thereof. Although these lines of evidence implicate that *Alu* elements, or at least a subset of them, can exert direct regulatory effects in gene transcription, direct characterization of the biochemical regulatory capacity of *Alus* (Chuong et al., 2017) and their function *in situ* need to be explored.

Here, we characterize the regulatory activity of an intronic *Alu* element of the *AluSx* family in the transcriptional gene regulation of the human *Linc000441-RB1* locus in the K562 erythroleukemic cell line. By employing plasmid-based reporter assays of stably transfected pools of cells and single clones coupled with Fluorescence-Activated Cell Sorting (FACS), we show that this *Alu* sequence behaves as an enhancer element protecting against epigenetic silencing for over 100 days of continuous cell culture. Remarkably, CRISPR-Cas9 deletion of the *Alu* element results in strong transcriptional upregulation of the long intergenic non-coding RNA (lincRNA) *Linc000441* with consequences in cell proliferation, suggesting that the *Alu* sequence behaves as a transcriptional *in situ* attenuator. Overall, our results reveal that a single *Alu* sequence can affect gene transcription and cell proliferation. Importantly, biochemical and *in situ* activities could differ and highlight the importance of analyzing both when characterizing repeat sequences. Furthermore, our results underscore the possibility that *Alu* elements have a more widespread role for transcriptional regulation of lincRNAs than previously anticipated.

## MATERIALS AND METHODS

### Plasmid Constructs

pRB1prom-GFP plasmid contains *RB1* promoter already characterized and cloned into pEGFP plasmid (pEGFP-1; Clontech, Palo Alto, CA; De La Rosa-Velazquez et al., 2007), which is the intergenic region between *Linc000441* and the human *RB1* gene (chr13:48,877,623-48,878,023; h19 version). The closest *Alu* repeat upstream of *RB1* gene promoter (*AluSx1* repeat chr13: 48,874,474-48,874,760) was amplified (Supplementary Table S1) from human lymphocyte genomic DNA and subcloned into the pRB1prom-GFP in two orientations to generate pAlu(5'-3')-RB1prom-GFP and pAlu(3'-5')-RB1prom-GFP plasmids. All the plasmids contain a neomycin-resistance cassette, which allows G418 selection of stably transfected cells. The integrity of all plasmid constructs was verified by DNA sequencing.

### Cell Culture

K562 human erythroleukemic cells were cultured in ISCOVE medium (Invitrogen). K562 cells (K562 ATCC®CCL-243™) were provided by Gary Felsenfeld (National Institutes of Health,

Bethesda, Maryland, USA) and were cultured in DMEM. All media contained 10% (v/v) fetal bovine serum (FBS) and 1% penicillin/streptomycin and were maintained in an incubator at 37°C with 5% CO<sub>2</sub>. Human lymphocytes were obtained from peripheral blood of a healthy donor, isolated with Ficoll-Paque Plus 2 (Amersham) following the manufacturer's instructions. Written informed consent was obtained from this healthy donor.

### Stable Transfection of K562 Cells

K562 cells were washed twice with phosphate buffered saline (PBS) and resuspended in DMEM. A total of  $3 \times 10^5$  K562 cells were then transferred to a 6-well plate and transfected with 1 µg (1 µg/µl) of corresponding linearized plasmids using Lipofectamine 2000 (Invitrogen, MA, USA) according to the manufacturer's instructions. After 6 h, 4 ml of non-selective medium was added to transfected cells. Following 48 h of cell recovery, the cells were transferred to media containing 0.9 mg/ml of G-418 (Geneticin, Calbiochem) for selection. Geneticin-resistant pools were analyzed by FACS at different time points (day 0, day 15, day 25, day 40, and day 60) to obtain the percentage (%) of GFP-positive cells and the mean fluorescence intensity from each construct at every time point. Data were analyzed with BD CellQuest Pro software (BD Biosciences). We performed four independent experiments and computed using the Graphpad Prisma Software 7.0. Statistically significant differences in mean fluorescence intensity values between the *Alu*-containing constructs and the one without it were computed using a two-tailed Mann-Whitney test ( $p < 0.05$ ).

### Cell Clone Isolation

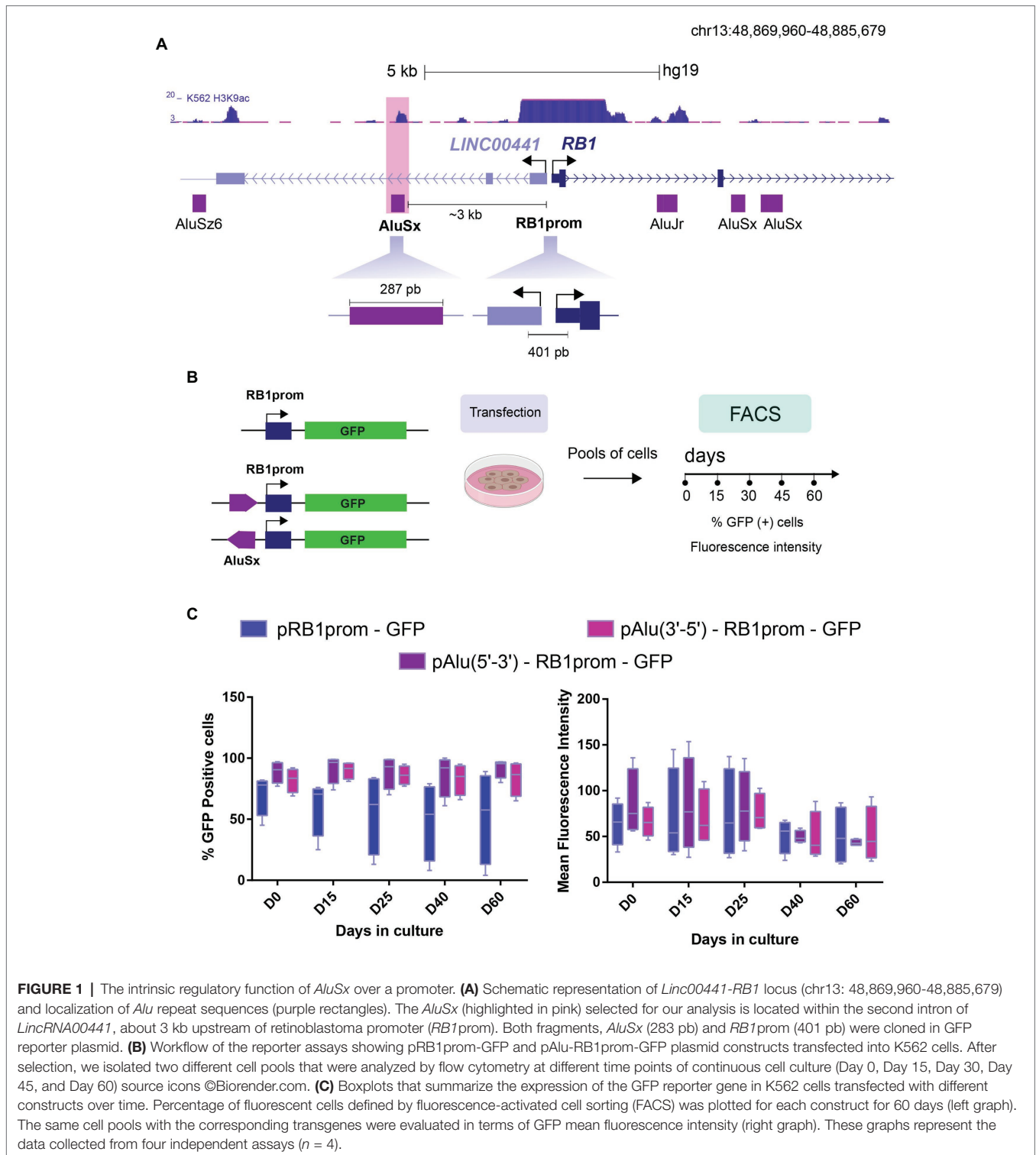
After 48 h of cell recovery,  $\sim 5 \times 10^5$  (500 µl) of transfected cells were transferred to a cellulose matrix (Methocel, FLuka) containing 0.9 mg/ml of G-418. Individual colonies were picked after 2–3 weeks and expanded in G-418 containing liquid DMEM to perform subsequent experiments. For each construct, we isolated and analyzed 14 independent clones at different points (day 0, day 15, day 30, day 45, day 60, day 80, and day 100) of continuous cell culture for 100 days. The integrity of the transgene was verified by PCR (Supplementary Figure S1C) and Southern blot (data not shown).

### CRISPR/Cas9-Mediated Targeted Deletion

*AluSx* element (chr13: 48,874,474-48,874,760, **Figure 1A**) upstream of the *RB1* gene promoter locus was deleted in K562 cells by co-transfecting two Cas9 containing plasmids plentiCRISPRv2 (Addgene Cat. 52961; Cambridge, MA, USA), each carrying a unique single guide RNA (sgRNA) flanking the *Alu* sequence.

We designed the CRISPR-Cas9 sgRNAs using CRISPOR<sup>1</sup>, as described (Haeussler et al., 2016), to minimize off-target effects. Bsmbl linkers were added to sgRNAs. The oligonucleotides were then annealed following a standard protocol (Sambrook and Russell, 2006), ligated into the vector and

<sup>1</sup><http://crispor.tefor.net/>



**FIGURE 1 |** The intrinsic regulatory function of *AluSx* over a promoter. **(A)** Schematic representation of *Linc00441-RB1* locus (chr13: 48,869,960-48,885,679) and localization of *Alu* repeat sequences (purple rectangles). The *AluSx* (highlighted in pink) selected for our analysis is located within the second intron of *LincRNA00441*, about 3 kb upstream of retinoblastoma promoter (*RB1*prom). Both fragments, *AluSx* (283 pb) and *RB1*prom (401 pb) were cloned in GFP reporter plasmid. **(B)** Workflow of the reporter assays showing pRB1prom-GFP and pAlu-RB1prom-GFP plasmid constructs transfected into K562 cells. After selection, we isolated two different cell pools that were analyzed by flow cytometry at different time points of continuous cell culture (Day 0, Day 15, Day 30, Day 45, and Day 60) source icons ©Biorender.com. **(C)** Boxplots that summarize the expression of the GFP reporter gene in K562 cells transfected with different constructs over time. Percentage of fluorescent cells defined by fluorescence-activated cell sorting (FACS) was plotted for each construct for 60 days (left graph). The same cell pools with the corresponding transgenes were evaluated in terms of GFP mean fluorescence intensity (right graph). These graphs represent the data collected from four independent assays ( $n = 4$ ).

confirmed by sequencing prior transfection. See Supplementary Table S1 for the list of sgRNAs sequences.

The plasmidic vectors were transfected into cells by using Lipofectamine 2000 (Invitrogen) as per manufacturer protocol. Cultures were then selected for 3–4 days with puromycin (5 µg/ml; Sigma). We obtained genomic DNA from CRISPR-Cas9

pools of transfected cells and screened by PCR to confirm the deletion of the evaluated fragment.

Then, CRISPR-Cas9 pools of transfected cells were seeded at low density to isolate monoclonal cell clones with respective mutations through serial dilution in a 96-well plate. Forty-eight randomly selected puromycin-resistant colonies were individually



expanded and splitted for future culture and genomic DNA isolation. A hundred nanograms of genomic DNA from these cell clones or K562 genomic DNA (control) were screened by PCR with primer pairs annealing to the region outside of the double-strand break (DSB) sites. See Supplementary Table S1 for the list of screening primers used to confirm targeted deletions. PCR reactions were evaluated on 1% agarose gels. Bands were excised from the gel and purified using the QIAquick Kit (QIAGEN) and the status of the deletion was confirmed by Sanger sequencing. We selected three homozygous clones for two different deletions of the *AluSx* repeat, for further analysis ( $\Delta$ Alu-C1,  $\Delta$ Alu-C2, and  $\Delta$ Alu-C3).

## RNA Extraction and Real-Time Quantitative PCR

Total RNA was extracted from K562 cells with TRIzol Reagent (Invitrogen) according to the manufacturer's instructions. RNA concentrations were determined using the NanoDrop ND-1000 spectrophotometer (NanoDrop Inc., DE, USA).

Real-time quantitative PCR (RT-qPCR) was carried out with KAPA SYBR® FAST One-Step qRT-PCR kit and specific primers for *RB1*, *Linc00441*, and  $\alpha$ -*tubulin* as an endogenous normalization control (Supplementary Table S1). The qPCR reactions were carried out in the StepOne detection system (Applied Biosystems) at 42°C for 5 min, 95°C for 30 s, followed by 40 three-step cycles of 95°C for 3 s, 62°C for 10 s, and 72°C for 10 s, in triplicate for each sample. Relative RNA levels were calculated using the comparative  $\Delta\Delta$ Ct method (Schmittgen and Livak, 2008). Statistically significant differences in gene expression between the wild-type and the CRISPR mutants were computed using a *t*-test ( $p < 0.05$ ) and the Graphpad Prisma Software 7.0.

## Cell Proliferation Assay

We used Trypan Blue assay to determine cell viability by counting viable cell numbers with a microscope in the following cell lines: K562 WT vs.  $\Delta$ Alu-C1,  $\Delta$ Alu-C2, and  $\Delta$ Alu-C3 (Strober, 2015). On day 0, we started with  $1 \times 10^5$  cells in a volume of 3 ml by triplicate. We counted in a hemocytometer the cell number for each condition every 24 h during 4 days. Finally, we plotted the average values from triplicates of cell number counts as a function of time for the different cell lines.

## Bioinformatic Analysis

### Motif Data Analysis

Identification of binding sites for TFs was done using the JASPAR (Fornes et al., 2020) Vertebrate Database using a threshold  $p < 0.0001$ .

## RESULTS

### An Intronic *Alu* Element Behaves as an Enhancer and Protects Against Epigenetic Silencing in Reporter Constructs

To investigate the regulatory function of *Alu* sequences, we chose the well-characterized *RB1* gene locus. We and others have previously shown, different epigenetic mechanisms are at

play to ensure proper control of *RB1* gene expression (De La Rosa-Velazquez et al., 2007; Dávalos-Salas et al., 2011). We hypothesized that the *Alu* sequences closest to the *RB1* gene promoter could impact its transcriptional regulation *via* two general mechanisms. Firstly, induced epigenetic silencing, as it has been reported that young *Alu* elements are epigenetically repressed (ref) and *Alu* sequences can gain DNA methylation in cancer cells (Akers et al., 2014; Bakshi et al., 2016; Jorda et al., 2017) and secondly, transcription boosting by acting as enhancers, as has been proposed for old *Alu* families (Su et al., 2014). Therefore, we retrieved the location of *Alu* sequences surrounding the minimal *RB1* gene promoter including sequences 5 kb upstream and downstream of the transcription start site (TSS), based on the annotation by Repeat Masker (Stirzaker et al., 1997; **Figure 1A**). The two *Alu* repeats closest to the *RB1* gene promoter are an *AluSx* and an *AluJr* element. The *AluSx* element is located 3 kb upstream of the *RB1* TSS and lies within the second intron of *Linc00441*, a lincRNA divergent to *RB1*, that has been shown to affect *RB1* transcription in cancer cells (Tang et al., 2017). The *AluJr* element is positioned in the first intron of *RB1* gene, about 2.5 kb downstream of the *RB1* TSS. Both *Alu* elements are characterized by multiple TFBS (**Supplementary Figure S1**) and are enriched for the euchromatin and associated histone post-translational modification H3K9ac. Notably, the H3K9ac, a histone mark associated with active enhancers, was recently found enriched in *Alu* elements expressed in a cell-type specific manner (Zhang et al., 2019). However, only *AluJr* partially overlaps with a region annotated as a promoter in the *RB1* locus based on Chromatin Segmentation by HMM from ENCODE (ENCODE Project Consortium, 2012). Additionally, it is immediately next to a *FLAM* SINE element, posing technical challenges for the manipulation of this sequence (**Supplementary Figure S1C**). Due to the *AluJr* sequence overlapping with other potential regulatory elements in the *RB1* locus, we dismissed working with the latter and decided to focus our study on the regulatory activity of the *AluSx* element upstream of *RB1* TSS.

Initially, to characterize the regulatory activity of the *AluSx* *in vitro*, we cloned the *Alu* element in both orientations (5'-3' and 3'-5') in a reporter plasmid containing the *RB1* promoter sequence (RBprom) and *GFP* as a reporter gene (**Figure 1B**). We have previously employed this reporter plasmid to monitor the epigenetic silencing of the *RB1* gene promoter (De La Rosa-Velazquez et al., 2007). Reporter constructs were transfected into K562 cell line and selected with geneticin to obtain pools of cells with stable integrants that were then evaluated by FACS. As expected, more than 50% of cells transfected with the pRB1prom-GFP construct were GFP-positive [GFP(+); **Figure 1C**, left, day 0], which is consistent with the promoter activity of this sequence. Unexpectedly, the mean number of GFP(+) cells carrying the *Alu*-containing construct in both orientations (pAlu(5'-3')-RB1prom-GFP and pAlu(3'-5')-RB1prom-GFP) was higher than the one observed in pRB1prom-GFP cells (89 and 82%, pAlu-RB1prom vs. 71%, pRB1prom; **Figure 1C**, left, day 0). A similar trend was observed when analyzing mean fluorescence intensity (86 and 66, pAlu-RB1prom vs. 64, pRB1prom **Figure 1C**, right, day 0). Importantly, we observed this regulatory effect in four

independent experiments, strongly suggesting that the *AluSx* sequence behaves as an enhancer element in this reporter assay. The increase in the number of GFP(+) cells irrespective of the *Alu* sequence orientation is an effect well-characterized for enhancer elements.

We have previously described the progressive epigenetic silencing of stably integrated transgenes over time in cell cultures (Dávalos-Salas et al., 2011). Since the acquisition of epigenetic silencing can be a time-dependent process, it was evaluated if the *Alu* element could still enhance transcription of the reporter gene despite their epigenetic silencing over time. Therefore, we followed pools of cells with stable integrants for the pRB1prom-GFP and pAlu-RB1prom-GFP transgenes over 60 days and quantified the number of GFP(+) cells. As expected for the *RB1* gene promoter, we observed a time-dependent reduction in the mean number of GFP(+) cells (71%, day 0 vs. 52%, day 60) and mean fluorescence intensity (64, day 0 vs. 50, day 60) indicative of epigenetic silencing of the *RB1* gene promoter as we have reported before (Dávalos-Salas et al., 2011). Remarkably, the presence of the *Alu* sequence upstream of the *RB1* gene promoter protected against epigenetic silencing. Accordingly, 92% of the pAlu-RB1prom-GFP cells were GFP(+) on day 60, in sharp contrast to just 52% GFP(+) cells containing the pRB1prom-GFP construct (Figure 1C, left). This effect is less evident at the level of mean fluorescence intensity, which is highly maintained in the pAlu-RB1prom-GFP cells during the first 25 days, and then decreases to the levels of pRB1prom-GFP cells (Figure 1C, right).

Finally, we also evaluated the regulatory activity of the *AluSx* in cell clones with stably integrated constructs and analyzed them for 100 days of continuous cell culture. Consistent with our results in pools of cells, we observed that the *Alu* sequence behaves as an enhancer and protects against epigenetic silencing although showing an increased variability, probably reflecting the effect of integration into different chromatin environments (Supplementary Figure S2). In summary, plasmid-based reporter assays suggest that the *Alu* sequence can act as an enhancer increasing the probability that more cells will become more transcriptionally active, opposed to promoting the number of transcription events in a specific population or acting as a protector against epigenetic silencing (Blackwood and Kadonaga, 1998).

### ***In situ* Deletion of the Intronic *Alu* Sequence Results in Changes in Transcription**

To assess the *in situ* function of the *AluSx* element, we employed the CRISPR-Cas9 system to generate a deletion of the *Alu* sequence in K562 cells. Using two sgRNAs targeting flanking sequences of the *AluSx* element, we generated a deletion of 353 bp (Figure 2A). After transfection, drug selection, and clonal dilution, we isolated three homozygous clonal cell lines that showed two different molecular lesions (Figures 2B,C). The mutant clone  $\Delta$ Alu-C1 is characterized by a deletion of 186 bp that removes 172 bp of the 3' region of the *AluSx*. In contrast, the mutant clones  $\Delta$ Alu-C2 and  $\Delta$ Alu-C3 have a deletion of 460 bp that removes the *Alu* sequence completely, as well as an additional 150 bp 5' upstream

of the repeat (Figure 2C and Supplementary Figure S2). Next, we evaluated the effect of the deletion on the transcription of *RB1* and the host gene *Linc00441* of the deleted *AluSx* sequence. We found that in the three mutant clones the transcription of *RB1* gene is only marginally affected, showing a tendency toward an increase that did not reach statistical significance. Unexpectedly, *Linc00441* expression was strongly upregulated; in particular, the increase was higher in the clones that lack the entire *AluSx* element (Figure 2D). Given our results in reporter constructs that suggest that *AluSx* possesses an intrinsic enhancer activity, the intragenic *AluSx* could be acting as an intragenic enhancer of *Linc00441*, which attenuates the host gene expression. Similar observations have been made for intragenic enhancers of protein-coding genes in humans (Cinghu et al., 2017).

Since *Linc00441* has been involved in cancer (Tang et al., 2017), we evaluated the effect of the *AluSx* element deletion on cell proliferation in mutant clones. The deletion of the *AluSx* element resulted in a decrease of K562 proliferation (Figure 2E). It is worth mentioning that when we analyzed the effect that *Alu* removal had on our three mutant clones, we observed different behaviors between  $\Delta$ Alu-C1 and  $\Delta$ Alu-C2/ $\Delta$ Alu-C3 both on the level of gene expression levels and cell proliferation.

It was proposed that *Alu* elements could also participate in transcription regulation by providing multiple TFBS when inserted in gene-rich regions (Polak and Domany, 2006). Therefore, we carried out an analysis of TFBS on the *AluSx*, identifying the ones that remained intact in clone C1 and not in clones C2 and C3 (Supplementary Figure S3). As we expected, we found several TFBS previously reported to be enriched in *Alu* sequences (Norris et al., 1995; Polak and Domany, 2006; Bouttier et al., 2016). Interestingly, the six TFBS that were absent in clones C2 and C3 was SREBF1, ESR1, Gfi1b, CTCFL, VDR, and TFAP2C that could have an essential role for these *AluSx* functioning in transcriptional regulation of *Linc00441*.

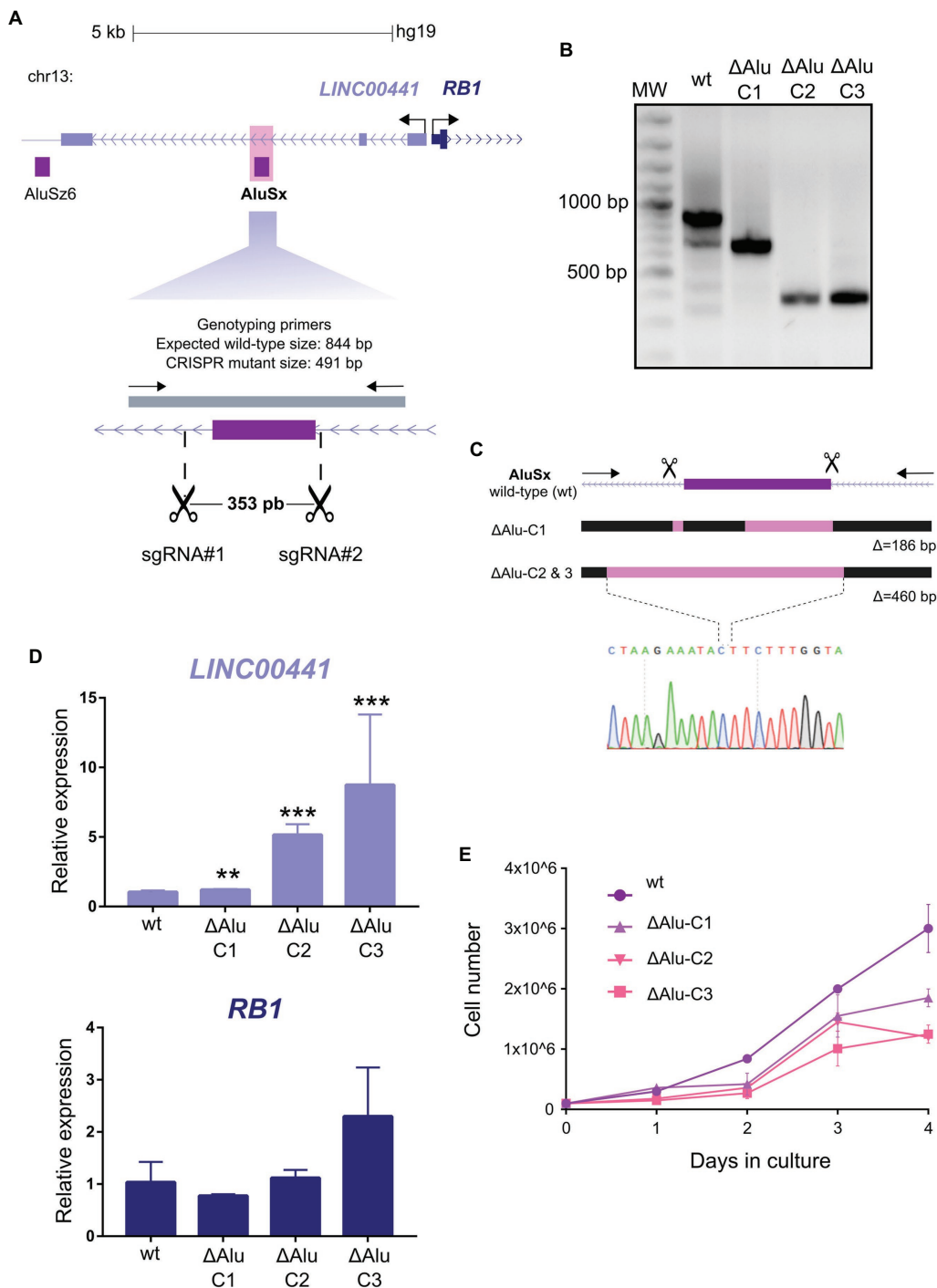
Taken together, the *in situ* deletion of the *AluSx* sequence promotes the transcription of the *Linc00441* gene, but not *RB1* gene, and importantly inhibits cell proliferation. This strongly suggests that *Alu* sequences can impact transcription and thereby change cellular phenotypes, such as cell proliferation.

## **DISCUSSION**

Recent studies have revealed a link between *Alu* elements and the control of gene expression (Hanke et al., 1995; Mallona et al., 2016; Chen et al., 2018). Many of these findings came from extensive computational analyses of genome-wide epigenomic and transcriptomic data (Goerner-Potvin and Bourque, 2018; Zhang et al., 2019); however, the direct functional testing of *Alu* sequences and their contribution repeat to gene regulatory networks have remained poorly explored.

Here, we provide evidence that an *Alu* repeat can behave as an enhancer and protect against epigenetic silencing using reporter assays. Furthermore, partial deletion or the complete removal of the *AluSx* from its endogenous locus increases the transcription of its host lincRNA, affecting cell proliferation. These findings contribute to our understanding of the regulatory





**FIGURE 2** | Functional contribution of *AluSx* to the *Linc00441* expression levels. **(A)** Design of CRISPR/Cas9 mediated deletion of *AluSx* in the *Linc00441-RB1* locus. The purple rectangle represents the *Alu* repeat. Scissors indicate the target sites of sgRNAs used to generate the deletion. The gray bar depicts the fragment amplified by the genotyping primers (arrows). **(B)** Genotyping of CRISPR mutant clones with deletions spanning the *AluSx* repeat. Expected wild type amplicon size, 844 base pairs (bp). Three CRISPR mutant clones  $\Delta$ Alu-C1,  $\Delta$ Alu-C2, and  $\Delta$ Alu-C3 homozygous for two different deletions of the *Alu* repeat. **(C)** Schematic representation of CRISPR mutant clones  $\Delta$ Alu-C1,  $\Delta$ Alu-C2, and  $\Delta$ Alu-C3 with deletions of the *AluSx* repeat. Pink rectangles represent the deleted sequence in each mutant. The electropherogram of the sequencing at the deletion breakpoints of  $\Delta$ Alu-C2 and  $\Delta$ Alu-C3 mutant clones is shown.  $\Delta$  = the number of base pairs deleted in each CRISPR mutant clone. **(D)** Gene expression analyses of *Linc00441* and *RB1* gene by real-time quantitative PCR (RT-qPCR). Transcriptional quantifications in each CRISPR mutant compared to wild type using *Linc00441* Exon 1 (upper graph) and *RB1* (bottom graph) gene specific primers. Significant differences between wild type and CRISPR mutants were calculated using a *t*-test  $n = 3$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Error bars represent the standard deviation (SD) of three independent experiments at least ( $n = 3$ ). **(E)** Quantification of cell proliferation by a trypan blue cell counting assay. The graph shows the average viable cell numbers of three replicates counted for 4 days on the three CRISPR mutant clones ( $\Delta$ Alu-C1,  $\Delta$ Alu-C2, and  $\Delta$ Alu-C3) compared to wild type.

potential these primate-specific sequences have in gene expression on the human genome.

In this study, we focused on the *AluSx* repeat located upstream of the *RBI* gene promoter and within *Linc00441*. The presence of this *Alu* repeat increases both the number of GFP-positive cells and the mean fluorescence intensity, suggesting that it acts as an enhancer element. In support of the aforementioned, a recent study analyzed genome-wide nucleosome occupancy, histone modification, and sequence motif features at *Alu* elements, concluding that *Alu* elements showed enhancer features (Su et al., 2014). Interestingly, the effect of the *AluSx* was more evident in the increase of GFP-positive cells instead of the mean fluorescence intensity. This supports the idea that *AluSx* repeat increases the burst frequency of transcription according to the binary model, where enhancers increase transcriptional levels of associated promoters (Blackwood and Kadonaga, 1998). Of note, while our data using plasmid-based reporters to test the regulatory potential of the *AluSx* suggest an effect in transcription and protection against epigenetic silencing, we cannot discard that the *AluSx* sequence, outside of its genomic context, could act as a non-specific DNA spacer that affect epigenetic silencing of the *RB-1* promoter.

Although we demonstrated the *cis*-regulatory effect of the *Alu* in reporter assays, these experiments are naturally limited by the fact that the repeat sequence is investigated independently of its native chromosomal context. Therefore, we chose to determine its role in the regulation of its host and neighboring genes and its implications in cell proliferation, through a combination of CRISPR-Cas9 mediated deletions and gene expression analysis. Contrary to the classical function of an enhancer element, we observed in mutants lacking the *AluSx* sequence a significant increase in the expression of *Linc00441*, but not *RBI*. From this, we conclude that *AluSx* must attenuate its host gene expression as reported for intragenic enhancers of protein-coding genes in humans (Cinghu et al., 2017). Notably, this attenuator effect was observed specifically for intragenic enhancer-containing genes with low-to-moderate expression levels in embryonic stem cells, reasoning that the enhancer's dominant function is presumably the one of an attenuator at these genes.

A recent report raised another important aspect for enhancer activity of TE that depends on the cooperative action of multiple TFs, whose binding motifs appear to have been already present in the corresponding ancestral TE insertions (Sundaram et al., 2017). We performed an analysis of the motifs of TFBS present in the *AluSx* repeat element. We found that many of the TFBS identified, are binding sites for nuclear factors, hormones ligands as well as other TFs related to differentiation processes, which correlates with motifs that have been reported to be enriched on *Alu* sequences (Polak and Domany, 2006). Interestingly, we found that six TFBS remain intact in the mutant that carry a partial deletion of the *AluSx*, probably related to the modest effect observed on expression and proliferation, compared with the mutants that have a complete removal of *AluSx*. Among these TFBS, we identified binding motifs related with metabolic pathways (SREBF1; sterol

biosynthesis), hormone response elements (ESR1 and VDR) and differentiation processes (Gfi1b in hematopoietic lineage and TFAP2C in early morphogenesis). Interestingly, we also identified a TFBS for CTCF paralog (CTCFL) that can be related with the protective effect against epigenetic silencing observed to this *AluSx*. However, how these TFs are involved in the enhancer activity of this *Alu* repeat is currently unknown. Trying to identify differential contribution of each TFBS related to the attenuator activity may help to get a better understanding of this novel function of repeat elements.

Given the recent observations that *Alu* sequences present enrichment of histone post-translational modifications associated with enhancer elements or the binding of RNA Pol II/III in a tissue or cell-type-specific manner (Su et al., 2014; Zhang et al., 2019), it would be of great interest to investigate the role of the *AluSx* in other cell-types. Such an experiment would inform on the presence of specific factors, such as cell-type specific transcription factors that impact the regulatory activity of an *Alu* sequence.

Further studies are required to investigate, on a genome-wide scale, the net impact of intragenic *Alu* in lincRNA expression. Nevertheless, our findings suggest that these repeat elements may be part of the complex machinery fine-tuning transcriptional regulation, highlighting the need for more functional assays to unravel the mechanisms of these enigmatic elements.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

Written informed consent was obtained from the donor.

## AUTHOR CONTRIBUTIONS

RP-M performed the experiments with the support of RA-M, EA-O, GG, and FS-O. RP-M, RA-M, and FR-T analyzed the data and performed the statistical analysis. RP-M, RA-M, and FR-T conceived the study, participated in its design, and wrote the manuscript. KM and EA-O critically revised the manuscript. All authors have read and approved the final version of the manuscript.

## FUNDING

This work is supported by a grant from Consejo Nacional de Ciencia y Tecnología (CONACyT: 128464 and 220503), Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica, UNAM (DGAPA-PAPIIT IN203811, IN201114 and IN203620) and Fundación Miguel Alemán to FR-T.

## ACKNOWLEDGMENTS

We are grateful with the Molecular Biology Unit at the Instituto de Fisiología Celular, UNAM for the services provided for the sequencing experiments. RP-M is a doctoral student from Programa de Doctorado en Ciencias Bioquímicas, Universidad Nacional Autónoma de México, and received fellowship 173754 from Consejo Nacional de Ciencia y Tecnología (CONACyT).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00928/full#supplementary-material>

**SUPPLEMENTARY FIGURE S1** | *Alu* elements surrounding the *Linc00441*-*RB1* locus are enriched for TF binding sites and show enrichment for H3K9ac.

**(A)** Genomic landscape surrounding the *Linc00441*-*RB1* locus. Data derived from the UCSC Genome Browser. **(B)** Genomic landscape of the *AluSx* element upstream of the *RB1* gene promoter and located in the intron 2 of *Linc00441*. **(C)** Genomic landscape of the *AluJr* element located in the intron 1 of *RB1*.

**SUPPLEMENTARY FIGURE S2** | The intrinsic regulatory effect of *AluSx* on a promoter. **(A)** Representative flow cytometry profiles of K562 cellular pools expressing the transgenes described in **Figure 1B** after 0, 15, 30, 45, and 60

days of continuous cell culture. **(B)** Schematic workflow of the isolation of individual cell lines carrying each one of the transgene reporter constructs. Fourteen cell independent clones were isolated after selection and the integrity of transgenes in each cell line was confirmed by PCR and Southern blotting (data not shown). Fourteen independent clones from each GFP-transgene were analyzed by FACS at different time points (Day 0, Day 15, Day 30, Day 45, Day 60, Day 80, and Day 100) of continuous cell culture (Source icons ©Biorender.com). **(C)** Amplification using three different primer pairs (depicted as blue, green, and orange arrows) was performed on genomic DNA obtained from each cell line. The blue and green primers were used to verify the integrity of the transgenes. Single and multi-copy integrants were determined using the orange primers. The expected amplicon lengths are indicated below the arrows. At the bottom a representative gel for the three PCRs is shown. **(D)** Boxplots that summarize the expression of the GFP reporter gene in the 14 isolated cell lines ( $n = 14$ ) obtained for each construct over time. The percentage (%) of fluorescent cells (upper graph) and the mean fluorescence intensity (bottom graph) determined by flow cytometry are shown. Significant differences between the different constructs were calculated using a Mann-Whitney test, with confidence level set as 95%. \*\* $p < 0.01$ .

**SUPPLEMENTARY FIGURE S3** | Analysis of motifs in *AluSx* and their removal in CRISPR mutants. Schematic representation of the region (chr13: 48,874,274-48,874,960) that contains the *AluSx* element upstream of the *RB1* gene promoter. Light purple rectangles represent the deleted region in each mutant **(top panel)**. Motif binding sites in *AluSx* (chr13: 48,874,474-48,874,760) by MEME ( $p < 0.0001$ ) are shown as boxes in light purple. Highlighted region in pink corresponds to the non-deleted sequence in the  $\Delta$ Alu-C1 mutant allele, with a partial deletion of the *AluSx*, containing six TFBS **(middle panel)**. Logo motifs from each TFBS with their corresponding  $p$  are shown **(bottom panel)**.

## REFERENCES


- Akers, S. N., Moysich, K., Zhang, W., Collamat Lai, G., Miller, A., Lele, S., et al. (2014). LINE1 and Alu repetitive element DNA methylation in tumors and white blood cells from epithelial ovarian cancer patients. *Gynecol. Oncol.* 132, 462–467. doi: 10.1016/j.ygyno.2013.12.024
- Bakshi, A., Herke, S. W., Batzer, M. A., and Kim, J. (2016). DNA methylation variation of human-specific Alu repeats. *Epigenetics* 11, 163–173. doi: 10.1080/15592294.2015.1130518
- Bannert, N., and Kurth, R. (2004). Retroelements and the human genome: new perspectives on an old relation. *Proc. Natl. Acad. Sci. U. S. A.* 101, 14572–14579. doi: 10.1073/pnas.0404838101
- Batzer, M. A., and Deininger, P. L. (2002). Alu repeats and human genomic diversity. *Nat. Rev. Genet.* 3, 370–379. doi: 10.1038/nrg798
- Baylin, S. B., Herman, J. G., Graff, J. R., Vertino, P. M., and Issa, J. P. (1998). Alterations in DNA methylation: a fundamental aspect of neoplasia. *Adv. Cancer Res.* 72, 141–196.
- Blackwood, E. M., and Kadonaga, J. T. (1998). Going the distance: a current view of enhancer action. *Science* 281, 60–63. doi: 10.1126/science.281.5373.60
- Bouttier, M., Laperriere, D., Memari, B., Mangiapane, J., Fiore, A., Mitchell, E., et al. (2016). Alu repeats as transcriptional regulatory platforms in macrophage responses to *M. tuberculosis* infection. *Nucleic Acids Res.* 44, 10571–10587. doi: 10.1093/nar/gkw782
- Chen, K., Wang, Y., and Sun, J. (2018). A statistical analysis on transcriptome sequences: the enrichment of Alu-element is associated with subcellular location. *Biochem. Biophys. Res. Commun.* 499, 397–402. doi: 10.1016/j.bbrc.2018.03.024
- Chuong, E. B., Elde, N. C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* 18, 71–86. doi: 10.1038/nrg.2016.139
- Cinghu, S., Yang, P., Kosak, J. P., Conway, A. E., Kumar, D., Oldfield, A. J., et al. (2017). Intragenic enhancers attenuate host gene expression. *Mol. Cell* 68, 104–117. doi: 10.1016/j.molcel.2017.09.010
- Dávalos-Salas, M., Furlan-Magaril, M., González-Buendía, E., Valdes-Quezada, C., Ayala-Ortega, E., and Recillas-Targa, F. (2011). Gain of DNA methylation is enhanced in the absence of CTCF at the human retinoblastoma gene promoter. *BMC Cancer* 11:232. doi: 10.1186/1471-2407-11-232
- De La Rosa-Velazquez, I. A., Rincón-Arano, H., Benitez-Bribiesca, L., and Recillas-Targa, F. (2007). Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Res.* 67, 2577–2585. doi: 10.1158/0008-5472.CAN-06-2024
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247
- Estecio, M. R., Gallegos, J., Dekmezian, M., Lu, Y., Liang, S., and Issa, J. P. (2012). SINE retrotransposons cause epigenetic reprogramming of adjacent gene promoters. *Mol. Cancer Res.* 10, 1332–1342. doi: 10.1158/1541-7786.MCR-12-0351
- Fornes, O., Castro-Mondragon, J. A., Khan, A., van der Lee, R., Zhang, X., Richmond, P. A., et al. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 48, D87–D92. doi: 10.1093/nar/gkz1001
- Goerner-Potvin, P., and Bourque, G. (2018). Computational tools to unmask transposable elements. *Nat. Rev. Genet.* 19, 688–704. doi: 10.1038/s41576-018-0050-x
- Graff, J. R., Herman, J. G., Myohanen, S., Baylin, S. B., and Vertino, P. M. (1997). Mapping patterns of CpG island methylation in normal and neoplastic cells implicates both upstream and downstream regions in de novo methylation. *J. Biol. Chem.* 272, 22322–22329. doi: 10.1074/jbc.272.35.22322
- Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J. -B., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* 17:148. doi: 10.1186/s13059-016-1012-2
- Hanke, J. H., Hambor, J. E., and Kavathas, P. (1995). Repetitive *Alu* elements form a cruciform structure that regulates the function of the human CD8 alpha T cell-specific enhancer. *J. Mol. Biol.* 246, 63–73. doi: 10.1006/jmbi.1994.0066
- Jorda, M., Diez-Villanueva, A., Mallona, I., Martin, B., Lois, S., Barrera, V., et al. (2017). The epigenetic landscape of *Alu* repeats delineates the structural and functional genomic architecture of colon cancer cells. *Genome Res.* 27, 118–132. doi: 10.1101/gr.207522.116
- Kim, E. Z., Wespiser, A. R., and Caffrey, D. R. (2016). The domain structure and distribution of Alu elements in long noncoding RNAs and mRNAs. *RNA* 22, 254–264. doi: 10.1261/rna.048280.114
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062

- Mallona, I., Jorda, M., and Peinado, M. A. (2016). A knowledgebase of the human *Alu* repetitive elements. *J. Biomed. Inform.* 60, 77–83. doi: 10.1016/j.jbi.2016.01.010
- Norris, J., Fan, D., Aleman, C., Marks, J. R., Futreal, P. A., Wiseman, R. W., et al. (1995). Identification of a new subclass of *Alu* DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers. *J. Biol. Chem.* 270, 22777–22782. doi: 10.1074/jbc.270.39.22777
- Polak, P., and Domany, E. (2006). *Alu* elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics* 7:133. doi: 10.1186/1471-2164-7-133
- Sambrook, J., and Russell, D. W. (2006). Purification of PCR products in preparation for cloning. *CSH Protoc.* 2006:pdb.prot3825. doi: 10.1101/pdb.prot3825
- Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.2008.73
- Stirzaker, C., Millar, D. S., Paul, C. L., Warnecke, P. M., Harrison, J., Vincent, P. C., et al. (1997). Extensive DNA methylation spanning the Rb promoter in retinoblastoma tumors. *Cancer Res.* 57, 2229–2237.
- Strober, W. (2015). Trypan blue exclusion test of cell viability. *Curr. Protoc. Immunol.* 111, A3.B.1–A3.B.3. doi: 10.1002/0471142735.ima03bs111
- Su, M., Han, D., Boyd-Kirkup, J., Yu, X., and Han, J. D. (2014). Evolution of *Alu* elements toward enhancers. *Cell Rep.* 7, 376–385. doi: 10.1016/j.celrep.2014.03.011
- Sundaram, V., Choudhary, M. N., Pehrsson, E., Xing, X., Fiore, C., Pandey, M., et al. (2017). Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. *Nat. Commun.* 8:14550. doi: 10.1038/ncomms14550
- Tang, J., Xie, Y., Xu, X., Yin, Y., Jiang, R., Deng, L., et al. (2017). Bidirectional transcription of Linc00441 and RB1 via H3K27 modification-dependent way promotes hepatocellular carcinoma. *Cell Death. Dis.* 8:e2675. doi: 10.1038/cddis.2017.81
- Zhang, X. O., Gingeras, T. R., and Weng, Z. (2019). Genome-wide analysis of polymerase III-transcribed *Alu* elements suggests cell-type-specific enhancer function. *Genome Res.* 29, 1402–1414. doi: 10.1101/gr.249789.119

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Pérez-Molina, Arzate-Mejía, Ayala-Ortega, Guerrero, Meier, Suaste-Olmos and Recillas-Targa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## Epigenetic Effects of an Adenosine Derivative in a Wistar Rat Model of Liver Cirrhosis

Jesús Rafael Rodríguez-Aguilera,<sup>1</sup> Carlos Guerrero-Hernández,<sup>1</sup> Rosario Pérez-Molina,<sup>2</sup> Carla Elizabeth Cadena-del-Castillo,<sup>2</sup> Rebeca Pérez-Cabeza de Vaca,<sup>1</sup> Nuria Guerrero-Celis,<sup>1</sup> Mariana Domínguez-López,<sup>1</sup> Adrián Rafael Murillo-de-Ozores,<sup>1</sup> Rodrigo Arzate-Mejía,<sup>2</sup> Félix Recillas-Targa,<sup>2</sup> and Victoria Chagoya de Sánchez <sup>1\*</sup>

<sup>1</sup>Departamento de Biología Celular y Desarrollo, Instituto de Fisiología Celular, UNAM, Circuito Exterior s/n Ciudad Universitaria, Coyoacán 04510, Cd.Mx., México

<sup>2</sup>Departamento de Genética Molecular, Instituto de Fisiología Celular, UNAM, Circuito Exterior s/n Ciudad Universitaria, Coyoacán 04510, Cd.Mx., México

### ABSTRACT

The pathological characteristic of cirrhosis is scarring which results in a structurally distorted and dysfunctional liver. Previously, we demonstrated that *Col1a1* and *Pparg* genes are deregulated in CCl<sub>4</sub>-induced cirrhosis but their normal expression levels are recovered upon treatment with IFC-305, an adenosine derivative. We observed that adenosine was able to modulate S-adenosylmethionine-dependent *trans*-methylation reactions, and recently, we found that IFC-305 modulates *HDAC3* expression. Here, we investigated whether epigenetic mechanisms, involving DNA methylation processes and histone acetylation, could explain the re-establishment of gene expression mediated by IFC-305 in cirrhosis. Therefore, Wistar rats were CCl<sub>4</sub> treated and a sub-group received IFC-305 to reverse fibrosis. Global changes in DNA methylation, 5-hydroxymethylation, and histone H4 acetylation were observed after treatment with IFC-305. In particular, during cirrhosis, the *Pparg* gene promoter is depleted of histone H4 acetylation, whereas IFC-305 administration restores normal histone acetylation levels which correlates with an increase of *Pparg* transcript and protein levels. In contrast, the promoter of *Col1a1* gene is hypomethylated during cirrhosis but gains DNA methylation upon treatment with IFC-305 which correlates with a reduction of *Col1a1* transcript and protein levels. Our results suggest a model in which cirrhosis results in a general loss of permissive chromatin histone marks which triggers the repression of the *Pparg* gene and the upregulation of the *Col1a1* gene. Treatment with IFC-305 restores epigenetic modifications globally and specifically at the promoters of *Pparg* and *Col1a1* genes. These results reveal one of the mechanisms of action of IFC-305 and suggest a possible therapeutic function in cirrhosis. *J. Cell. Biochem.* 119: 401–413, 2018. © 2017 Wiley Periodicals, Inc.

**KEY WORDS:** LIVER; DNA METHYLATION; CHROMATIN; *Col1a1*; HEPATOPROTECTOR

Cirrhosis originates from different mechanisms of liver injury that lead to necroinflammation and fibrogenesis. Histologically, it is characterized by diffuse nodular regeneration surrounded by dense fibrotic septa with subsequent parenchymal extinction and collapse of liver structures, with pronounced distortion of hepatic vascular architecture [Tsochatzis et al., 2014]. Fibrosis is largely

asymptomatic and its progression to cirrhosis is slow, developing over 20–40 years in patients with chronic liver injury. The scarring response of liver includes activation of hepatic stellate cells (HSCs), which represent resident mesenchymal cells, into contractile myofibroblasts that encapsulate the injury, and induce loss of vitamin A droplets. These activated cells enhance migration and

**ABBREVIATIONS:** 5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine; H4ac, histone H4 acetylation; H3K27me3, histone H3 lysine 27 trimethylation; PTM, post-translational modification.

**Conflicts of interest:** The authors who have taken part in this study declared that they do not have anything to disclose regarding conflict of interest with respect to this manuscript.

**Grant sponsor:** Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de México; **Grant numbers:** IN201114, IN203811, IN207012, IN225909; **Grant sponsor:** Consejo Nacional de Ciencia y Tecnología; **Grant numbers:** 128464, 220503; **Grant sponsor:** PAEP-UNAM, Programa de Maestría y Doctorado en Ciencias Bioquímicas, UNAM; **Grant number:** 30479367-5; **Grant sponsor:** PhD Fellowship from CONACyT; **Grant number:** CVU 508509.

\*Correspondence to: Victoria Chagoya de Sánchez, Circuito Exterior s/n, Ciudad Universitaria, Coyoacán, 04510 México, Cd.Mx., México. E-mail: vchagoya@ifc.unam.mx

Manuscript Received: 27 February 2017; Manuscript Accepted: 6 June 2017

Accepted manuscript online in Wiley Online Library (wileyonlinelibrary.com): 7 June 2017

DOI 10.1002/jcb.26192 • © 2017 Wiley Periodicals, Inc.



deposition of extracellular matrix components, like collagen I, the principal protein that triggers fibrosis [Hernandez-Gea and Friedman, 2011]. A widely used model to study cirrhosis is chronic administration of CCl<sub>4</sub> into rats [Paquet and Kamphausen, 1975] which has been considered an acceptable model to study human cirrhosis.

Chronic CCl<sub>4</sub>-intoxication of rats results in cirrhosis at 8–12 weeks [Iredale, 2007]. The CCl<sub>4</sub> metabolism produces lipid peroxidation and membrane damage, which are the main causes of hepatocellular alteration [Varela-Moreiras et al., 1995]. Furthermore, it causes chromosome instability suggesting involvement of DNA damage [Michalopoulos, 1990]. Previously, we demonstrated in this animal model that treatment with IFC-305, an adenosine derivative, reverses liver fibrosis and ameliorates hepatic function [Perez-Carreón et al., 2010]. At the cellular level, it prevents HSC activation *in vivo* and *in vitro* [Perez-Carreón et al., 2010; Velasco-Loyden et al., 2010], promotes recovery of the cell cycle inhibition [Chagoya de Sanchez et al., 2012], and re-establishes the expression of around 150 genes deregulated in cirrhosis, among them *Col1a1* and *Pparg* [Perez-Carreón et al., 2010].

With an increased understanding of chromatin organization of the eukaryotic genome, it has become clear that not only genetic but also epigenetic processes influence both normal biology and diseases [Recillas-Targa, 2014]. The peroxisome proliferator-activated receptor gamma (PPAR $\gamma$ ) is one of those factors whose activity is diminished in activated HSCs; in fact, its expression, *per se*, even maintains HSCs in a quiescent state [Hazra et al., 2004a]. PPAR $\gamma$  can be regulated by the methyl-CpG-binding protein 2 (MeCP2), which co-localizes with the deposition of the histone modification H3K27me3 and forms a repressive chromatin structure in the coding region of *Pparg* gene [Mann et al., 2010]. One of the most antifibrotic effects of PPAR $\gamma$  is its ability to inhibit collagen I gene transcription. This inhibition is mediated by the ability of the nuclear receptor to compete with NF- $\kappa$ B/p300 association to the *Col1a1* gene promoter in HSCs [Yavrom et al., 2005].

As IFC-305 affects the expression of a sub-set of genes during fibrosis reversion, we decided to evaluate if epigenetic mechanisms are involved. To this aim, we first evaluated global levels of methylated DNA (5mC), 5-hydroxymethylcytosine (5hmC), and histone H4 acetylation (H4ac). In addition, we analyzed the deposition of histone H4ac and H3K27me3 marks in the *Pparg* gene promoter and the DNA methylation state of the *Col1a1* gene promoter. We found that IFC-305 exerts its effect by regulating chromatin marks globally and in some genes directly involved in the fibrogenic process. We propose a key role of IFC-305 in fibrosis reversion that involves an increase in DNA methylation around the transcription start site of *Col1a1* gene and the acquisition of a transcriptional favoring combination of histone post-translational modifications (PTMs) at the *Pparg* gene promoter. These results can be understood as the molecular basis of hepatoprotective properties of IFC-305, and suggest it might have the potential to function as a therapeutic alternative for cirrhosis treatment; moreover, here we highlight the importance of epigenetic changes in cirrhosis establishment *in vivo*.

## RESULTS

### IFC-305 PROTECTS FROM CCl<sub>4</sub>-INDUCED LIVER FIBROSIS

The CCl<sub>4</sub>-mediated liver surface distortion was observed both in rats treated with the toxic compound for 10 weeks and in rats receiving in addition saline for another 5 weeks after CCl<sub>4</sub> (Supplementary Fig. S1B). We found nodules, an increment in alanine aminotransferase activity (ALT), and hepatomegaly (Fig. S1B–D). Treatment of cirrhotic rats for 5 weeks with IFC-305 triggered a rescue with a liver phenotype similar to animals without cirrhosis (Supplementary Fig. S1B). IFC-305 treatment also led to a decrease in ALT activity (Supplementary Fig. S1C), which suggests a hepatoprotective function by IFC-305 as it has been previously reported [Perez-Carreón et al., 2010]. Interestingly, the treatment of healthy rats with IFC-305 over 10 weeks does not generate any physiological effect denoted by no alteration of ALT as a biochemical parameter (Supplementary Fig. S1C: 10IFC). These results indicate that IFC-305 is capable of normalizing the hepatic function altered by cirrhosis and to accelerate the recovery of the liver macroscopic phenotype shown before the 5-week treatment.

To demonstrate that alterations in the liver phenotype can be attributed to collagen accumulation, mimicking human cirrhosis, we performed Western blot assays against collagen I, and found that there is a nearly 50% increment of collagen I in cirrhosis (Supplementary Fig. S2A). Treatment with IFC-305 diminished the amount of the protein as compared to the cirrhosis and saline groups. A similar effect was observed with Masson Trichrome staining (Supplementary Fig. S2B). Together, these data confirmed that IFC-305 has hepatoprotective properties and reduces the amount of collagen I in the liver.

### IFC-305 REGULATES S-ADENOSYLMETHIONINE (SAM) LEVELS AND DNA METHYLATION

We have previously shown that adenosine can modulate *trans*-methylation reactions, like methylation of phospholipids, via regulation of SAM levels [Chagoya de Sanchez et al., 1991]. Therefore, to assess if IFC-305 has an impact on DNA methylation, we first quantified the level of SAM and found that the amount of this molecule is diminished in cirrhosis, whereas IFC-305 treatment restored the physiological levels (Fig. 1A).

The effects of cirrhosis at the epigenetic level are little understood, but there are few studies that have demonstrated that the DNA methylation status was reduced throughout the genome because of CCl<sub>4</sub> treatment in the early-stage liver fibrosis [Varela-Moreiras et al., 1995; Komatsu et al., 2012]. To determine whether the hepatoprotective effects of IFC-305 involve changes in DNA methylation, we evaluated global DNA methylation levels in the cirrhosis model. We found a reduction in DNA methylation at the cirrhosis state and a regaining of the 5mC modification with the treatment with IFC-305 (Fig. 1B and C). These samples were analyzed by ELISA which revealed a twofold increase in the level of DNA methylation, when comparing IFC-305 treatment against healthy and cirrhotic liver states (Fig. 1D). These results demonstrate that IFC-305, as adenosine itself, regulates the *trans*-methylation reactions through modulating the levels of SAM and suggest an epigenetic effect of IFC-305 in cirrhotic rats.

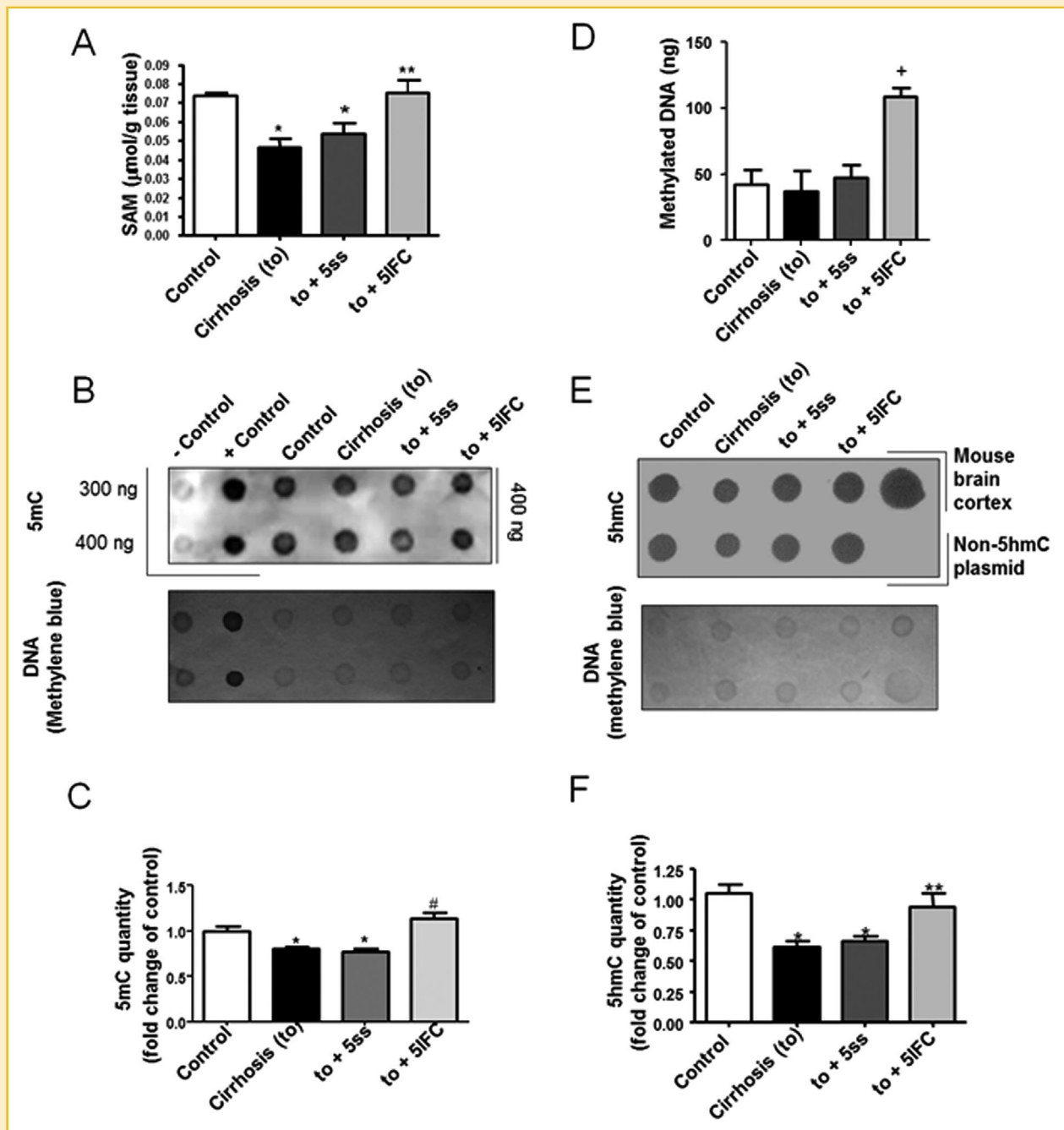


Fig. 1. Timelapse analysis of SAM levels and DNA methylation dynamics in the CCl<sub>4</sub>-induced cirrhosis model. (A) SAM quantification. (B) Global DNA methylation analysis. Upper panel shows a representative image of a biological triplicate of dot-blot; Lower panel represents the membrane stained with methylene blue. (C) 5mC dot-blot quantification. (D) DNA methylation analysis by ELISA assay (E) 5hmC quantification. Upper panel shows a representative image of a biological triplicate; Lower panel represents the membrane stained with methylene blue. (F) 5hmC dot-blot quantification. Data represent mean  $\pm$  SEM from three rats/group. \*Statistical difference ( $P < 0.05$ ) compared to Control group. \*\*Statistical difference ( $P < 0.05$ ) when compared to Cirrhosis (to). #Statistical difference ( $P < 0.05$ ) compared to Cirrhosis (to) and to +5ss. +Statistical difference ( $P < 0.05$ ) compared to Control, Cirrhosis (to) and to +5ss.

### 5-HYDROXYMETHYLCYTOSINE IS REDUCED DURING CIRRHOSIS AND IFC-305 RESTORES ITS PHYSIOLOGICAL LEVELS

Next, we decided to decipher in more detail other epigenetic processes in response to IFC-305 treatment. The presence of DNA methylation has been considered a synonym of gene expression

silencing and repressive chromatin structure [Recillas-Targa, 2014]. The discovery of an active DNA demethylation pathway, that involves the conversion of 5mC to oxidized forms, like 5hmC, and DNA repair through the base excision system, incorporates a dynamic reversibility of DNA methylation

[Kriaucionis and Heintz, 2009; Tahiliani et al., 2009; Bhutani et al., 2011; Kohli and Zhang, 2013]. Considering the significant number of genes modified during cirrhosis [Perez-Carreón et al., 2010] and the effect of IFC-305 on global DNA methylation (Fig. 1B–D), we analyzed global levels of 5hmC.

We performed dot-blot assays using an anti-5hmC antibody and found that 5hmC is diminished in cirrhosis. Interestingly, we observed a rescue of 5hmC levels with IFC-305 treatment in cirrhotic rats similar to the one observed in control rats (Fig. 1E and F). Consequently, it seems that IFC-305 can modulate the active DNA demethylation process, thereby restoring the amount of 5hmC to levels found in healthy liver. Taken together, these data suggest that IFC-305 treatment results in a restoration of 5hmC to physiological levels, while increasing the amount of 5mC genome wide.

### HISTONE H4 ACETYLATION DECREASES IN CIRRHOSIS AND IFC-305 REESTABLISHES ITS PHYSIOLOGICAL LEVEL

DNA methylation is associated with gene silencing and the establishment of a close chromatin conformation, in part, by modulation of histone acetylation levels. Our previous transcriptome analysis showed that HDAC3, a histone deacetylase, is overexpressed in cirrhotic liver in rats [Perez-Carreón et al., 2010]. Administration of IFC-305 for 5 weeks triggers a 30% decrease of the HDAC3 expression level [Perez-Carreón et al., 2010], which suggests that IFC-305 could restore histone acetylation levels in part by down regulating *HDAC3* gene expression. To validate this, we first quantified HDAC3 protein levels and found a twofold increase in the cirrhotic state compared to non-treated rats and healthy rats treated for 10 weeks with IFC-305 (10IFC). In contrast, we found that the amount of HDAC3 after treating the cirrhotic state for 5 weeks with IFC-305 was similar to the one observed in a healthy liver (Fig. 2A). To investigate if the HDAC3 protein levels correlate with changes in

histone acetylation we carried out Western blots of nuclear protein extracts using an antibody against acetylated histone H4. Importantly, we found a correlation between increased levels of HDAC3 and diminishment of hyperacetylated histone H4 in cirrhosis, while IFC-305 was able to restore the healthy liver level of this histone mark (Fig. 2B). This change is HDAC3 specific as no differences in the protein level of HDAC1 could be observed in cirrhosis or 5 weeks of treatment with IFC-305 (Supplementary Fig. S4). Together, these data demonstrate that treatment of cirrhotic rats with IFC-305 modulates the level of DNA methylation modifications 5mC and 5hmC, as well as the PTM histone H4ac, which regulate epigenetically gene expression.

### THE OVEREXPRESSION OF *Pparg* GENE BY IFC-305 IS MEDIATED BY THE GENERATION OF OPEN CHROMATIN IN THE GENE PROMOTER

Once we established that cirrhosis generates global chromatin alterations which are to some extent reversed by IFC-305 administration, we decided to analyze in detail the chromatin configuration of gene promoters that exhibit an altered expression in cirrhosis and are involved in the fibrogenic process, such as *Col1a1* and *Pparg*.

A few years ago, it has been demonstrated that PPAR $\gamma$  is a negative regulator of *Col1a1* gene expression in HSCs [Yavrom et al., 2005]. Taking this into account, we first analyzed by Western blot the amount of nuclear receptor in the cirrhotic model and found that protein levels correlated with transcript levels previously described [Perez-Carreón et al., 2010]. In cirrhosis we observed a decrease of the protein level, whereas IFC-305 increased its concentration beyond the normal liver level (Fig. 3A). In order to know which type of cells were responsible for the expression changes of PPAR $\gamma$ , we performed immunofluorescence assays and found that PPAR $\gamma$  signal was detected in a considerable number of cells in healthy liver. In

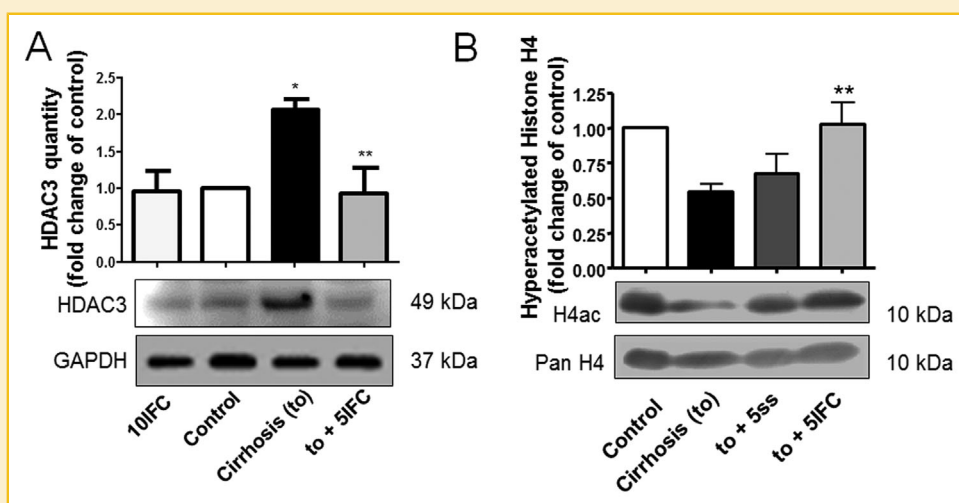


Fig. 2. Histone H4 acetylation levels are recovered after IFC-305 treatment of cirrhotic rats. (A) HDAC3 protein levels in  $\text{CCl}_4$  model. Representative Western blots are shown at the bottom of densitometric analysis. GAPDH immunodetection was used as a loading control. Data represent mean  $\pm$  SEM from three rats/group. (B) Immunodetection of global hyperacetylated histone H4 levels. Representative Western blot images are shown at the bottom of densitometric analysis. Total H4 immunodetection using a pan-Histone H4 antibody was used as a loading control. Data represent mean  $\pm$  SEM from three rats/group. \*Statistical difference ( $P < 0.05$ ) compared to Control group. \*\*Statistical difference ( $P < 0.05$ ) when compared to Cirrhosis (to).



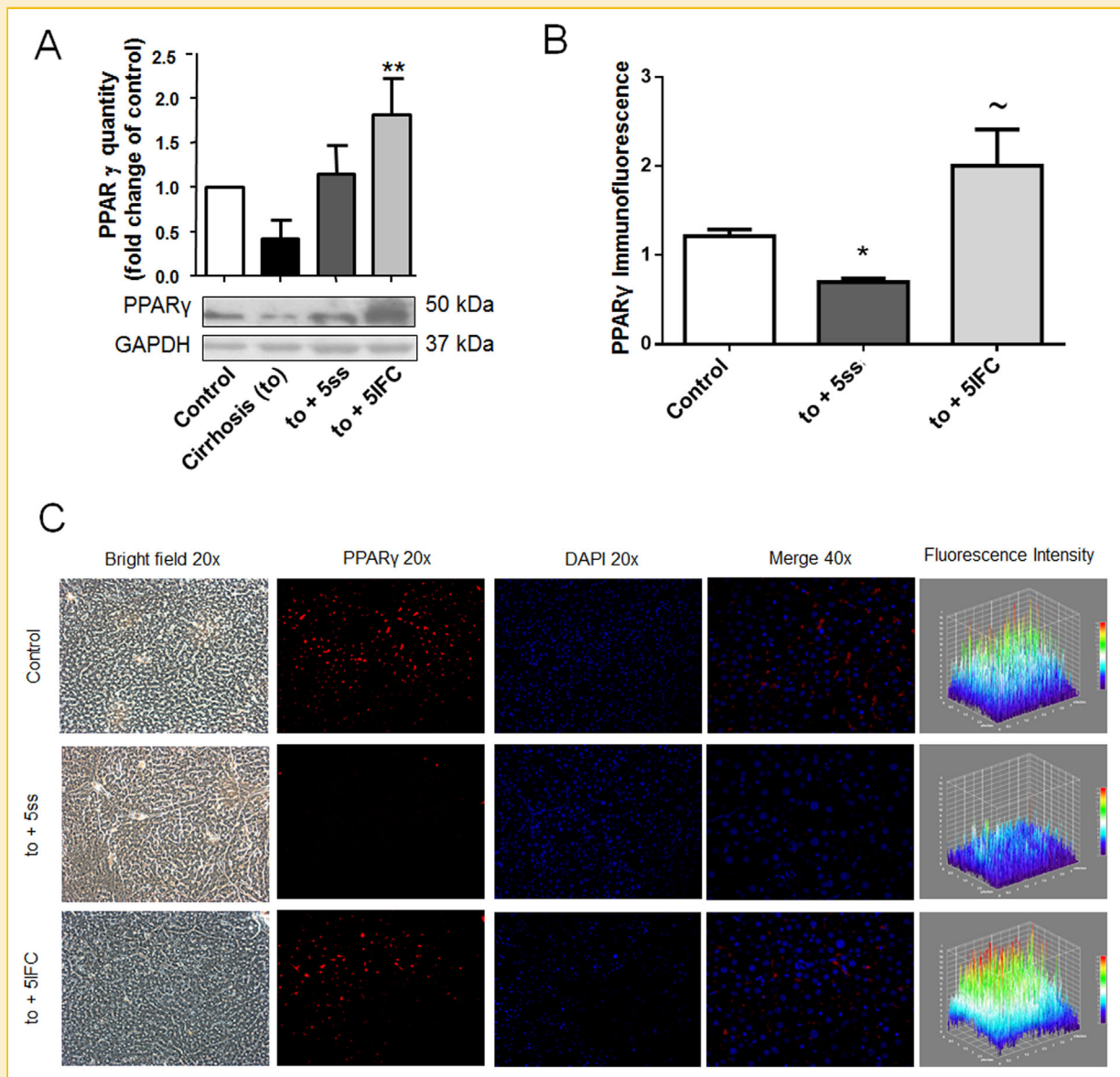


Fig. 3. The nuclear receptor PPAR $\gamma$  levels increase in cirrhotic liver treated with IFC-305. (A) Representative Western blot images are shown at the bottom of densitometric analysis. Data represent mean  $\pm$  SEM from four rats/group. \*\*Statistical difference ( $P < 0.05$ ) when compared to Cirrhosis (to). (B) Quantification of immunofluorescence of the positive signal to PPAR $\gamma$ . Data represent mean  $\pm$  SEM from three fields/group. \*Statistical difference ( $P < 0.05$ ) when compared to control group. ~Statistical difference ( $P < 0.05$ ) when compared to +5ss. (C) Immunofluorescence of PPAR $\gamma$  in whole liver histological sections. Representative images are shown.

case of the cirrhotic liver (to +5ss) the positive signal decreases strongly, whereas the IFC-305 treatment leads to a recovery of the levels of PPAR $\gamma$  signal (Fig. 3B and C). Analyzing the location of cells with a positive PPAR $\gamma$  signal in the to +5IFC condition, we observed that immunoreactive cells were found in regions where there is not residual of fibrosis septa. When we immunostained sections of fibrotic tissue with an antibody against  $\alpha$ SMA, a marker of activated HSCs, we observed in to +5ss an enrichment of  $\alpha$ SMA positive cells around the fibrosis septa, but with the IFC-305 treatment the immunoreactivity decreased (Supplementary Fig. S3). The localization of  $\alpha$ SMA positive cells is consistent with our previous data [Perez-Carreón et al., 2010].

To determine if the increased amount of PPAR $\gamma$  after treatment of cirrhotic rats with IFC-305 is mediated by changes of histone PTM at the *Pparg* gene promoter, we performed chromatin immunoprecipitations (ChIP) using an antibody against acetylated histone H4. We observed that the acetylated histone H4 mark decreased in the liver of cirrhotic rats, whereas in rats treated with IFC-305, the histone H4ac modification recovered reaching normal levels (Fig. 4A). Additionally, we analyzed chromatin samples from a group of healthy rats treated with IFC-305 over 10 weeks (10 IFC), and we did not observe a significant difference compared with control rats, only when compared with cirrhosis (to) (Fig. 4A: 10 IFC). Together, these data suggest that the increase in

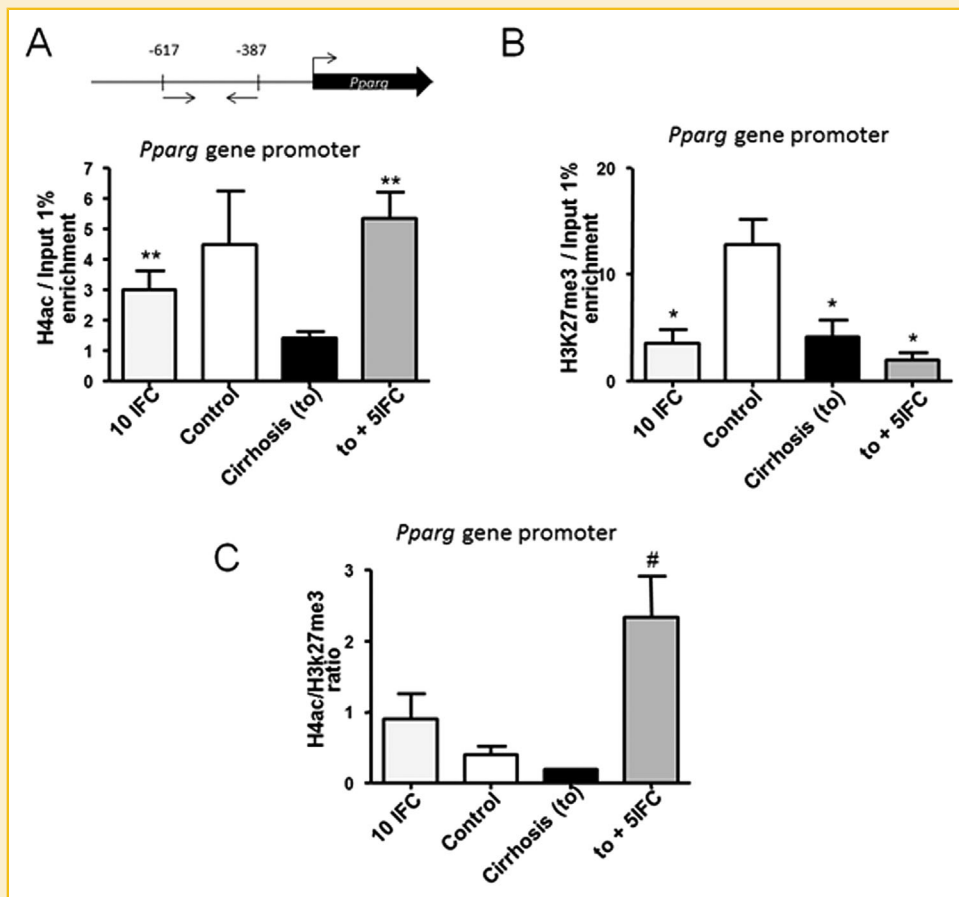


Fig. 4. Combination histone PTMs on the *Pparg* gene promoter and deposition of PPAR $\gamma$  at the *Col1a1* gene promoter in the CCl $_4$ -induced cirrhosis model. (A) Scheme of *Pparg* gene promoter depicting the positions of primers used in A–C; enrichment of permissive hyperacetylated histone H4 mark on *Pparg* gene promoter in a pre-established cirrhosis treated with the IFC-305 compound. (B) Deposition of repressive histone mark H3K27me3 on *Pparg* gene promoter. (C) Ratio of H4ac and H3K27me3 histone marks enrichment. Data represent mean  $\pm$  SEM of two biological samples analyzed in triplicate. \*Statistical difference ( $P < 0.05$ ) compared to Control (to) group. \*\*Statistical difference ( $P < 0.05$ ) when compared to Cirrhosis (to). #Statistical difference ( $P < 0.05$ ) when compared to Control and Cirrhosis (to).

transcription and consequently in PPAR $\gamma$  protein levels result from a restoration of open chromatin marks at the promoter of the *Pparg* gene mediated by IFC-305 treatment of the cirrhotic liver (Fig. 4A: to +5IFC). To further validate the open chromatin state of the *Pparg* gene promoter we performed ChIP assays using an antibody against the repressive histone mark H3K27me3 deposited by the Polycomb repressive complex 2. H3K27me3 ChIP assays revealed a reduction in the deposition of the repressive histone mark on the *Pparg* gene promoter region in all treatments compared to the healthy liver (Fig. 4B). We calculated the ratio of enrichment between H4ac and H3K27me3, and determined that the *Pparg* promoter is preferentially found in a transcriptionally permissive chromatin structure in the cirrhotic liver treated with IFC-305 (Fig. 4C: to +5IFC).

Next, we evaluated the DNA methylation state of the promoter and 5' region of the *Pparg* gene, that includes a segment of the first exon, by sodium bisulfite conversion and sequencing. Previously, the group of Mann et al. [2010] described that in myofibroblasts derived from HSCs the promoter and first exon of the *Pparg* are enriched for MeCP2 binding, and the first exon additionally for

MBD2 and HP1 $\alpha$ , proteins directly linked to DNA methylation and heterochromatin, respectively. Under our experimental conditions promoter and 5' region of the *Pparg* gene were not DNA methylated (Supplementary Fig. S5). Therefore, these results suggest that over expression of the *Pparg* gene induced by IFC-305 [Perez-Carreón et al., 2010], in a pre-established cirrhosis, is mediated not by DNA methylation, but mainly by histone modifications, favoring its transcription.

#### REPRESSION OF THE *Col1a1* GENE BY IFC-305 IS MAINLY MEDIATED BY AN INCREMENT IN DNA METHYLATION AROUND THE TRANSCRIPTION START SITE OF THE GENE

Previous work published by Tsukamoto and coworkers demonstrated that PPAR $\gamma$  suppresses the proximal *Col1a1* promoter [Yavrom et al., 2005]. When we evaluated the deposition of nuclear receptor on the *Col1a1* gene promoter, we found in cirrhosis no difference in the deposition of the nuclear receptor as compared to a healthy liver (Supplementary Fig. S6). These data suggest that the repression mechanism of *Col1a1* gene expression mediated by PPAR $\gamma$  is particular for HSCs, as previously shown [Yavrom et al., 2005].

To investigate whether the cause for collagen I over expression during cirrhosis is related to a reduction of DNA methylation on its gene promoter we applied sodium bisulfite conversion and sequencing to the *Col1a1* gene promoter (Fig. 5). This regulatory sequence is known to contain two protected footprints as determined by DNase footprinting analysis of activated HSCs (–133 to –79 bp) [Nehls et al., 1991; Rippe et al., 1995; Yavrom et al., 2005]. Furthermore, it contains the sites via which PPAR $\gamma$  renders a mayor inhibitory effect on the promoter and in vitro the transcription factors NF-1 and Sp1 bind to it in activated HSCs [Yavrom et al., 2005] (Fig. 5A). We found that in the *Col1a1* gene promoter and first exon, the CpGs in positions 1, 11, and 17 lose DNA methylation in cirrhosis (Fig. 5B: Cirrhosis [to]). Interestingly, the loss of 5mC is sustained during the next 5 weeks in the absence of CCl $_4$  (Fig. 5B; to +5ss; CpGs 1, 4, 11, and 17). On the other hand, treatment with IFC-305 increased CpG methylation around the transcription start site (TSS) and fostered recovery of CpG methylation at the first position (Fig. 5B; to +5IFC; CpG 1). The previously described behavior at specific CpG dinucleotides is represented by the trend line in the graphic of % of DNA methylation, which depicts the level of methylation of each CpG dinucleotide. Comparing the percentage of DNA methylation in this trend line, we

observed that the treatment with IFC-305 triggers a significant increment in DNA methylation compared with to +5ss (Fig 5B: % of DNA methylation, trend line). This suggests that *Col1a1* gene expression which is overexpressed in cirrhosis and returns to control levels with IFC-305 treatment [Perez-Carreon et al., 2010] shows an inverse correlation with the DNA methylation status in the *Col1a1* gene promoter in the used cirrhosis model.

#### THE PREVENTION OF IN VITRO HSC ACTIVATION BY IFC-305 APPEARS TO BE COORDINATED BY EPIGENETIC MECHANISMS

The liver tissue is composed of different cell types, whereby the hepatocytes represent the vast majority of the cells of the organ. Taking this into account, it is probable that the observations described above result from changes that occur primarily in these parenchymal cells. Moreover, the cells that are mainly responsible for the development of fibrosis are HSCs, which represent 8% of the cellular population of the liver. Therefore, in order to validate the origin of epigenetic changes mediated by IFC-305 treatment we assessed the levels of hyperacetylated histone H4 on a global level in a previously described model of prevention of HSC activation in vitro [Velasco-Loyden et al., 2010].

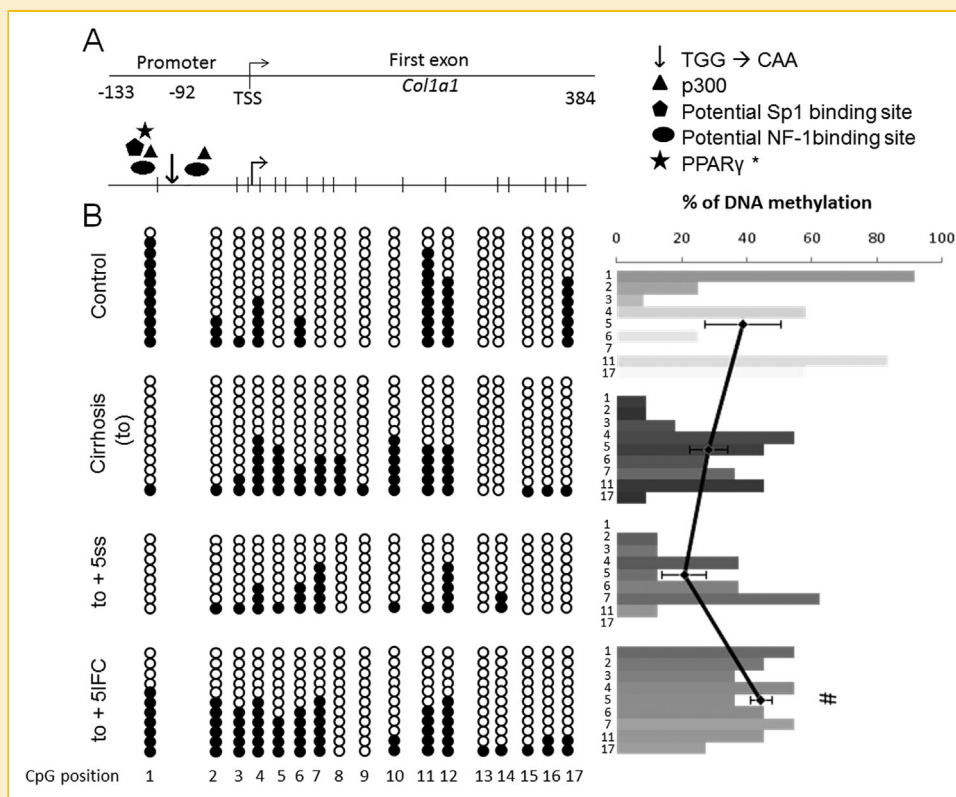


Fig. 5. DNA methylation analysis of the *Col1a1* gene promoter and first exon. (A) Scheme of the analyzed region of the *Col1a1* gene promoter. Pentagon and oval indicate putative Sp1- and NF-1-binding sites, respectively; the triangle denotes the region to which p300 is recruited and the star corresponds to the region where PPAR $\gamma$  inhibits p300-facilitated NF-1 binding to DNA in HSCs to abrogate *Col1a1* gene expression. Vertical arrow marks a sequence which is critical for promoter function, a 3-nucleotide mutation results in an almost complete loss of promoter activity [Nehls et al., 1991; Yavrom et al., 2005]. (B) Left panel: DNA methylation status, white (○) and black (●) dots represent unmethylated and methylated CpG, respectively. Columns represent each CpG site and rows show all individual clones analyzed. For the scheme two independent biological samples were analyzed. Right panel: Percentage of DNA methylation, which was calculated as the ratio between methylated CpGs and the number of analyzed clones for each condition. The trend line was obtained from the mean of ratios of all analyzed CpGs of each condition. #Statistical difference ( $P < 0.05$ ) compared to Cirrhosis (to) and to +5ss.

In this model we have previously demonstrated that IFC-305 prevents the activation of isolated HSCs [Velasco-Loyden et al., 2010]. Here, we isolated primary cultures of HSCs and observed that the activation process directs an increment of hyperacetylated histone H4. Additionally, we found an enrichment of this histone modification at 7 days of activation, whereas the culture treated with 5 mM IFC-305 prevented an acquisition of the histone H4 acetylation (Supplementary Fig. S7). These results propose that the *in vitro* activation of HSC presents a different behavior compared to the whole liver. Due to this and considering the localization of PPAR $\gamma$  detected by immunofluorescence in liver sections, it is most likely that the epigenetic changes observed in the cirrhosis model can be mainly attributed to hepatocytes.

## DISCUSSION

Cirrhosis is the 14th cause of death worldwide [Tsochatzis et al., 2014] and today no curative treatment exists, except for liver transplantation. We propose IFC-305, an adenosine derivative compound, as a promising alternative to treat liver injury. The aim of the present study is to explore the mechanism of action of IFC-305 in cirrhotic rats and the contribution of epigenetic mechanisms implicated in fibrosis reversion.

Based on our cirrhotic model we observed in dot-blot assay, a diminishment in DNA methylation during cirrhosis which is in concordance with other reports [Varela-Moreiras et al., 1995; Komatsu et al., 2012]. On the other hand, we obtained restoration of global genomic DNA methylation levels with the administration of the hepatoprotector IFC-305 (Fig. 1B and C). In contrast, in ELISA assay we observed a hypermethylation of the genomic DNA in response to IFC-305 treatment (Fig. 1D). It is conceivable that this variation can be attributed to the difference in DNA samples used for both assays, as the DNA for the first assay was sonicated to enhance the recognition of the antibody to the 5mC epitope and for the ELISA assay integral DNA was used as the kit's manufacturer indicates. One possible explanation for the regulation of DNA methylation by IFC-305 lies in its ability to restore SAM levels decreased at cirrhosis (Fig. 1A). From its chemical structure the IFC-305 compound is a derivative of adenosine, which has previously been shown to have a role in the regulation of SAM-dependent methylation reactions [Chagoya de Sanchez et al., 1991]. Furthermore, it may be explained by the increment of energy charge in hepatocytes mediated by adenosine [Chagoya de Sanchez et al., 1972] favoring ATP-dependent reactions such as transformation of methionine to SAM by methionine adenosyltransferases. In addition, there are reports that indicate that treatment of preneoplastic rats with hepatic nodules with SAM triggers a modification of DNA methylation leading to changes in gene expression [Garcea et al., 1989]. Therefore, the increase in DNA methylation induced by IFC-305 administration could be mediated through a regulation of SAM levels (Fig. 1A). Both increments mediated by IFC-305 could counteract the depletion of DNA methylation in the cirrhotic tissue, which is linked to a decrease of genomic instability, that is a characteristic of cirrhosis [Michalopoulos, 1990] and contributes to carcinogenesis [Hamilton, 2011].

In an attempt to find a correlative relationship with DNA methylation, we analyzed the incorporation of 5hmC and found a decrease of this DNA modification in cirrhotic samples, an effect that was abrogated by IFC-305 administration (Fig. 1E and F). As an approach to gain an insight into the effects caused by IFC-305 on an epigenetic level, we used assays to detect global 5mC and 5hmC levels that revealed an average of all the changes taking place genome wide, and did not give us information about the alterations of these DNA modifications at specific loci. For this reason, as we observe a mean increment in both marks in response to IFC-305 treatment, instead of a negative correlation, we suppose that the changes of each cytosine variant occur in different regions of the genome. It is possible, as suggested by the group of Mann, that specific CpGs become *de novo* methylated in order to recruit chromatin remodeling complexes that silence gene expression, while other CpGs become *de novo* methylated to subsequently become hydroxymethylated which would then instead lead to transcriptional activation or an increase in transcriptional elongation [Page et al., 2016]. However, oxidative bisulfite assays coupled to next-generation sequencing are needed to prove this hypothesis. Furthermore, it is important to keep in mind that during cirrhosis we observed a decrease in SAM, which is the universal donor of methyl groups, and therefore, this change may reflect a diminishment of methyl groups that are available for DNA methyltransferases (DNMTs). Thus, we can argue that the diminishment of 5mC can be attributed to the deficiency of SAM and not to active demethylation. This could also explain the levels of 5hmC which is generated from 5mC, a DNA modification that is already low under this condition. The initial description of 5hmC incorporation to DNA was reported in the liver and brain of rat [Penn et al., 1972], and recently a global reduction of 5hmC and its regulatory enzymes that accompany liver fibrosis and HSC *trans*-differentiation was documented [Page et al., 2016]. When we view cirrhosis as a preneoplastic state, reduction of 5hmC is consistent with studies in patients where hepatic, breast, lung, and pancreas cancer samples present a drop in 5hmC levels compared to adjacent tissue or healthy tissue samples [Yang et al., 2013].

The quantification of acetylated histone H4 revealed a 50% reduction during cirrhosis and IFC-305 treatment led to a recovery to normal levels (Fig. 2B). Histone acetylation generates an open chromatin structure that facilitates the association of transcription factors to their DNA binding sites in gene regulatory regions favoring controlled gene expression [Recillas-Targa, 2014]. Thus, this could be a mechanism by which IFC-305 treatment reduces the amount of deregulated genes in cirrhotic rats from 413 to 263 [Perez-Carreón et al., 2010]. The histone acetylation dynamics in our cirrhosis model could be particularly explained by the altered transcription [Perez-Carreón et al., 2010] and protein levels of HDAC3, as the amount of enzyme increases in cirrhosis and decreases with IFC-305 administration about 50% (Fig. 2A). Notably, the treatment of healthy rats with IFC-305 for 10 weeks generates a decrease in HDAC1 protein levels (Fig. S4: 10 IFC). Interestingly, however, no alteration of the liver physiology can be observed, when IFC-305 is administered to healthy rats, despite the global change in HDAC1 levels (Supplementary Fig. S1C: 10IFC).



The study of deposition of histone marks in the *Pparg* gene promoter demonstrated that, in cirrhosis, the chromatin presents a combinatorial code of histone PTMs at this regulatory sequence that does not favor transcription, in comparison to a healthy liver. In contrast, the treatment with IFC-305 triggers a ratio between the active histone mark H4ac and the repressive histone modification H3K27me3 (Fig. 4C) that matches with a permissive chromatin structure and correlates with an increase of transcription [Perez-Carreón et al., 2010] and protein levels of the *Pparg* gene (Fig. 3A). Importantly, our results are in agreement with a previously published report that described the enrichment of repressive histone marks (H3K27me2 and H3K9me3) on the *Pparg* promoter and gene in myofibroblast derived from *trans*-differentiated HSC, an in vitro cell system that is to a certain extent equivalent to cirrhosis [Mann et al., 2010].

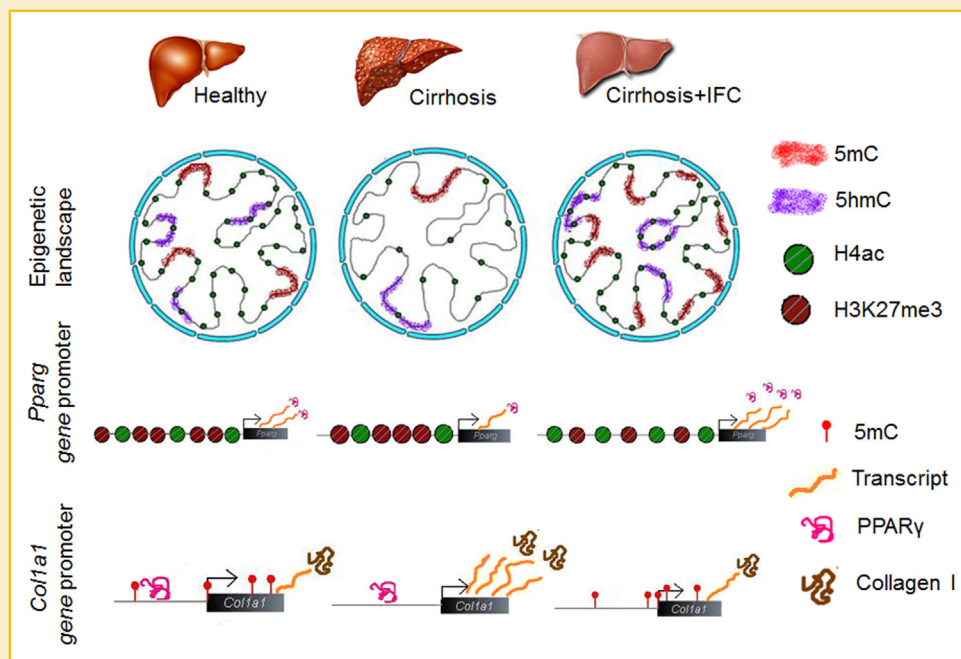
In addition, we evaluated the DNA methylation status of the *Pparg* gene promoter and found that this region is devoid of DNA methylation under all conditions of the CCl<sub>4</sub> model (Supplementary Fig. S5). This result goes hand in hand with the fact that H3K27me3 and DNA methylation are mutually exclusive marks on CpG islands [Brinkman et al., 2012]. We also addressed the PPAR $\gamma$  deposition on the *Colla1* gene promoter, where the nuclear receptor inhibits *trans*-activation of the fibrogenic gene in HSC. We found no changes in the deposition of PPAR $\gamma$  between healthy and cirrhotic rats. Interpreting this observation we have to take into account that the HSC population corresponds only to approximately 8%, while the hepatocytes are the most abundant cells in the liver. PPAR $\gamma$  positive immunoreactivity in response to IFC-305 treatment was detected in the latter cell type, as the signal was present (Fig. 3C) in non-septa regions (that were negative for the activated HSC marker  $\alpha$ SMA [Supplementary Fig. S3]). It is possible that the mechanism of *Pparg* gene regulation, involving DNA methylation, as well as the PPAR $\gamma$ -mediated regulation of *Colla1* are HSC-specific. Therefore, as the parenchymal cells outnumber the HSCs, mechanisms that take place only in HSCs are hidden, when analyzing the whole tissue. Our results suggest that on an epigenetic level transcription of the *Pparg* gene in CCl<sub>4</sub>-induced cirrhosis are mediated principally by histone PTMs. It is important to mention that the levels of PPAR $\gamma$  protein are similar to the control when CCl<sub>4</sub> is suspended (Fig. 3A: to +5ss), nevertheless the collagen quantity is equivalent to cirrhosis in this state (Supplementary Fig. S2: to +5ss). Maybe the PPAR $\gamma$  overproduction mediated by the IFC-305 (Fig. 3A: to +5IFC) is necessary to trigger the binding of the nuclear receptor to the *Colla1* gene promoter in activated HSCs to abrogate the *Colla1* overexpression. In this context, it has been described that PPAR $\gamma$  is necessary to revert activated HSCs to a quiescent state, and inhibit the expression of collagen I, TGF $\beta$ , and  $\alpha$ SMA in HSCs [Hazra et al., 2004b; Tsuchida and Friedman, 2017].

Concerning the DNA methylation pattern of the *Colla1* gene promoter, we identified four CpG dinucleotides which specifically decreased their DNA methylation level in cirrhosis (Fig. 5B). The difference in DNA methylation correlates with an over-expression of the transcript [Perez-Carreón et al., 2010] and an increment in the protein level of collagen I (Supplementary Fig. S2). It is noteworthy that CpG at position 1 is in close proximity to some

relevant regulatory sequences, like the NF-I binding site, and the relevant TGG-triplet that is critical for *Colla1* gene expression, which is abrogated when mutated to CCA (Fig. 5A: vertical arrow) [Yavrom et al., 2005]. At this CpG we identified the strongest reduction in DNA methylation in cirrhosis that was largely recovered by IFC-305 treatment (Fig. 5B; CpG 1). Furthermore, we want to point out that the administration of IFC-305 to cirrhotic rats triggers an increase of DNA methylation at the CpGs surrounding the *Colla1* gene TSS. Therefore, it would be interesting to investigate the dynamics of histone PTMs in this region to understand if they cooperate with DNA methylation to regulate *Colla1* gene expression.

In order to know if HSC is the cell type that is responsible for the changes described in the cirrhosis model, we determined the global levels of histone H4 acetylation. We observed that contrary to the cirrhosis state, which is characterized by a loss of histone H4 acetylation, there is a gain of this epigenetic modification in the activated state of HSCs (7d), (Supplementary Fig. S7). It has been described previously that the developmental cell specification is accompanied by a progressive chromatin restriction as the default state transitions from dynamic remodeling to generalized compaction [Zhu et al., 2013]. Since the activation process of HSCs can be considered a *trans*-differentiation process from a specialized cell to a myofibroblast, it is understandable that the chromatin of these cells acquires characteristics of decompaction, such as histone H4 acetylation. Previously, we showed that culturing HSCs in the presence of IFC-305 blocks their activation in vitro [Velasco-Loyden et al., 2010]. When we analyzed the histone H4 acetylation in this context, we noticed that the hepatoprotective drug retains the HSCs in an intermediate state with H4ac levels between the ones of a recently isolated cell (3 h) and an activated HSC (7 d). This suggests that the activation impairment mediated by IFC-305 is partially linked to chromatin modifications. Interestingly, the global changes described on whole tissue differed from the observations made in HSCs in vitro. The most likely explanation for this discrepancy is that data obtained from the whole liver tissue will primarily reflect changes that occur in the most abundant cell type, namely parenchymal cells, whereas HSCs only account for 8% of the total cell population in the liver. Thus, if IFC-305 is able to generate different chromatin changes depending on the metabolism characteristics of the cell type, it would be relevant to explore other functions of the adenosine derivative in different liver cell types.

In summary, we propose a model which suggests that changes in gene expression observed during treatment with the hepatoprotector IFC-305 are driven by epigenetic mechanisms, like DNA methylation and histone H4 acetylation, at a global scale, as well as in specific genes critical for cirrhosis development, such as *Pparg* and *Colla1* (Fig. 6). The model further proposes that chromatin modulation coordinated by the adenosine derivative plays a role in the prevention of HSC activation in vitro. These findings raise further questions concerning the molecular mechanisms involved in the regulation of other genes and hepatic functions altered during cirrhosis, which need to be addressed in future studies.



**Fig. 6.** Integrative model of the effects of the hepatoprotector IFC-305 in the CCl<sub>4</sub>-induced cirrhosis rat model. Top panel: Global epigenetic changes; in cirrhosis there is a decrease in global DNA methylation (5mC), 5hmC, and histone H4 acetylation levels; IFC-305 treatment increases DNA methylation and re-establishes the levels of 5hmC and histone H4 acetylation. Middle panel: Changes in the *Pparg* gene promoter; in cirrhosis, the chromatin configuration of the promoter is more compact than in the control liver and this correlates with decreased transcript and protein levels. In contrast, administration of the hepatoprotector IFC-305 triggers chromatin relaxation at the *Pparg* gene promoter, mediating transcript and protein over expression. Bottom panel: Changes in the *Col1a1* gene promoter; four methylated CpG dinucleotides in the healthy liver were found to lose DNA methylation in cirrhosis, this change correlates with an increment of *Col1a1* transcription and as a consequence generates fibrosis. Administration of IFC-305 induces an increase in CpG methylation around the TSS and at the CpG at position 1.

## CONCLUSIONS

This study provides a general description of dynamics of DNA methylation and histone PTMs (H4ac and H3K27me3) in a CCl<sub>4</sub>-induced cirrhosis animal model and the changes that occur upon treatment with the adenosine derivative IFC-305 (Fig. 6). We find a decrease of SAM levels, as well as global 5mC, 5hmC, and histone H4 acetylation modifications in cirrhosis. Treatment with IFC-305 restores levels of SAM, 5hmC, and histone H4 acetylation to physiological levels and increases the amount of 5mC. The hepatoprotector establishes a combination of histone PTMs at the *Pparg* gene promoter that triggers an increment of its transcription and protein levels. However, further studies are required in order to decipher the global dynamics of chromatin conformation in cirrhosis and upon administration of IFC-305. Finally, we propose that the reduction of collagen I, mediated by IFC-305, is induced by a recovery of DNA methylation in the *Col1a1* gene promoter, and this could be a key mechanism of action to reverse cirrhosis. Thus, the results of this study provide an important foundation to understand the fibrosis process at a molecular level.

## MATERIALS AND METHODS

### CHEMICALS

CCl<sub>4</sub> was purchased from Merk<sup>®</sup> México (México). IFC-305 is the aspartate salt of adenosine: 2-aminosuccinic acid-2-(6-amino-9H-

purin-9-yl)-5-(hydroxymethyl) tetrahydrofuran-3,4-diol (1:1). It was synthesized in the laboratory in agreement with UNAM patent 207422. Salts were from MP Biomedicals, LLC (Illkirch, France) and MP Biomedicals, Inc. (Eschwege, Germany).

### ANIMAL TREATMENT AND INDUCTION OF CIRRHOSIS WITH CCl<sub>4</sub>

Male Wistar rats (n = 20) weighing 100–110 g were rendered cirrhotic by chronic treatment with CCl<sub>4</sub>. Animals were intraperitoneally injected with a solution of 1:6 of CCl<sub>4</sub> in vegetable oil (0.4 g/kg weight) for 10 weeks. Cirrhosis-induced rats were divided in three groups, one was euthanized after CCl<sub>4</sub> withdrawal (Cirrhosis [to]), while other groups were treated with saline solution (to +5ss) or with IFC-305 (to +5IFC) for 5 weeks (50 mg/kg weight, three times/week), a group of non-treated rats was used as control and an additional group of healthy rats treated for 10 weeks with IFC-305 (10 IFC) were used in some assays to compare the effect of the drug with CCl<sub>4</sub> (Supplementary Fig. S1A). Animals were euthanized with sodium pentobarbital (Sedalpharma, Pets Pharma de México S.A. de C.V.). Liver was photographed (Supplementary Fig. S1B), recovered, weighed, and frozen or fixed with formaldehyde, and routinely processed for Masson Trichrome staining (Supplementary Fig. S2B). Animals were obtained from animal care facility of the Instituto de Fisiología Celular, UNAM. All procedures were conducted according to institutional guidelines for care and use of laboratory animals and were approved with protocol number VCHH53-14, valid to October 2019.

## SEROLOGICAL PARAMETERS OF LIVER FUNCTION

Alanine aminotransferase (ALT, EC 2.6.1.2) activity was determined as previously described [Hernandez-Munoz et al., 1984].

## S-ADENOSYLMETHIONINE QUANTIFICATION

SAM levels were determined as previously described [Hernandez-Munoz et al., 1984].

## TOTAL LIVER AND NUCLEAR PROTEINS ISOLATION

Liver tissue (100 mg) was manually homogenized in a Potter tube with a Teflon pistil with 1 mL PBS 1× and a protease inhibitors cocktail (Complete, Roche; PMSF, Sigma), incubated 20 min on ice, and centrifuged at 14,000 rpm, 4°C for 5 min (GS-15R Centrifuge, F2402H rotor, Beckman). The supernatant was recovered and centrifuged a second time. The nuclear fraction was obtained as previously described [Chagoya de Sanchez et al., 2012].

## WESTERN BLOT ANALYSIS

Total liver proteins or nuclear proteins were quantified with the Bradford reagent (Bio-Rad), and 30 µg of protein/well was run by electrophoresis in SDS-polyacrylamide gels (12% for total protein extracts and 20% for nuclear protein extracts) and transferred (1 h for total protein extracts and 30 min for nuclear protein extracts in humidity chamber) to an Immobilon<sup>®</sup>-P transfer membrane (Millipore). The blots were incubated with primary antibodies anti-collagen I (Santa Cruz Biotechnology sc-25974, Lot #F1207, RRID:AB\_2081880), anti-PPAR $\gamma$  (Santa Cruz Biotechnology sc-7196, Lot #E0611, RRID:AB\_654710), anti-GAPDH (Millipore MAB374, Lot #NG1722294, RRID:AB\_2107445), anti-HDAC3 (Abcam ab47237, Lot #GR29265-1, RRID:AB\_732779), anti-HDAC1 (Cell Signaling #5356, RRID:AB\_10612242, kindly donated by Dr. Alfonso León del Rio, IIB-UNAM) for total proteins and anti-hyperacetylated Histone H4 (Millipore 06-946, Lot #29860, RRID:AB\_310310) and anti-histone H4 (Millipore 04-858, Lot #1951953, RRID:AB\_1977264) for nuclei proteins and detected with the respective secondary antibody. Protein bands were visualized by using SuperSignal<sup>®</sup> West Femto (Thermo Scientific). Densitometric analyses of bands were done with Quantity One software (Bio-Rad<sup>®</sup>, Hercules, CA). The biological replicates of Western blots were performed independently and each series-assay was normalized to its own liver control and then compared between replicates.

## DOT-BLOT ANALYSIS

Genomic DNA was isolated from 200 mg of frozen rat liver [Strauss, 1998], then, genomic DNA was sonicated to obtain 100–12,000 kb DNA fragments and 400 ng were used per assay. For 5mC detection we used as positive control commercially available methylated-DNA (Zymo Research D5325-5-1) and for 5hmC we employed genomic DNA isolated from mice brain cortex. Plasmid DNA was used as negative control for both assays. Each DNA was hybridized onto a DNA hybridization membrane (HybondN+ GE) and then cross-linked using UV light [Valdes-Quezada et al., 2013]. The membrane was incubated overnight with antibody against 5mC (Zymo Research A3001-30, Lot #ZRC175458) or 5hmC (ActiveMotif<sup>®</sup> 39769, Lot #10310001, RRID:AB\_10013602) and detected with secondary antibody. DNA dots were visualized with SuperSignal<sup>®</sup> West Femto

(Thermo Scientific<sup>®</sup>). The membrane was methylene blue stained as loading control. Densitometric analyses of dots were done with the Quantity One software (Bio-Rad<sup>®</sup>).

## QUANTITATIVE EVALUATION OF GLOBAL GENOMIC DNA METHYLATION

Global DNA methylation was determined by using 100 ng of genomic DNA with the Imprint Methylated DNA Quantification Kit (Sigma<sup>®</sup> MDQ1) following the manufacturer's instruction.

## CHROMATIN IMMUNOPRECIPITATION ASSAY

Briefly, 200 mg of frozen liver tissue was fractionated and treated with 1% formaldehyde in PBS 1× and protease inhibitors in agitation for 10 min. The crosslinking was stopped with 125 mM glycine for 10 min, centrifuged 5 min at 5,000 rpm at 4°C (GS-15R Centrifuge, F2402H rotor, Beckman) [Valdes-Quezada et al., 2013]. The sonicated chromatin fragments were between 500 and 800 bp. The samples were diluted 1:5 in a dilution buffer and the diluted chromatin was pre-cleared [Valdes-Quezada et al., 2013]; 800 µg of chromatin were incubated with 4 µg of antibodies anti-hyperacetylated histone H4 (Millipore 06-946, Lot #29860, RRID:AB\_310310), anti-H3K27me3 (Abcam<sup>®</sup> AB6002, Lot #GR137554-2, RRID:AB\_305237), anti-PPAR $\gamma$  (Abcam<sup>®</sup> AB41928, Lot #GR26030-60, RRID:AB\_777392), normal rabbit anti-IgG (Millipore 12-370) and normal mouse anti-IgG (Millipore 12-371). The recovered DNA was assayed by qPCR (StepOne<sup>™</sup> detection system, Applied Biosystems) using the following primers: *Pparg* gene promoter previously described by Mann et al. [2010]; we used the *GAPDH* gene as a positive control of an open chromatin region, forward 5'-GGCACTGCACAAGAAGATG-3', reverse 5'-GTAAAGCCCGAGTAGCTG-3'; and a satellite repeated sequence as a positive control of a compacted chromatin region [Hernandez-Hernandez et al., 2008]; PPAR $\gamma$  response element in *aP2* gene, as positive control of PPAR $\gamma$  binding region, and *Col1a1* gene promoter were taken from Tsukamoto's group report [Yavrom et al., 2005]. For the open chromatin histone mark (acetylated histone H4), the enrichment was calculated over a rat satellite region. Inversely, for the repressive chromatin histone mark (H3K27me3), the enrichment was calculated over a constitutively open region corresponding to the *GAPDH* gene. For PPAR $\gamma$  the enrichment was calculated over GAPDH region, where the nuclear receptor is not bound. The enrichment calculations were performed as follows: Enrichment = antibody ( $C_{T_T} - C_{T_C}$ ) / Input 1% ( $C_{T_T} - C_{T_C}$ ); where  $C_{T_T}$  = amplification of the test region and  $C_{T_C}$  = amplification of the negative control region.

## SODIUM BISULFITE DNA CONVERSION AND SEQUENCING

Genomic DNA (3 µg) was processed with sodium bisulfite [Valdes-Quezada et al., 2013], DNA fragments of interest were PCR-amplified using the following primers: *Pparg* gene promoter and first exon 5'-end, forward 5'-GGTATGGGTATTTGTTGAGG-3', reverse 5'-ATCAAAAATAACTTCTCAACCC-3'; *Pparg* first exon, forward 5'-GGGTTGAGAAGTTATTTTTGAT-3', reverse 5'-TAAA AAAAACCTAAAATAACTCCAC-3'; *Col1a1* gene promoter, forward 5'-GAGGGTT TAGTTATATTAGTGAT-3', reverse 5'-CAAACTCCACTCTTAAATCTAC-3'. The amplified DNA fragments were cloned into

the pGEM-T Easy system (Promega<sup>®</sup>), and sequenced using SP6 sequence primers.

### HEPATIC STELLATE CELL ISOLATION AND CULTURE

HSCs were isolated from normal male Wistar rats (600–800 g body weight) by in situ enzymatic digestion of the liver with collagenase/pronase and density gradient ultracentrifugation with Accudenz<sup>®</sup> (Accurate Chemical, Westbury, NY) as previously described [Velasco-Loyden et al., 2010]. HSCs were cultured on plastic cell culture dishes in DMEM low glucose (LG) (Invitrogen, Carlsbad, CA) containing 10% fetal bovine serum (FBS), Antibiotic-Antimycotic, and 2 µg/mL gentamicin (Gibco, Life Technologies, Grand Island, NY) (DMEM LG 10% FBS AB/AM + genta). We seeded 250,000 cells (3 h, 1 day [1 d] and 7 days with IFC-305 5 mM [7 d + 5IFC]) or 125,000 cells (7 days [7 d]) by duplicate in a 12 wells plate. For 7 d and 7 d + 5IFC, after 24 h of isolation, the culture medium was replaced with DMEM LG 10% FBS AB/AM + genta, containing the indicated concentration of IFC-305. The medium was changed every other day for a total of 7 days.

### IMMUNOFLUORESCENCE

Histological sections of liver were used for PPAR $\gamma$  and  $\alpha$ -smooth muscle actin ( $\alpha$ SMA) immunofluorescence, with anti-PPAR $\gamma$  (Santa Cruz Biotechnology sc-7196, Lot #E0611, RRID:AB\_654710) diluted at 1:50, incubating overnight (4°C), with anti- $\alpha$ SMA (clone 1A4 Sigma Chemical Co. A 2547, Lot 035K4846, RRID:AB\_476701) diluted at 1:700 and secondary Alexa Fluor 568 goat anti-rabbit IgG (Invitrogen) antibody diluted 1:400 for anti-PPAR $\gamma$ , and secondary Fluorescein horse anti-mouse IgG (Vector Laboratories, Inc., Burlingame, CA) antibody diluted 1:400 for anti- $\alpha$ SMA, incubating for 2 h in obscurity at room temperature. We used the components of DAKO EnVision + System, HRP (DAB) (DAKO Corporation, Carpinteria, CA. K400711) to block the proteins and to dilute the antibodies. Coverslips were mounted on the histological sections with mounting medium for fluorescence with DAPI (Vectashield, Burlingame, CA). Fluorescence was visualized with an Olympus Inverted Microscope model IX71 and images captured with Evolution/QImaging Digital Camera (Media Cybernetics, Bethesda, MD). Data analysis was performed using ImageJ (National Institutes of Health, Bethesda, MD).

For cell cultures, once the growing times have elapsed, HSC were washed twice with PBS 1 $\times$ , fixed with 4% paraformaldehyde for 5 min, and washed three more times. The wells were blocked with block solution (1% Triton X-100 and albumin fraction V from bovine blood (USB Corporation, Cleveland, OH) in PBS 1 $\times$ ) for 20 min. The cells were washed once with block solution without Triton X-100. Cells were incubated overnight at 4°C with 1:400 primary antibody anti-hyperacetylated histone H4 in block solution without Triton X-100 (Millipore 06-946, Lot #29860, RRID:AB\_310310). Then, the cells were incubated 1 h at room temperature, washed three times, and then incubated with 1:400 anti-rabbit Alexa Fluor goat anti-rabbit IgG (Invitrogen) in PBS 1 $\times$  or block solution in accordance with primary antibody solution. Finally the cells were washed three times and a circle coverslip was mounted on the well with mounting medium for fluorescence with DAPI (Vectashield). Fluorescence was visualized with an Olympus Inverted Microscope model IX71 and images captured with Evolution/QImaging Digital Camera (Media

Cybernetics). Data analysis was performed using ImageJ (National Institutes of Health).

### STATISTICAL ANALYSIS

Statistical analysis was performed using the unpaired, Student's *t*-test, using Graph Pad Prism 5.0 (Graph Pad Software, Inc., La Jolla, CA) for Windows.

### AUTHORS' CONTRIBUTIONS

JRRA: Performed the great majority of experiments and prepared the manuscript. CAGH: Provided the conceptual framework and performed some experimental assistance. RPM: Provided experimental assistance performing sodium bisulfite DNA conversion, chromatin immunoprecipitation assay and comments on the manuscript. CECC: Provided assistance for dot blot assay and comments on the manuscript. RPCV: Was responsible for animal experimental procedures, immunohistochemistry, Western blot assay and comments on the manuscript. NGC: Contributed to some experiments and comments on the manuscript. MDL: Performed enzymatic assay determination, technical assistance in HPLC and comments on the manuscript. ARMO: Contributed to some experiments and comments on the manuscript. RAM: Contributed with bioinformatics analysis and made remarks on the manuscript. FRT: Contributed to design of epigenetic experiments, discussions of the results and comments on the manuscript. VCS: Provided insights and design of the project and prepared the manuscript with assistance from JRRA and FRT.

### ACKNOWLEDGMENTS

We acknowledge the technical assistance of Lucía Yáñez-Maldonado, Lidia Martínez-Pérez, Susana Vidrio-Gómez, Gabriela Velasco-Loyden, Georgina Guerrero-Avendaño, Fernando Suaste-Olmos, and the Molecular Biology Unit of IFC-UNAM. We thank Alberto Aranda-Fraustro and Olga Lidia Pérez Reyes from the Instituto Nacional de Cardiología "Ignacio Chávez" for obtaining the histological sections for this work. We wish to thank Emilio Rojas del Castillo, Mónica Liliana Guzmán-López, Karin Meier, and the members of the Recillas-Targa research group for their constant comments and suggestions. JRRA appreciates the article shared by Dr. James P. Hamilton. JRRA dedicates this work to his PhD advisor, Dr. Victoria Chagoya, within the framework of the celebration of the 50th years of her studies about adenosine and to his co-advisor Dr. Félix Recillas-Targa. This work has been supported by DGAPA-UNAM, grant numbers: IN225909 and IN207012 to VCS; IN203811 and IN201114 to FRT; CONACyT México, grant numbers: 128464 and 220503 to FRT; Fellowship: PAEP-UNAM, Programa de Maestría y Doctorado en Ciencias Bioquímicas, UNAM; No. Cta. 30479367-5 to JRRA; fellowship: PhD Fellowship from CONACyT; CVU 508509 to JRRA.

### REFERENCES

Bhutani N, Burns DM, Blau HM. 2011. DNA demethylation dynamics. *Cell* 146:866–872.



- Brinkman AB, Gu H, Bartels SJ, Zhang Y, Matarese F, Simmer F, Marks H, Bock C, Gnirke A, Meissner A, Stunnenberg HG. 2012. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res* 22:1128–1138.
- Chagoya de Sanchez V, Brunner A, Pina E. 1972. In vivo modification of the energy charge in the liver cell. *Biochem Biophys Res Commun* 46:1441–1445.
- Chagoya de Sanchez V, Hernandez-Munoz R, Sanchez L, Vidrio S, Yanez L, Suarez J. 1991. Twenty-four-hour changes of S-adenosylmethionine, S-adenosylhomocysteine adenosine and their metabolizing enzymes in rat liver; possible physiological significance in phospholipid methylation. *Int J Biochem* 23:1439–1443.
- Chagoya de Sanchez V, Martinez-Perez L, Hernandez-Munoz R, Velasco-Loyden G. 2012. Recovery of the cell cycle inhibition in CCl(4)-induced cirrhosis by the adenosine derivative IFC-305. *Int J Hepatol* 2012:212530.
- Garcea R, Daino L, Pascale R, Simile MM, Puddu M, Ruggiu ME, Seddaiu MA, Satta G, Sequenza MJ, Feo F. 1989. Protooncogene methylation and expression in regenerating liver and preneoplastic liver nodules induced in the rat by diethylnitrosamine: Effect of variations of S-adenosylmethionine: S-adenosylhomocysteine ratio. *Carcinogenesis* 10:1183–1192.
- Hamilton JP. 2011. Epigenetics: Principles and practice. *Dig Dis* 29:130–135.
- Hazra S, Miyahara T, Rippe RA, Tsukamoto H. 2004a. PPAR gamma and hepatic stellate cells. *Comp Hepatol* 3:57.
- Hazra S, Xiong S, Wang J, Rippe RA, Krishna V, Chatterjee K, Tsukamoto H. 2004b. Peroxisome proliferator-activated receptor gamma induces a phenotypic switch from activated to quiescent hepatic stellate cells. *J Biol Chem* 279:11392–11401.
- Hernandez-Gea V, Friedman SL. 2011. Pathogenesis of liver fibrosis. *Annu Rev Pathol* 6:425–456.
- Hernandez-Hernandez A, Rincon-Arano H, Recillas-Targa F, Ortiz R, Valdes-Quezada C, Echeverria OM, Benavente R, Vazquez-Nin GH. 2008. Differential distribution and association of repeat DNA sequences in the lateral element of the synaptonemal complex in rat spermatocytes. *Chromosoma* 117:77–87.
- Hernandez-Munoz R, Glender W, Diaz Munoz M, Adolfo J, Garcia-Sainz JA, Chagoya de Sanchez V. 1984. Effects of adenosine on liver cell damage induced by carbon tetrachloride. *Biochem Pharmacol* 33:2599–2604.
- Iredale JP. 2007. Models of liver fibrosis: Exploring the dynamic nature of inflammation and repair in a solid organ. *J Clin Invest* 117:539–548.
- Kohli RM, Zhang Y. 2013. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502:472–479.
- Komatsu Y, Waku T, Iwasaki N, Ono W, Yamaguchi C, Yanagisawa J. 2012. Global analysis of DNA methylation in early-stage liver fibrosis. *BMC Med Genomics* 5:5.
- Kriaucionis S, Heintz N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324:929–930.
- Mann J, Chu DC, Maxwell A, Oakley F, Zhu NL, Tsukamoto H, Mann DA. 2010. MeCP2 controls an epigenetic pathway that promotes myofibroblast transdifferentiation and fibrosis. *Gastroenterology* 138:705–714, e1–e4.
- Michalopoulos GK. 1990. Liver regeneration: Molecular mechanisms of growth control. *FASEB J* 4:176–187.
- Nehls MC, Rippe RA, Veloz L, Brenner DA. 1991. Transcription factors nuclear factor I and Sp1 interact with the murine collagen alpha 1 (I) promoter. *Mol Cell Biol* 11:4065–4073.
- Page A, Paoli P, Moran Salvador E, White S, French J, Mann J. 2016. Hepatic stellate cell transdifferentiation involves genome-wide remodeling of the DNA methylation landscape. *J Hepatol* 64(3):661–673.
- Paquet KJ, Kamphausen U. 1975. The carbon-tetrachloride-hepatotoxicity as a model of liver damage. First report: Long-time biochemical changes. *Acta Hepatogastroenterol (Stuttg)* 22:84–88.
- Penn NW, Suwalski R, O'Riley C, Bojanowski K, Yura R. 1972. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem J* 126:781–790.
- Perez-Carreón JI, Martínez-Pérez L, Loredó ML, Yanez-Maldonado L, Velasco-Loyden G, Vidrio-Gómez S, Ramírez-Salcedo J, Hernández-Luis F, Velázquez-Martínez I, Suárez-Cuenca JA, Hernández-Munoz R, de Sánchez VC. 2010. An adenosine derivative compound, IFC305, reverses fibrosis and alters gene expression in a pre-established CCl(4)-induced rat cirrhosis. *Int J Biochem Cell Biol* 42:287–296.
- Recillas-Targa F. 2014. Interdependency between genetic and epigenetic regulatory defects in cancer. *Methods Mol Biol* 1165:33–52.
- Rippe RA, Almounajed G, Brenner DA. 1995. Sp1 binding activity increases in activated Ito cells. *Hepatology* 22:241–251.
- Strauss WM. 1998. Preparation of genomic DNA from mammalian tissue. *Curr Protoc Mol Biol (Suppl42)*:2.2.1–2.2.3.
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324:930–935.
- Tsochatzis EA, Bosch J, Burroughs AK. 2014. Liver cirrhosis. *Lancet* 383:1749–1761.
- Tsuchida T, Friedman SL. 2017. Mechanisms of hepatic stellate cell activation. *Nat Rev Gastroenterol Hepatol* <https://doi.org/10.1038/nrgastro.2017.38> [Epub ahead of print].
- Valdes-Quezada C, Arriaga-Canon C, Fonseca-Guzman Y, Guerrero G, Recillas-Targa F. 2013. CTCF demarcates chicken embryonic alpha-globin gene autonomous silencing and contributes to adult stage-specific gene expression. *Epigenetics* 8:827–838.
- Varela-Moreiras G, Alonso-Aperte E, Rubio M, Gasso M, Deulofeu R, Alvarez L, Caballeria J, Rodes J, Mato JM. 1995. Carbon tetrachloride-induced hepatic injury is associated with global DNA hypomethylation and homocysteinemia: Effect of S-adenosylmethionine treatment. *Hepatology* 22:1310–1315.
- Velasco-Loyden G, Perez-Carreón JI, Aguero JF, Romero PC, Vidrio-Gomez S, Martinez-Perez L, Yanez-Maldonado L, Hernandez-Munoz R, Macias-Silva M, de Sanchez VC. 2010. Prevention of in vitro hepatic stellate cells activation by the adenosine derivative compound IFC305. *Biochem Pharmacol* 80:1690–1699.
- Yang H, Liu Y, Bai F, Zhang JY, Ma SH, Liu J, Xu ZD, Zhu HG, Ling ZQ, Ye D, Guan KL, Xiong Y. 2013. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* 32:663–669.
- Yavrom S, Chen L, Xiong S, Wang J, Rippe RA, Tsukamoto H. 2005. Peroxisome proliferator-activated receptor gamma suppresses proximal alpha 1(I) collagen promoter via inhibition of p300-facilitated NF-I binding to DNA in hepatic stellate cells. *J Biol Chem* 280:40650–40659.
- Zhu J, Adli M, Zou JY, Verstappen G, Coyne M, Zhang X, Durham T, Miri M, Deshpande V, De Jager PL, Bennett DA, Houmard JA, Muoio DM, Onder TT, Camahort R, Cowan CA, Meissner A, Epstein CB, Shores N, Bernstein BE. 2013. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* 152:642–654.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

RESEARCH ARTICLE

Open Access



# Epigenetic silencing of miR-181c by DNA methylation in glioblastoma cell lines

Erandi Ayala-Ortega, Rodrigo Arzate-Mejía, Rosario Pérez-Molina, Edgar González-Buendía, Karin Meier, Georgina Guerrero and Félix Recillas-Targa\*

## Abstract

**Background:** Post-transcriptional regulation by microRNAs is recognized as one of the major pathways for the control of cellular homeostasis. Less well understood is the transcriptional and epigenetic regulation of genes encoding microRNAs. In the present study we addressed the epigenetic regulation of the *miR-181c* in normal and malignant brain cells.

**Methods:** To explore the epigenetic regulation of the *miR-181c* we evaluated its expression using RT-qPCR and the *in vivo* binding of the CCCTC-binding factor (CTCF) to its regulatory region in different glioblastoma cell lines. DNA methylation survey, chromatin immunoprecipitation and RNA interference assays were used to assess the role of CTCF in the *miR-181c* epigenetic silencing.

**Results:** We found that *miR-181c* is downregulated in glioblastoma cell lines, as compared to normal brain tissues. Loss of expression correlated with a notorious gain of DNA methylation at the *miR-181c* promoter region and the dissociation of the multifunctional nuclear factor CTCF. Taking advantage of the genomic distribution of CTCF in different cell types we propose that CTCF has a local and cell type specific regulatory role over the *miR-181c* and not an architectural one through chromatin loop formation. This is supported by the depletion of CTCF in glioblastoma cells affecting the expression levels of *NOTCH2* as a target of *miR-181c*.

**Conclusion:** Together, our results point to the epigenetic role of CTCF in the regulation of microRNAs implicated in tumorigenesis.

**Keywords:** Glioblastoma cells, CCCTC-binding factor (CTCF), DNA methylation, RNA interference, Epigenetics

## Background

MicroRNAs (miRNAs) are small non-coding RNAs that participate in the control of many cellular processes such as stress response, cell differentiation, cell-cycle regulation, stem cell biology, apoptosis among many others [1]. MicroRNAs exert their regulatory effect post-transcriptionally by inducing RNA degradation or translation inhibition, and their expression can be deregulated in cancer by genetic and epigenetic mechanisms [2–4]. MicroRNAs can also affect gene expression of many genes by direct regulation of the epigenetic machinery. For example, microRNAs like *miR-101*, *miR-205* and *miR-26a* regulate chromatin modifiers in cancer such as the Polycomb associated

histone methyltransferase EZH2 [2, 3]. The DNA methylation maintenance enzyme Dnmt1 is regulated in different cell- types by the *miR-126* and *miR-152*, as well as the *de novo* methyltransferases Dnmt3a and Dnmt3b by the *miR-29* family members *miR-29a*, *-29b* and *-29c* [5]. Overexpression of *miR-29a*, *-29b* and *-29c* cause abnormal downregulation of the Dnmt3a and Dnmt3b, which is associated with development of lung cancer and acute myeloid leukemia [6, 7].

DNA methylation can regulate microRNAs gene expression in cancer [8]. In particular, repression of gene expression by DNA methylation of promoter associated CpG islands has been reported for several microRNAs in glioblastoma cells like *miR-211*, *miR-204*, *miR-145*, *miR-137* among others [9–12]. For example, *miR-145* was shown to be downregulated in glioblastoma cells and low expression of *miR-145* was found to

\* Correspondence: frecilla@ifc.unam.mx  
Instituto de Fisiología Celular, Departamento de Genética Molecular,  
Universidad Nacional Autónoma de México, Ciudad de México, México

be correlated with poor prognosis in patients [11]. Overexpression of *miR-145* reduced cell proliferation, migration and invasion in glioblastoma cells by suppressing SOX9 and ADD3 [13]. Thus, DNA methylation of CpG-rich microRNAs promoters in glioblastoma cells seems to be an important process for tumour development and maintenance.

CTCF is a ubiquitous, highly-conserved 11-zinc finger nuclear protein [14, 15], which is subjected to different post-translational modifications [16, 17] and has been implicated in a broad range of functions including higher-order chromatin organization by favoring inter- and intra-chromosomal interactions [18–20]. The combinatorial usage of different zinc-fingers confers CTCF the capacity to bind complex sequences, interact with other proteins and with ncRNAs [14, 21–23]. CTCF is also important to maintain, CpG-rich promoter regions of tumour suppressor genes, like *BRCA1*, *retinoblastoma*, and others, in an unmethylated state [24, 25]. Importantly, DNA methylation can affect CTCF binding in part because of the presence of CpGs in the CTCF binding motif [26]. For example, increased methylation at the promoter of the brain-derived neurotrophic factor (*BDNF*) triggered the dissociation of CTCF which resulted in gene silencing [27]. In fact 41 % of cell-type specific CTCF binding sites show differential DNA methylation [28].

In addition, several reports have implicated CTCF in the regulation of microRNAs expression [29]. *MiR-125b* expression is decreased in breast cancer, partly, through CTCF dissociation from its promoter region [30]. In addition, ER $\alpha$  positive breast cancer cells overexpress *miR-375* concomitantly with promoter DNA hypermethylation and CTCF depletion [31]. Furthermore, CTCF and pluripotency maintenance factors are depleted in the *miR-290* regulatory region in differentiated embryonic stem cells, together with increased DNA methylation and deposition of the repressive histone mark H3K27me3 [32].

The *miR-181c* is a member of the miR-181 family of microRNAs involved in the development of glioblastoma multiforme (GBM), which is one of the most frequent and malignant primary brain tumours [33, 34]. *MiR-181c* is downregulated in GBM, and its expression levels correlate with tumour progression, suggesting that its epigenetic regulation could be affected [33]. In contrast, *miR-181c* is overexpressed in gastric cancer, skin basal cell carcinoma, and in osteosarcomas [35–37].

Here we explored the epigenetic regulatory processes responsible for the deregulation of *miR-181c* in glioblastoma cells; in particular, we asked whether the nuclear factor CTCF participates in its epigenetic regulation. We first confirmed that *miR-181c* is differentially expressed in glioblastoma cell lines. We analyzed ChIP-seq data

sets from different cell-types and identified a DNA region located in the 5' non-coding region of the *miR-181c* enriched in histone marks characteristic of promoter regions. We confirmed binding of CTCF to the promoter region of *miR-181c* in the glioblastoma cell line U87MG and K562 cells. In contrast, CTCF does not bind the promoter region of the aggressive glioblastoma cell line T98G. Absence of CTCF correlates with gain of DNA methylation and *miR-181c* downregulation. Furthermore, we show that depletion of CTCF in glioblastoma cells affects the expression levels of *NOTCH2* a target of *miR-181c*. Together, these results implicate CTCF and DNA methylation in the epigenetic regulation of *miR-181c* in cancer cells.

## Methods

### Cell culture

K562 human erythroleukemic cells were cultured in ISCOVE medium (Invitrogen). K562 cells (K562 ATCC<sup>®</sup> CCL-243<sup>™</sup>) were provided by Gary Felsenfeld (National Institutes of Health, Bethesda, Maryland, US); human glioblastoma-astrocytoma grade IV U87MG cells (U87MG ATCC<sup>®</sup> HTB-14<sup>™</sup>), human glioblastoma multiforme T98G cells (T98G ATCC<sup>®</sup> CRL-1690<sup>™</sup>) and human acute T cell leukemia Jurkat cells (Jurkat ATCC<sup>®</sup> TIB-152<sup>™</sup>) were cultured in RPMI-1640 medium (Invitrogen); all media contained 10 % (v/v) fetal bovine serum (FBS) and 1 % penicillin/streptomycin. T98G, U87MG and Jurkat cells were provided by Manel Esteller (Centro Nacional de Investigaciones Oncológicas (CNIO) and Cancer Epigenetics and Biology Program (PEBC), Spain). All cell lines were purchased from the American Type Culture Collection (Manassas, VA) and were previously authenticated by STR profiling. Cells were maintained at 37 °C in a humidified 5 % CO<sub>2</sub>-containing atmosphere. Human lymphocytes were obtained from peripheral blood of a healthy donor, isolated with Ficoll-Paque Plus (Amersham) following the manufacturer's instructions. Written informed consent was obtained from this healthy donor.

### Quantitative real time PCR

Total RNA from Human Hypothalamus and Orbital Frontal Cortex were purchased from Ambion (First Choice<sup>®</sup> Total RNA AM6786 and AM6864). Total RNA was extracted from lymphocytes, K562, Jurkat, U87MG and T98G cells with TRIzol Reagent (Invitrogen) according to manufacturer's instructions. RNAs were treated with DNase I (RQ1, Promega) followed by Random Primer cDNA generation from 1  $\mu$ g DNase I treated RNA (Reverse Transcription System, Promega). Real-Time quantitative PCR (qPCR) was carried out with SYBR Green (Sigma) and specific primers for *primiR-181c* (Forward: 5'-CCCATCTCAGCCTCCTAAGT-3' and Reverse:

5'-GACCAACCTGAGCAACATAG-3'), *NOTCH2* (Forward: 5'-CCTTCCACTGTGAGTGTCTGA-3' and Reverse: 5'-AGGTAGCATCATTCTGGCAGG-3') and *GAPDH* as an endogenous normalization control (Forward: 5'-CCACTCCTCCACCTTTGAC-3' and Reverse: 5'-ACCCTGTTGCTGTAGCCA-3'). In order to analyze *miR-181c* mature transcript levels, first strand cDNA was generated using Taqman® MicroRNA Reverse Transcription Assay (Applied Biosystems) with specific primers provided by the manufacturer and U6 RNA was used as an endogenous normalization control. *MiR-181c* mature transcript levels were measured with Taqman® MicroRNA Assay primers (Applied Biosystems). The qPCR reactions were carried out in the StepOne detection system (Applied Biosystems) at 95 °C for 2 min, followed by 40 two-step cycles of 95 °C for 30 s and 60 °C for 45 s, triplicates were made for each sample. Relative RNA levels were calculated using the comparative  $\Delta\Delta C_t$  method. Significant differences on gene expression were evaluated by a *t*-Student test.

#### DNA sodium bisulfite conversion

Genomic DNA was extracted from indicated cells by phenol-chloroform technique, and 1.5  $\mu$ g were cut with HindIII previous to bisulfite conversion. Bisulfite conversion was performed as described previously [38]. Specific primers for converted promoter region were used to generate PCR product (Forward: 5'-GTTTTAGATAGAGGGGTGGG-3' and Reverse: 5'-CAATCCTCAAAAAA CCAACTC-3'). PCR products were cloned in pGEMT-easy (Promega) followed by sequencing with Sp6 primer. Culture recuperation after transformation for plasmid enrichment was carried out at 30 °C to avoid recombination as much as possible.

#### Chromatin immunoprecipitation assay

The ChIP assay was performed as previously reported with 4  $\mu$ g of antibody against CTCF (Millipore 07-729) [38]. Immunoprecipitated DNA was evaluated by PCR using specific amplification primers for CTCF downstream (Forward: 5'-GTCTCAACTTCTGGGCTCC-3' and Reverse: 5'-GAAGAGAAATAGGCGGTGG-3'), Upstream (Forward: 5'-CTCCCATCTCAGCCTCCTA-3' and Reverse: 5'-CAAGCCAAGCAGTGACGAC-3') regions, and *Igf2/H19* Differential Methylation Region (DMR) as a positive control (Forward: 5'-CAGGCTCC CCAAAATCTA-3' and Reverse: 5'-GGGAACATAG AGAAAGAGG-3').

#### CTCF knockdown with lentivirus expressing shRNAi

CTCF knockdown was performed essentially as described [39]. HEK293FT cells were used to produce pLL3.7 control and CTCF shRNAi (5'-GGACAGTGT TGACAACACTAA-3') lentiviruses with generation III

packaging vectors. pLL3.7 and shRNAi CTCF plasmid were kindly provided by Joaquín Espinosa [39]. For the tetracycline inducible system the pTRIPZ lentiviral vector (Open Systems) was used with shRNAi (5'-AGGAC AGTGTTGACAACACTA-3') targeting CTCF. U87MG cells were transduced with virus for 8 h in the presence of polybrene (8  $\mu$ g/ml; Sigma). Cultures were then selected for 3–4 days with puromycin (5  $\mu$ g/ml; Sigma) and then harvested for the experiments detailed in this article. Doxycycline induction was carried out with 2  $\mu$ g/ml for 72 h, and cells were harvested for the corresponding experiments.

#### Bioinformatic analysis

All ChIP-seq and RRBS data was downloaded from the Analysis/Data hub from the ENCODE project (<https://genome.ucsc.edu/ENCODE/analysis.html>) and displayed on the IGV genome browser (<https://www.broadinstitute.org/igv/node/250>). CTCF Motif analysis was performed with JASPAR using the human motif as query (<http://jaspar.genereg.net/>). CpG islands were downloaded from the UCSC genome browser hg19 (<https://genome.ucsc.edu/>). *In situ* Hi-C data from GM12878 cell line at 5 kb resolution was analyzed by using the JuiceBox software (<http://www.aidenlab.org/juicebox/>).

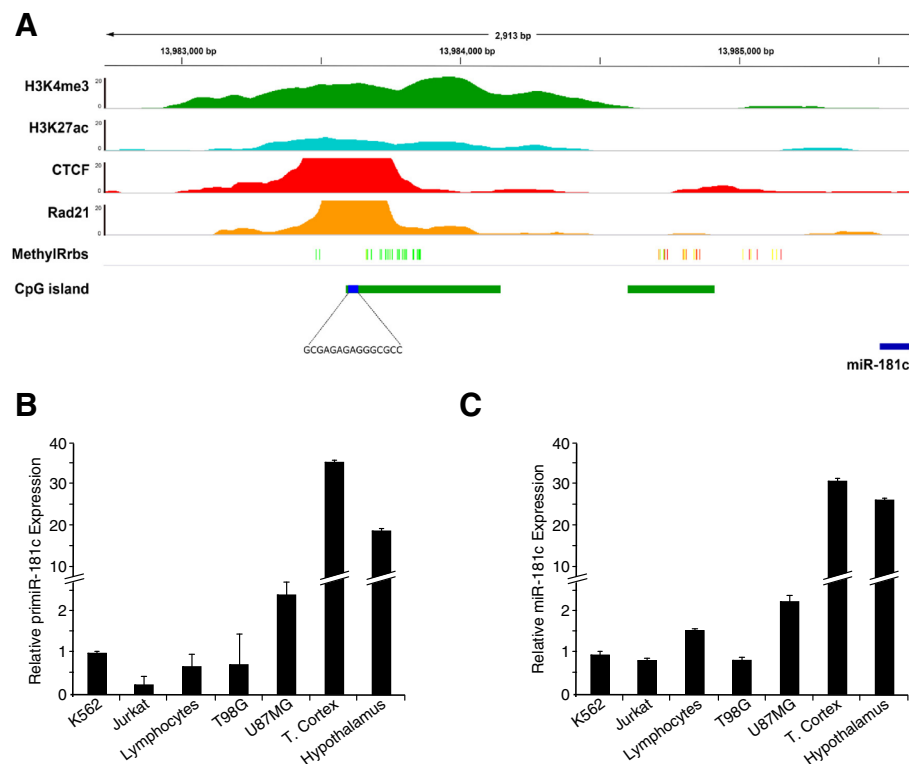
## Results

### Differential expression of miR-181c in brain and human glioblastoma cells

The human *miR-181c* is frequently downregulated in Glioblastoma Multiforme (GBM) and its downregulation has been linked to tumour progression [33]. However, the mechanisms controlling its expression are unknown. To identify the regulatory region of *miR-181c* we analyzed ChIP-seq data for the promoter-associated histone marks H3K4me3 and H3K27ac generated by ENCODE in the erythroleukemic K562 cell line. We identified a DNA region occupied by H3K4me3 and H3K27ac located 2 k bases (kb) upstream of the sequence corresponding to the mature *miR-181c* (Fig. 1a and Additional file 1: Figure S1). This DNA region was previously reported to act as a promoter of *miR-181c* [40]. The identified region overlaps with a CpG Island of 0.5 kb containing a CTCF binding motif. Indeed, ChIP-seq data shows that CTCF binds the promoter region (Fig. 1a). Since CTCF binds to a CpG rich region in the promoter of *miR-181c* we hypothesized that DNA methylation and CTCF could be critical regulators of *miR-181c* expression in glioblastoma.

As a first step to uncover the mechanisms controlling *miR-181c* we first evaluated transcript levels for the *pri-miR-181c* and mature *miR-181c* by RT-qPCR in two glioblastoma cell lines T98G and U87MG, erythroleukemic K562 cells, lymphoblastic Jurkat cells, peripheral





**Fig. 1** CTCF binds to the promoter of *miR-181c*. **a** IGV genome browser screenshot for ChIP-seq data of H3K4me3, H3K27ac, CTCF and Rad21 from K562 cells and Reduced Representation Bisulfite Sequencing (RRBS) data from the same cell line. Green bars, 0 % molecules sequenced are methylated; Yellow bars, 50 % molecules sequenced are methylated; Red bars, 100 % molecules sequenced are methylated. CTCF binding motif with the highest score is shown with reference to one CpG island. The region depicted is chr19:13,982,729-13,985,645. Data was downloaded from the Analysis/Data hub by the ENCODE project. **b** *primier-181c* expression levels in different cells measured by RT-qPCR with SYBR Green. **c** *miR-181c* expression levels in different cells measured by Taqman assay

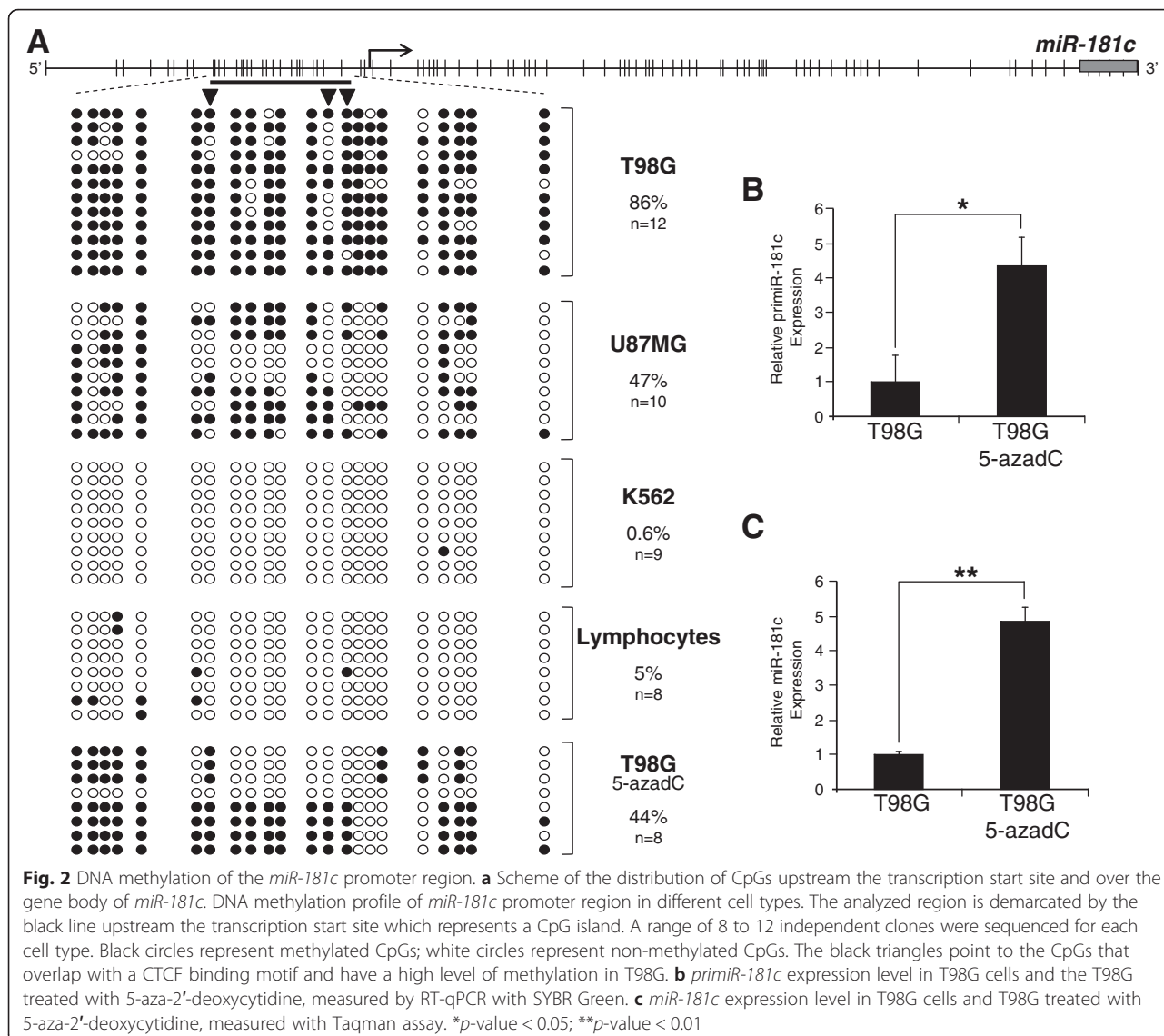
blood lymphocytes, frontal cortex and hypothalamus-derived primary cells (Fig. 1b and c). The highest level of expression of the *primier-181c* and the mature *miR-181c* transcript were found in cells from the frontal cortex and hypothalamus, which is consistent with previous reports showing that *miR-181c*, is mainly expressed in brain cells in human, mouse and rat (Expression Atlas EMBL-EBI). Intermediate levels of expression were found in U87MG glioblastoma cell line, and very low levels in the aggressive glioblastoma cell line T98G and the rest of the analyzed cells (Fig. 1b and c). Thus *miR-181c* is expressed at low levels in glioblastoma cell lines as compared with brain primary cells.

#### Low levels of *miR-181c* in glioblastoma cells correlate with DNA hypermethylation of its promoter region

In order to characterize the DNA methylation profile of the *miR-181c* promoter region we performed DNA bisulfite conversion coupled to sequencing in T98G and U87MG glioblastoma cell lines, K562 cells and primary lymphocytes (Fig. 2a). The highest DNA methylation levels of the *miR-181c* promoter, with 86 % of methylated CpGs, were found in T98G (Fig. 2a). Intermediate

DNA methylation, with 47 % of methylated CpGs, was found in U87MG. The promoter region was almost unmethylated, with 0.6 and 5 % of methylated CpGs, in K562 cells and lymphocytes, respectively (Fig. 2a). Hypermethylation of the *miR-181c* promoter region correlates with the low level of transcript detected in T98G cells. A 2-fold increase in the expression of the *primier-181c* in U87MG cells, as compared with T98G cells, correlates with a 50 % reduction in the methylation of the promoter. Low levels of expression of *miR-181c* and *primier-181c* do not correlate with absence of DNA methylation in K562 cells and lymphocytes. This is probably due to the tissue-specific expression of the *miR-181c* and the absence in K562 cell of a particular set of transcription factors and co-factors needed for *miR-181c* gene transcription.

DNA hypermethylation of promoter regions of microRNAs has been linked to transcriptional repression [8]. To determine if the hypermethylation of the promoter region of *miR-181c* cells promotes transcriptional repression, we treated T98G cells with the DNA methylation inhibitor 5-aza-2'-deoxycytidine (5-azadC) for 72 h and then analyzed the expression of the primary

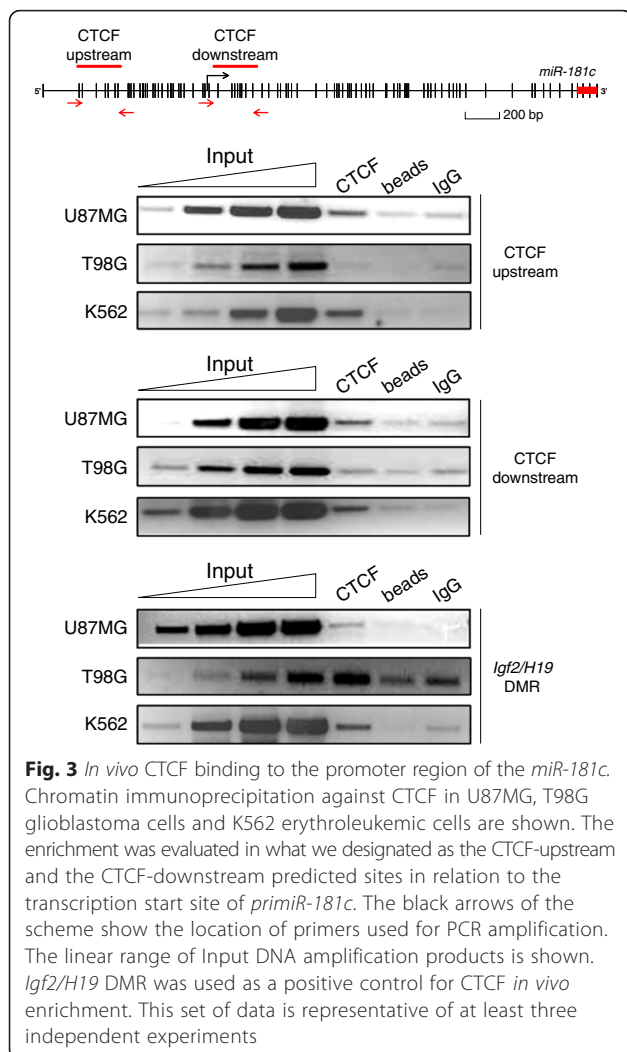


*primiRNA-181c* and mature *miR-181c*. To confirm these results we performed DNA bisulfite genomic DNA conversion and sequencing in T98G genomic DNA previously treated with 5-azadC. More than 40 % of the CpGs are demethylated, and of note, the CTCF binding site seems preferentially unmethylated (Fig. 2a). In line with this, we found that the expression levels of *miR-181c* increased after treatment (Fig. 2b and c), suggesting silencing of the *miR-181c* by DNA methylation in glioblastoma cells.

**Interaction of CTCF with the regulatory region of *miR-181c***

The region of the *miR-181c* occupied by CTCF in K562 cells as indicated by ENCODE data, spans the 7<sup>th</sup>, 13<sup>th</sup> and 14<sup>th</sup> CpGs of the analyzed CpG-island (Fig. 2a). Importantly, methylation levels of these CpGs are different

between T98G and U87MG cells (Fig. 2a arrowheads). In particular the 7<sup>th</sup> and 14<sup>th</sup> CpGs show a 45 and 61.6 % increase in DNA methylation, in T98G as compared to U87MG, opening the possibility that DNA methylation may affect CTCF binding in this region. Analysis of ChIP-seq data from ENCODE shows interaction of CTCF with the promoter region of *miR-181c* in 23 of the 47 cell lines analyzed, suggesting cell-type specific binding of CTCF to this region. However, if CTCF interacts with the *miR-181c* promoter in the glioblastoma cell lines used in this study is not known (Additional file 1: Figure S1 and Table S1). To determine if CTCF interacts with the *miR-181c* promoter in T98G and U87MG cells we performed chromatin immunoprecipitation (ChIP) (Fig. 3). As a positive control we looked at CTCF enrichment on the human *Igf2/H19* Differential Methylation Region (DMR) in K562 cells



[26, 41]. We found that CTCF interacts with the *miR-181c* promoter in U87MG, but not in T98G cells (Fig. 3). Motif analysis suggests that additional CTCF binding sites are present over the gene body of *miR-181c*, however, they are not bound by CTCF in U87MG, T98G and K562 cells, as revealed by ChIP assays (Additional file 1: Figure S2).

The interaction of CTCF with the *miR-181c* promoter correlates with moderate gene expression in U87MG cells. Absence of CTCF interaction with the *miR-181c* promoter correlates with DNA hypermethylation and very low expression levels in T98G cells. Thus, CTCF may be associated with expression regulation and protection against DNA methylation of the *miR-181c* promoter in U87MG cells.

#### CTCF occupancy correlates with *miR-181c* expression

To further characterize the contribution of CTCF to *miR-181c* regulation, CTCF was knocked down in U87MG cells by transduction with a doxycycline inducible lentivirus

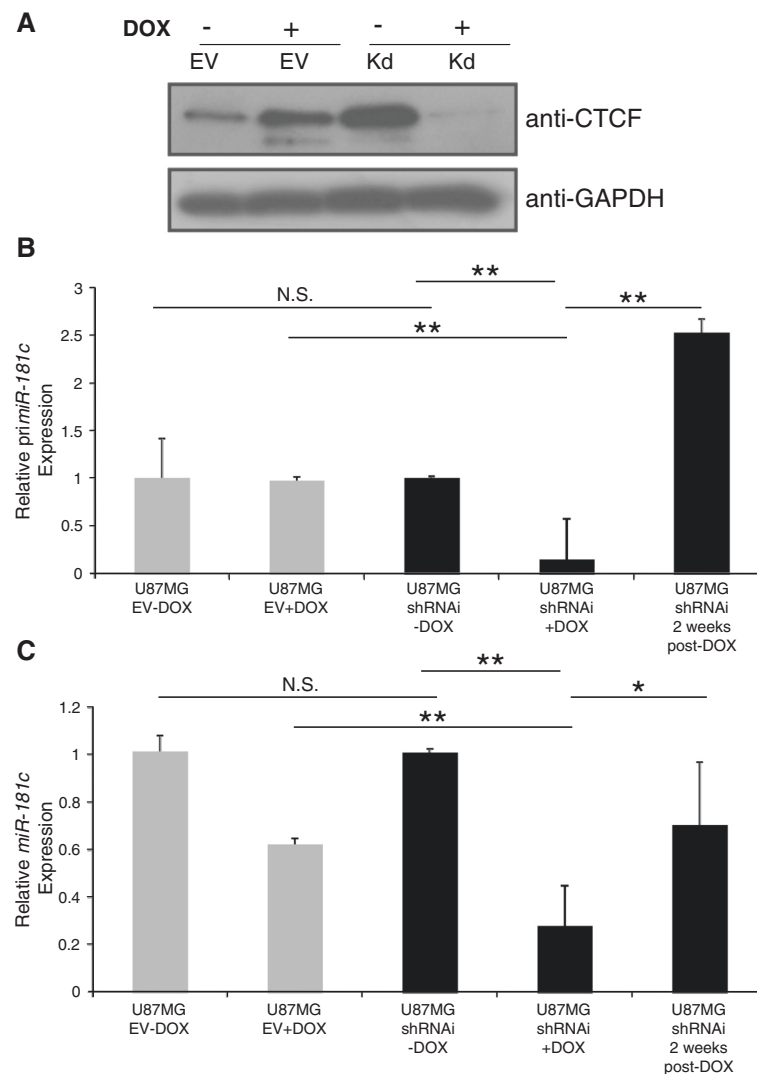
containing small-hairpin interference RNA (shRNAi) against CTCF (Fig. 4 and Additional file 1: Figure S3). Cells were treated with doxycycline or vehicle for 72 h and the expression levels of *primiR-181c* and mature *miR-181c* were assessed by RT-qPCR. *MiR-181c* was significantly downregulated upon CTCF knockdown (Fig. 4b and c). Doxycycline withdrawal for two weeks resulted in upregulation of *miR-181c* (Fig. 4b and c). This data suggests that CTCF promotes the expression of *miR-181c* in U87MG cells.

#### CTCF depletion results in increased promoter methylation and decreased expression of *miR-181c*

To test the function of CTCF in the protection against DNA methylation of the *miR-181c* promoter we knocked down CTCF in U87MG cells and assessed the level of DNA methylation (Fig. 5a). After 5 days of transduction with the CTCF shRNAi we observed a 20 % increase in the DNA methylation level of the *miR-181c* promoter region (Fig. 5b). This increase in DNA methylation was accompanied with a reduction of *primiR-181c* and *miR-181c* (Fig. 5c and d). These results suggest that CTCF could protect the *miR-181c* promoter from DNA methylation in U87MG cells.

#### *miR-181c* is flanked by two chromatin loops in GM12878 cells

The three-dimensional organization of the genome is critical to establish proper programs of gene expression through the formation of chromatin loops that bring together distal regulatory regions [42]. CTCF is a key mediator of chromatin looping, and novel techniques like *in situ* Hi-C coupled with deep sequencing allow the identification of all long-range chromatin interactions in a given cell-type [43]. To gain insight on whether the CTCF binding site in the promoter region of *miR-181c* is implicated in loop formation we took advantage of published data of high resolution *in situ* Hi-C generated in GM12878 cells [43]. The *in situ* Hi-C data set represents the highest resolution (1 kb) map of chromatin interactions ever published and identified chromatin loops at a genome wide scale. *In situ* Hi-C data for GM12878 cells suggest the presence of two chromatin loops (100 kb and 200 kb in size, respectively) flanking the *miR-181c* locus and part of the *Nanos3* locus (Fig. 6a). The anchor sites for chromatin loops frequently overlap with binding sites for CTCF in convergent orientation (92 %) [43]. In fact, the anchor sites for the two chromatin loops flanking the *miR-181c* locus correspond to constitutive binding sites for CTCF in convergent orientation (Fig. 6b). In contrast, the CTCF binding site on the promoter region of *miR-181c*, which is occupied only in a subset of cell lines, is not involved in chromatin looping (Fig. 6b). Thus, the CTCF binding site in the



**Fig. 4** Inducible knockdown of human CTCF in U87MG glioblastoma cell line affects *miR-181c* expression. **a** Western blot shows CTCF protein levels in U87MG cells transduced with an inducible Empty Vector (EV) with and without Doxycycline (DOX) induction as controls. U87MG cells were also transduced with an inducible shRNAi against CTCF without (-DOX) and with (+DOX) Doxycycline (DOX). **b** *primiR-181c* expression levels in cell pools containing the inducible shRNAi against CTCF. No treatment (U87MG/shRNAi), treatment (U87MG/shRNAi+DOX) and 2 weeks after Doxycycline deprivation (U87MG/shRNAi/2 weeks/post-DOX) were measured by RT-qPCR with SYBR Green. Empty vector controls are shown (U87MG/EV-DOX and U87MG/EV + DOX). N.S., not significant. **c** The *miR-181c* expression levels were evaluated under the same experimental conditions as in **(b)** using the Taqman assay. \**p*-value < 0.05 and \*\**p*-value < 0.01

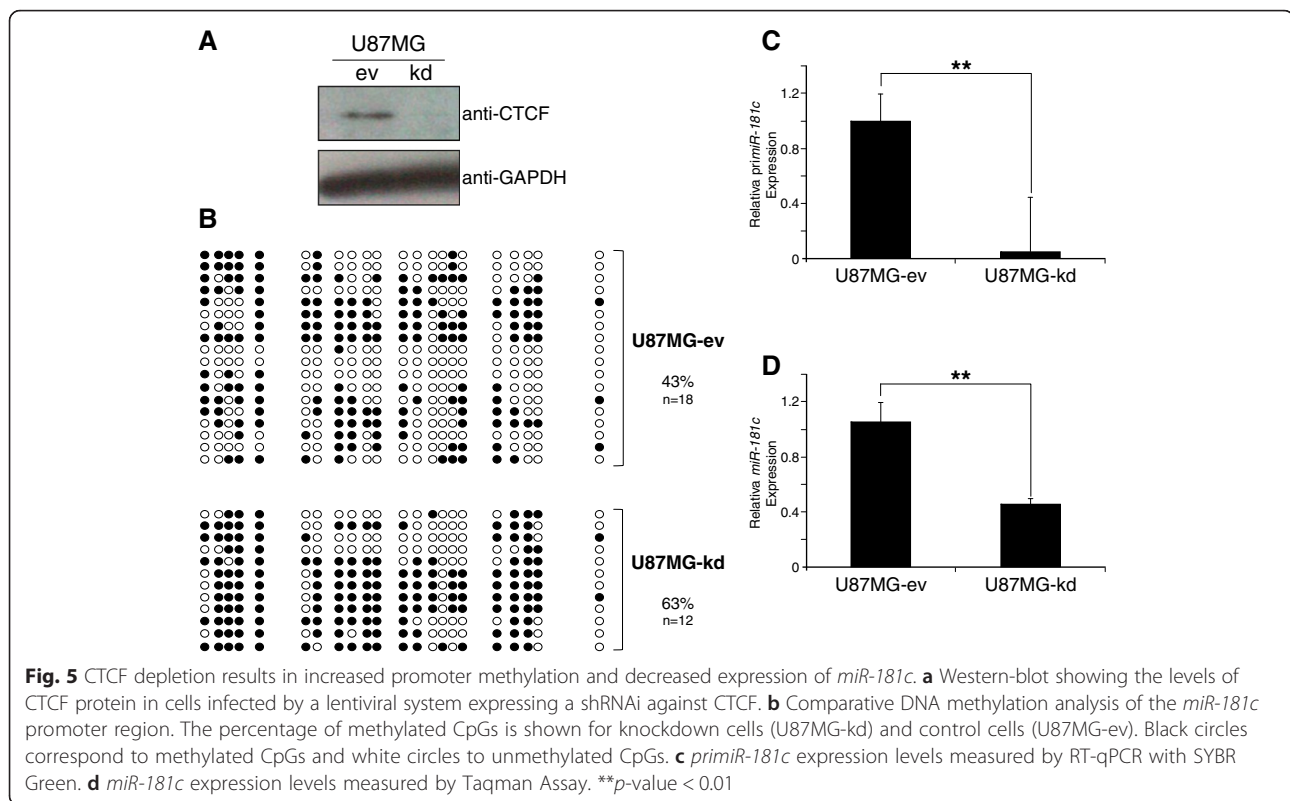
promoter region of *miR-181c* protects against DNA methylation, and we speculate that this particular site does not participate in chromatin looping.

#### Depletion of CTCF in glioblastoma cells affects the expression levels of *NOTCH2*

Our results suggest that CTCF participates in the transcriptional regulation of *miR-181c* by protecting its promoter against silencing by DNA methylation in U87MG cells. Therefore, we asked if reduced levels of CTCF could affect the transcript levels of *miR-181c*

targets like *NOTCH2* [34]. We infected the glioblastoma U87MG cell line with a lentivirus expressing a doxycycline inducible shRNAi against human CTCF. Quantitative RT-PCR of *NOTCH2* was performed after 3 and 30 days of induction with doxycycline. Knockdown of CTCF at 3 days after induction results in an increase of *NOTCH2* mRNA (Fig. 7a). This trend is more evident with cells that have been on doxycycline during 30 days. Therefore CTCF loss causes the epigenetic silencing of *miR-181c* by DNA methylation and the inability of the *miR-181c* to diminish the





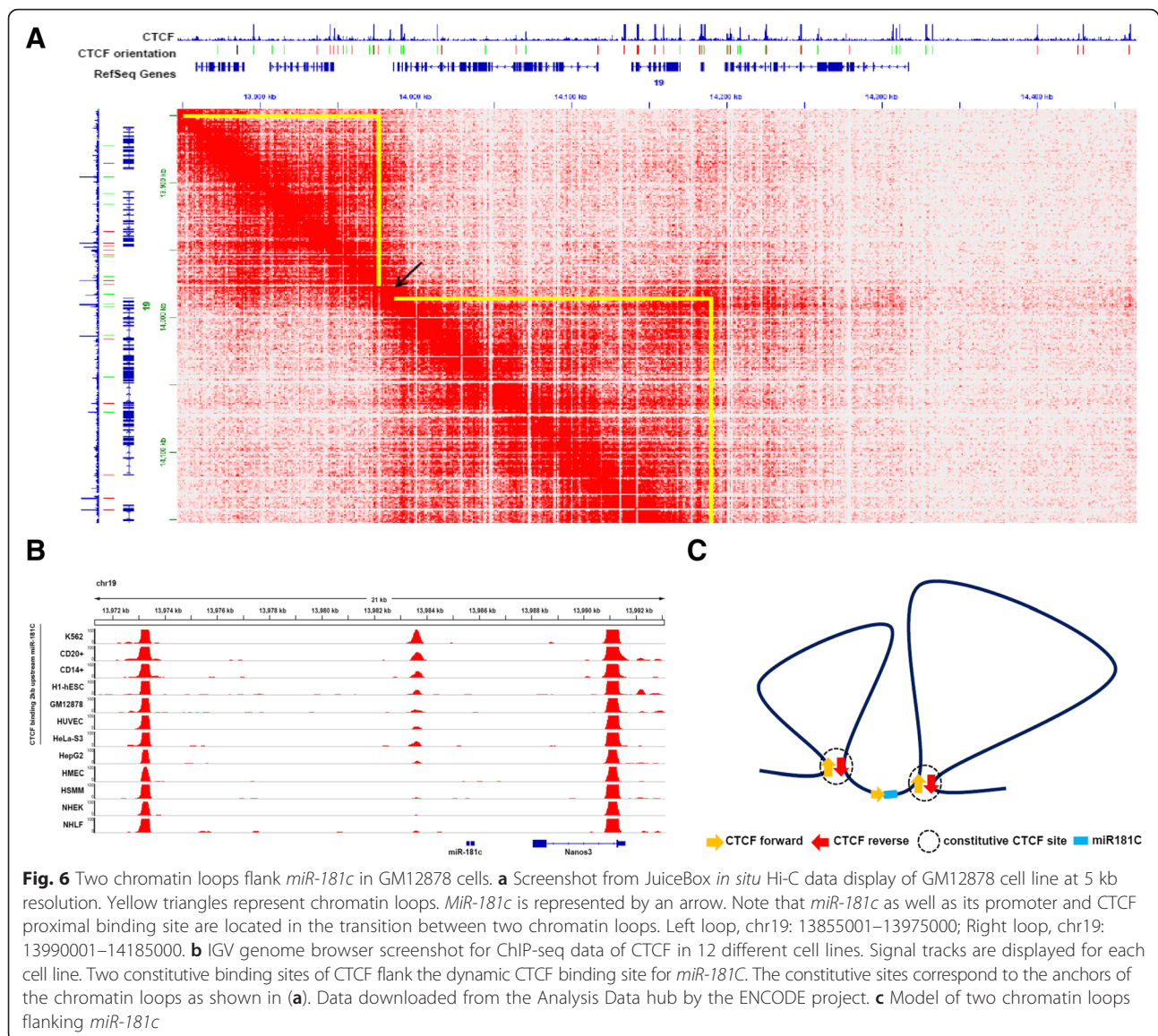
levels of *NOTCH2* transcripts in glioblastoma U87MG cells (Fig. 7b).

## Discussion

Cancer is a multistep disease that includes many interdependent components at the cellular level [44]. There are also molecular components that include genotypic abnormalities but more recently epigenotypic deregulation [45]. In particular, and based on the relevance of the post-transcriptional regulatory function of microRNAs over different types of genes we studied here how epigenetic regulatory processes can dysregulate microRNAs transcription in cancer. We asked how a microRNA, the *miR-181c*, involved in the regulation of brain specific genes can be epigenetically deregulated in glioblastoma cell lines, one of the more frequently occurring primary malignant brain tumours. We focused on the glioblastoma cell lines, T98G and U87MG, where the *miR-181c* is downregulated in comparison to normal brain tissues. This microRNA loss of gene expression correlated with a strong gain of DNA methylation in the *miR-181c* promoter region. Importantly, this aberrant DNA hypermethylation apparently interferes with the binding of the chromatin associated CTCF nuclear factor. CTCF depletion confirmed a gain of DNA methylation in U87MG cells supporting a previously reported

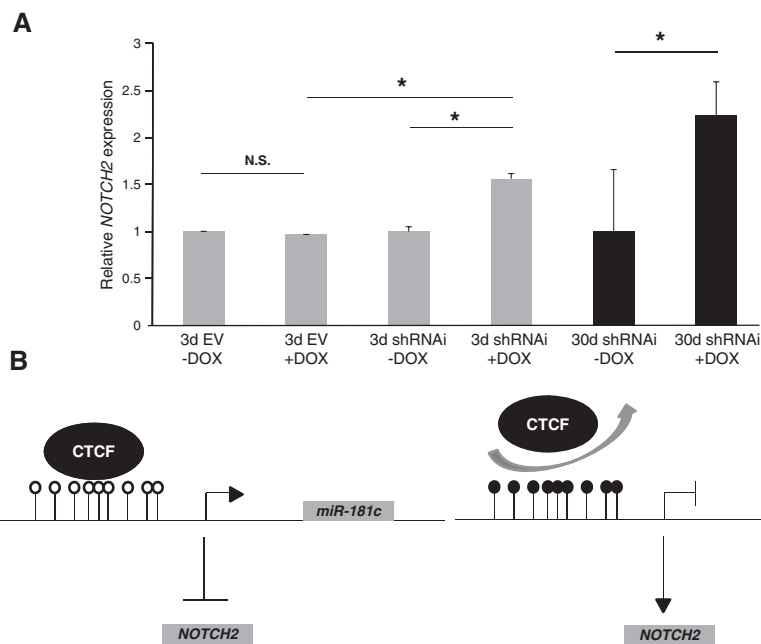
protective role of CTCF in tumour suppressor genes [25]. Finally, CTCF knockdown induces the upregulation of *NOTCH2* a target of *miR-181c*.

Concerning the transcriptional regulation of microRNAs an important sub-group is annotated as intergenic, but others are intronic and/or exonic, either in sense or antisense orientations presenting a more complex regulatory context. Genetic disruption of microRNAs has been documented in cancer, but there are some evidences that suggest that epigenetic alterations can be one of the major mechanisms for microRNA deregulation in cancer and other diseases [2]. There is a growing list of microRNAs that are subjected to epigenetic abnormal influence, including gain or loss of DNA methylation, histone covalent modifications, and more recently, the topological organization of the genome (see below). For example, it is well documented how members of the miR-34 family are involved in cancer through cell cycle arrest, cell invasion, apoptosis or even cancer metastasis [2]. These microRNAs are mainly silenced by DNA methylation of their promoter regions. Concerning the role of CTCF in microRNAs, a recent report showed that the miR-125b1 is aberrantly silenced by DNA methylation in breast cancer cells [30]. In such context, CTCF binding to the promoter region of the *miR-125b1* is disrupted and a gain in the repressive histone modification H3K9me3 and H3K27me3 is detected in cancer



cells [30]. Interestingly, alternative epigenetic silencing mechanisms exist, like the overexpression of EZH2, a key member of the Repressive Polycomb Complex, PRC2, that in addition to silence many genes, including tumour suppressor genes, can also silence different microRNAs in cancer cells [46]. It has been documented by several research groups that EZH2 is overexpressed in different cancers, and found to repress abnormally different microRNAs, including the miR-181c in prostate and breast cancer cells [47]. Then, based on our observation and the differential binding of CTCF to the miR-181c in different cell-types we propose that EZH2 and Polycomb proteins may be responsible for silencing the miR-181c in cell-types where the miR-181c is normally not expressed, like in the human erythroleukemic K562 cells or primary lymphocytes (Fig. 1).

An important aspect that is to a certain extent underestimated is the possibility that in glioblastoma cells CTCF is affected by mutations. Nowadays, there is a repertoire of different CTCF mutations, comprising somatic mutations, resulting in nonsense, missense, frameshift and splice site mutations [48]. Some of these mutations have been identified in different cancer types. From a functional point of view, a large proportion of mutations are found in the zinc-fingers that are critical for CTCF binding to DNA [48, 49]. Therefore, in glioblastoma cells and in regulatory regions as for the miR-181c, CTCF disruption can be caused by specific mutations that affect its binding to DNA. This view is further supported by a recent report in which *ctcf* hemizygous knockout mice predisposes to cancer, under certain inducible conditions, promoting tumour



**Fig. 7** Depletion of CTCF in U87MG cells affects the transcription levels of *NOTCH2*. **a** *NOTCH2* mRNA levels in U87MG cells transduced with an inducible shRNAi against CTCF after 3 and 30 days with or without DOX induction (3d shRNAi-DOX; 3d shRNAi + DOX; 30d shRNAi-DOX; 30d shRNAi + DOX). Empty vector controls are shown (3d EV-DOX and 3d EV + DOX). \**p*-value < 0.05. N.S., not significant. This set of data is representative of three-independent experiments. **b** In U87MG cells CTCF binds to the promoter region of *miR-181c* (Left). Loss of CTCF causes epigenetic silencing of *miR-181c* by DNA methylation and an increase of *NOTCH2* mRNA (Right)

aggressive invasion and metastatic dissemination [50]. What is even more relevant, in the context of the present study, is the fact that CTCF haploinsufficient mice destabilize genome-wide DNA methylation patterns supporting the relationship between CTCF and DNA methylation in certain genomic regions [50]. In the same study point mutations have been correlated with abnormal gain of DNA methylation. Therefore, CTCF is now considered as a tumour suppressor gene in human cancers and is significantly mutated gene in different types of cancers [50, 51].

Based on the recent series of publications and given the architectural role attributed to CTCF we cannot discard, that the CTCF located in the promoter region of the *miR-181c* plays a structural role [52]. Due to this possibility we analyzed the genomic distribution of CTCF, and its relationship with the three-dimensional architecture of the genome taking advantage of the newly, high resolution, genome-wide mapping of chromatin loops by *in situ* Hi-C [43]. *In situ* Hi-C series of experiments have reached up to 1 kb resolution. As shown in Fig. 6, the CTCF site associated with the *miR-181c* promoter does not seem to correspond to a loop anchor site (Fig. 6c). We believe that this is relevant, and we propose that this CTCF site is not a structural one, instead we suggest a local regulatory function, in

particular, protection against DNA methylation. In addition, Lieberman Aiden and collaborators demonstrated that more than 90 % of the CTCF sites at loop anchors, at the DNA binding sequence level, are positioned in a convergent orientation [43]. This is extremely relevant since this type of sequence convergence orientation for CTCF binding sites turns out to be an excellent predictor of chromatin loop formation. Based in such prediction we propose a model in which the *miR-181c*, and its adjacent gene *Nanos3*, are not included in a loop and their location correspond to a genomic region between two large chromosomal loops (Fig. 6c).

In glioblastoma the Notch signaling pathway is aberrantly activated [53]. *NOTCH2* is one of the receptors of the Notch pathway and was recently shown to be important for proliferation, invasion and self-renewal of glioblastoma U87MG cells [34]. The *NOTCH2* gene is also a post-transcriptionally target of *miR-181c* and a negative correlation between *NOTCH2* gene expression and *miR-181c* was found in glioblastoma samples [34]. In the present study we observed that CTCF knockdown induces overexpression of *NOTCH2* gene in U87MG glioblastoma cells possibly as a consequence of the epigenetic silencing by DNA methylation of *miR-181c* (Fig. 7a). This finding highlights the importance of

CTCF as a regulator of gene expression for tumour suppressor genes. In conclusion, microRNAs are subjected to multiple levels of regulation and there are few examples of how they are regulated transcriptionally, and even fewer how they are deregulated epigenetically. Due to their critical role during animal development it is important to continue exploring how these regulatory genes are controlled by a multitude of mechanisms.

## Conclusions

Cancer is a disease that combines genetic and epigenetic defects, in addition, to an active participation of microRNAs. The biogenesis and ways of action of microRNAs are relatively well known but their transcriptional regulation is an aspect that is poorly understood. Here we show that the *miR-181c* is differentially expressed in glioblastoma cell lines. As seen in some tumour suppressor genes CTCF is found in promoter regions protecting them against epigenetic silencing. The absence of CTCF correlates with gain of DNA methylation and the down-regulation of the *miR-181c* expression. Our results support the epigenetic role of CTCF in the regulation of microRNAs implicated in tumorigenesis.

## Additional file

**Additional file 1: Figure S1.** CTCF binds to the promoter of *miR-181c* in different cell lines. **Figure S2.** *In vivo* CTCF association in the promoter region of the *miR-181c*. **Figure S3.** Quantization of the inducible CTCF knockdown in the U87MG glioblastoma cells. **Table S1.** CTCF binds to the promoter of *miR-181c* in different cell lines. (PDF 902 kb)

## Abbreviations

5-azadC: 5-aza-2'-deoxycytidine; CTCF: CCCTC-binding factor; DMR: differential methylation region; DOX: doxycycline; ENCODE: encyclopedia of DNA elements; GBM: glioblastoma multiforme; RT-qPCR: quantitative reverse transcription-polymerase chain reaction; shRNAi: small-hairpin interference RNA.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

EA-O and FR-T designed the study and wrote the manuscript. EA-O, RA-M, RP-M, EG-B, GG and KM performed the experiments. FR-T, RA-M and EA-O performed the bioinformatic analysis of the corresponding genomic region. All authors read and approved the final version of the manuscript.

## Acknowledgements

We thank Paul Delgado-Olguín for critical reading of the manuscript, Gabriel Cuevas from the Instituto de Química (UNAM) for his constant support, Fernando Suaste-Olmos and Guadalupe Códiz Huerta for technical assistance. Ph.D. fellowships from CONACyT and Dirección General de Estudios de Posgrado-Universidad Nacional Autónoma de México (DGEP) (EA-O; RA-M; RP-M; EG-B). Additional support was provided by the PhD Graduate Program, "Doctorado en Ciencias Biomédicas y Ciencias Bioquímicas" and the Instituto de Fisiología Celular from the Universidad Nacional Autónoma de México. We acknowledge the post-doctoral fellowship from DGAPA-UNAM to KM. This work was supported by the DGAPA-PAPIIT, UNAM (IN209403, IN203811 and IN201114), CONACyT (42653-Q, 128464 and 220503) and Join Partnership Program UNAM-UIUC.

Received: 8 September 2015 Accepted: 10 March 2016

Published online: 16 March 2016

## References

- Jansson MD, Lund AH. MicroRNA and cancer. *Mol Oncol*. 2012;6(6):590–610.
- Suzuki H, Maruyama R, Yamamoto E, Kai M. Epigenetic alteration and microRNA deregulation in cancer. *Front Genet*. 2013;4:258.
- Benetatos L, Voulgaris E, Vartholomatos G, Hatzimichael E. Non-coding RNAs and EZH2 interactions in cancer: long and short tales from the transcriptome. *Int J Cancer*. 2013;133(2):267–74.
- Di Leva G, Garofalo M, Croce CM. MicroRNAs in cancer. *Ann Rev Pathol Mech Dis*. 2014;9:287–314.
- Denis H, Ndlovu MN, Fuks F. Regulation of mammalian DNA methyltransferases: a route to new mechanisms. *EMBO Rep*. 2011;12(7):647–56.
- Fabbri M, Garzon R, Cimmino A, Liu Z, Zanesi N, Callegari E, et al. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferase 3A and 3B. *Proc Natl Acad Sci U S A*. 2007;104(40):15805–10.
- Garzon R, Liu S, Fabbri M, Liu Z, Heaphy CE, Callegari E, et al. MicroRNA-29b induces global DNA hypomethylation and tumor suppressor gene reexpression in acute myeloid leukemia by targeting directly DNMT3A and 3B and indirectly DNMT1. *Blood*. 2009;113(25):6411–8.
- Zhao B, Bian EB, Li J, Li J. New advances of microRNAs in glioma stem cells, with special emphasis on aberrant methylation of microRNAs. *J Cell Physiol*. 2014;229(9):1141–7.
- Asuthkar S, Velpula KK, Chetty C, Gorantla B, Rao JS. Epigenetic regulation of miRNA-211 by MMP-9 governs glioma cell apoptosis, chemosensitivity and radiosensitivity. *Oncotarget*. 2012;3(11):1439–54.
- Ying Z, Li Y, Wu J, Zhu X, Yang Y, Tian H, et al. Loss of miR-204 expression enhances glioma migration and stem cell-like phenotype. *Cancer Res*. 2013;73(2):990–9.
- Lee HK, Bier A, Cazacu S, Finniss S, Xiang C, Twito H, et al. MicroRNA-145 is downregulated in glial tumors and regulates glioma cell migration by targeting connective tissue growth factor. *PLoS ONE*. 2013;8(2), e54652.
- Bier A, Giladi N, Kronfeld N, Lee HK, Cazacu S, Finniss C, et al. MicroRNA-137 is downregulated in glioblastoma and inhibits the stemness of glioma stem cells by targeting RTVP-1. *Oncotarget*. 2013;4(5):665–76.
- Lee HK, Finniss S, Cazacu S, Bucris E, Ziv-Av A, Xiang C, et al. Mesenchymal stem cells deliver synthetic microRNA mimics to glioma cells and glioma stem cells and inhibit their cell migration and self-renewal. *Oncotarget*. 2013;4(2):346–61.
- Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, et al. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol*. 1996;16(6):2802–13.
- Klenova EM, Nicolas RH, Paterson HF, Carne AF, Heath CM, Goodwin GH, et al. CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol Cell Biol*. 1993;13(12):7612–24.
- MacPherson MJ, Beatty LG, Zhou W, Du M, Sadowski PD. The CTCF insulator protein is posttranslationally modified by SUMO. *Mol Cell Biol*. 2009;29(3):714–25.
- Yu W, Ginjala V, Pant V, Chernukhin I, Whitehead J, Docquier F, et al. Poly(ADP-ribosyl)ation regulates CTCF-dependent chromatin insulation. *Nat Genet*. 2004;36(10):1105–10.
- Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell*. 2009;137(7):1194–211.
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485(7398):376–80.
- Zuin J, Dixon JR, van der Reijden MI, Ye Z, Kolovos P, Brouwer RW, et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A*. 2014;111(3):996–1001.
- Zlatanova J, Caiafa P. CTCF and its protein partners: divide and rule? *J Cell Sci*. 2009;122(Pt 9):1275–84.
- Choi NM, Feeney AJ. CTCF and ncRNA regulate the three-dimensional structure of antigen receptor loci to facilitate V(D)J recombination. *Front Immunol*. 2014;5:49.
- Saldaña-Meyer R, González-Buendía E, Guerrero G, Narenda V, Bonasio R, Recillas-Targa F, et al. CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, *Wrap53*. *Genes Dev*. 2014;28(7):723–34.



24. Recillas-Targa F, De La Rosa-Velázquez IA, Soto-Reyes R, Benítez-Bribiesca L. Epigenetic boundaries of tumour suppressor gene promoters: the CTCF connection and its role in carcinogenesis. *J Cell Mol Med*. 2006;10(3):554–68.
25. Recillas-Targa F, De La Rosa-Velázquez IA, Soto-Reyes E. Insulation of tumor suppressor genes by the nuclear factor CTCF. *Biochem Cell Biol*. 2011;89(5):479–88.
26. Engel N, West AG, Felsenfeld G, Bartolomei MS. Antagonism between DNA hypermethylation and enhancer-blocking activity at the *H19* DMD is uncovered by CpG mutations. *Nat Genet*. 2004;36(8):883–8.
27. Chang J, Zhang B, Heatch H, Galjart N, Wang X, Milbrandt J. Nicotinamide adenine dinucleotide (NAD)-regulated DNA methylation alters CTCF-binding factor (CTCF)/cohesin binding and transcription at the *BDNF* locus. *Proc Natl Acad Sci U S A*. 2010;107(50):21836–41.
28. Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, et al. Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res*. 2012;22(9):1680–8.
29. Saito Y, Saito H. Role of CTCF in the regulation of microRNA expression. *Front Genet*. 2012;3:186.
30. Soto-Reyes E, Recillas-Targa F. Epigenetic regulation of the human *p53* gene promoter by the CTCF transcription factor in transformed cell lines. *Oncogene*. 2010;29(15):2217–27.
31. de Souza Rocha Simonini P, Breiling A, Gupta N, Malekpour M, Youns M, Omrani-pour R, et al. Epigenetically deregulated microRNA-375 is involved in a positive feedback loop with estrogen receptor alpha in breast cancer cells. *Cancer Res*. 2010;70(22):9175–84.
32. Tata PR, Tata NR, Kühl M, Sirbu IO. Identification of a novel epigenetic regulatory region within the pluripotency associated microRNA cluster, EEmiRC. *Nucleic Acids Res*. 2011;39(9):3574–81.
33. Lakomy R, Sana J, Hankeova S, Fadrus P, Kren L, Lzicarova E, et al. MiR-195, miR-196b, miR-181c, miR-21 expression levels and O-6-methylguanine-DNA methyltransferase methylation status are associated with clinical outcome in glioblastoma patients. *Cancer Sci*. 2011;102(12):2186–90.
34. Ruan J, Lou S, Dai Q, Mao D, Ji J, Sun X. Tumor suppressor miR-181c attenuates proliferation, invasion and self-renewal abilities in glioblastoma. *NeuroReport*. 2015;26(2):66–73.
35. Sand M, Skrygan M, Sand D, Georgas D, Hahn SA, Gambichler T, et al. Expression of microRNAs in basal cell carcinoma. *Br J Dermatol*. 2012;167(4):847–55.
36. Jones KB, Salah Z, Del Mare S, Galasso M, Gaudio E, Nuovo GJ, et al. miRNA signature associate with pathogenesis and progression of osteosarcoma. *Cancer Res*. 2012;72(7):1865–77.
37. Cui MH, Hou XL, Lei XY, Mu FH, Yang GB, Yue L, et al. Upregulation of microRNA 181c expression in gastric cancer tissues and plasma. *Asian Pac J Cancer Prev*. 2013;14(5):3063–6.
38. Dávalos-Salas M, Furlan-Magaril M, González-Buendía E, Valdes-Quezada C, Ayala-Ortega E, Recillas-Targa F. Gain of DNA methylation is enhanced in the absence of CTCF at the human retinoblastoma gene promoter. *BMC Cancer*. 2011;11:232.
39. Gomes NP, Espinosa JM. Gene-specific repression of the p53 target gene *PUMA* via intergenic CTCF-cohesin binding. *Genes Dev*. 2010;24(10):1022–34.
40. Oszlak F, Poling LL, Wang Z, Liu H, Liu XS, Roeder RG, et al. Chromatin structure analyses identify miRNA promoters. *Genes Dev*. 2008;22(22):3172–83.
41. Cui H, Niemitz EL, Ravenel JD, Onyango P, Brandenburg SA, Lobanenko VV, et al. Loss of imprinting of insulin-like growth factor-II in Wilms' tumor commonly involves altered methylation but not mutations of CTCF or its binding site. *Cancer Res*. 2001;61(13):4947–50.
42. Sexton T, Cavalli G. The role of chromatin domains in shaping the functional genome. *Cell*. 2015;160(6):1049–59.
43. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinsin JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159(7):1665–80.
44. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646–74.
45. Recillas-Targa F. Interdependency between genetic and epigenetic regulatory defects in cancer. *Methods Mol Biol*. 2014;1165:33–52.
46. Au SL, Wong CC, Lee JM, Fan DN, Tsang FH, Ng IO, et al. Enhancer of zeste homolog 2 epigenetically silences multiple tumor suppressor microRNAs to promote liver cancer metastasis. *Hepatology*. 2012;56(2):622–31.
47. Cao Q, Manu RS, Russo N, Scanlon CS, Tsodikov A, Jing X, et al. Coordinated regulation of polycomb group complexes through microRNAs in cancer. *Cancer Cell*. 2011;20(2):187–99.
48. Tiffen JC, Bailey CG, Marshall AD, Metierre C, Feng Y, Wang Q, et al. The cancer-testis antigen BORIS phenocopies the tumor suppressor CTCF in normal and neoplastic cells. *Int J Cancer*. 2013;133(7):1603–14.
49. Nakahashi H, Kwon KR, Resch W, Vian L, Dose M, Stavreva D, et al. A genome-map of CTCF multivalency redefines the CTCF code. *Cell Rep*. 2013;3(5):1678–89.
50. Kemp CJ, Moore JM, Moser R, Bernard B, Teater M, Smith LE, et al. CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. *Cell Rep*. 2014;7(4):1020–9.
51. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. 2014;505(7484):495–501.
52. Ong CT, Corces VG. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet*. 2014;15(4):234–46.
53. Stockhausen MT, Kristoffersen K, Poulsen HS. The functional role of Notch signaling in human gliomas. *Neuro Oncol*. 2010;12(2):199–211.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)



## INCREASED EXPRESSION OF BDNF TRANSCRIPT WITH EXON VI IN HIPPOCAMPI OF PATIENTS WITH PHARMACO-RESISTANT TEMPORAL LOBE EPILEPSY

G. A. MARTÍNEZ-LEVY,<sup>a</sup> L. ROCHA,<sup>b</sup> F. D. LUBIN,<sup>c</sup>  
M. A. ALONSO-VANEGAS,<sup>d</sup> A. NANI,<sup>a</sup>  
R. M. BUENTELLO-GARCÍA,<sup>d</sup> R. PÉREZ-MOLINA,<sup>e</sup>  
M. BRIONES-VELASCO,<sup>a</sup> F. RECILLAS-TARGA,<sup>e</sup>  
A. PÉREZ-MOLINA,<sup>a</sup> D. SAN-JUAN,<sup>f</sup>  
J. CIENFUEGOS<sup>d</sup> AND C. S. CRUZ-FUENTES<sup>a\*</sup>

<sup>a</sup> Department of Genetics, National Institute of Psychiatry “Ramón de la Fuente Muñiz” (INPRFM), Mexico DF, Mexico

<sup>b</sup> Department of Pharmacobiology, Center for Research and Advanced Studies (CINVESTAV), Mexico City, Mexico

<sup>c</sup> Department of Neurobiology, Lubin Lab, University at Birmingham in Alabama, USA

<sup>d</sup> Neurosurgery Section, National Institute of Neurology and Neurosurgery “Manuel Velasco Suárez” (INNNMVS), Mexico City, Mexico

<sup>e</sup> Institute of Cell Physiology, National University of Mexico, UNAM, Mexico City, Mexico

<sup>f</sup> Clinical Research Department, National Institute of Neurology and Neurosurgery “Manuel Velasco Suárez” (INNNMVS), Mexico City, Mexico

**Abstract**—A putative role of the brain-derived neurotrophic factor (BDNF) in epilepsy has emerged from *in vitro* and animal models, but few studies have analyzed human samples. We assessed the BDNF expression of transcripts with exons I (BDNFI), II (BDNFII), IV (BDNFIV) and VI (BDNFVI) and methylation levels of promoters 4 and 6 in the hippocampi of patients with pharmaco-resistant temporal lobe epilepsy (TLE) ( $n = 24$ ). Hippocampal sclerosis (HS) and pre-surgical pharmacological treatment were considered as clinical independent variables. A statistical significant increase for the BDNFVI ( $p < 0.05$ ) was observed in TLE patients compared to the autopsy control group ( $n = 8$ ). BDNFVI was also increased in anxiety/depression TLE ( $N = 4$ ) when compared to autopsies or to the remaining group of patients ( $p < 0.05$ ). In contrast, the use of the antiepileptic drug Topiramate (TPM) ( $N = 3$ ) was associated to a decrease in BDNFVI expression ( $p < 0.05$ ) when compared to the remaining group of patients. Methylation levels at the BDNF promoters 4 and 6 were similar between TLE and autopsies and in relation to

the use of either Sertraline (SRT) or TPM. These results suggest an up-regulated expression of a specific BDNF transcript in patients with TLE, an effect that seems to be dependent on the use of specific drugs. © 2015 IBRO. Published by Elsevier Ltd. All rights reserved.

**Key words:** pharmaco-resistant temporal lobe epilepsy, BDNF, mRNA expression, epigenetics, promoter methylation, human.

### INTRODUCTION

Epilepsy affects 1–2% of the worldwide population. The most prevalent form of epilepsy is the one that affects the temporal lobe epilepsy (TLE) (Engel et al., 1997; Goldenberg, 2010); very often these patients are refractory to pharmacological treatment (70% of TLE patients) and in most cases surgery is indicated (Jardim et al., 2012).

Seizures in TLE are produced by the synchronized hyperactivity of neuron populations due to the disruption of the balance between excitatory and inhibitory synaptic transmission (Casillas-Espinosa et al., 2012). Cumulative evidence supports a role for the brain-derived neurotrophic factor (BDNF), involved in synaptic plasticity, as a critical mediator in this pathological process (Koyama and Ikegaya, 2005; Cunha et al., 2010).

Several studies have evaluated BDNF expression in the brain tissue of TLE patients. For instance, in comparison to autopsy controls, Takahashi et al. (1999) found an increment of BDNF protein in the epileptic cortex. Similarly, an increased immunoreactivity of BDNF and its associated Tyrosine receptor kinase B (TrkB) was observed in the TLE hippocampus (Hou et al., 2010). Moreover, an increased hybridization of the BDNF mRNA is detected in hippocampal granule cells from affected individuals (Murray et al., 2000).

It has also been reported that expression of BDNF is affected by: hippocampal sclerosis (HS) (Mathern et al., 1997; Wang et al., 2011), duration and severity of seizures (Hong et al., 2014), fluoxetine treatment in patients with comorbid depression (Kandratavicius et al., 2013) and by antiepileptic drugs (Hou et al., 2010). Therefore, these variables warrant further consideration when the expression of this neurotrophin is being assessed.

\*Corresponding author. Tel: +52-55-41605074.

E-mail address: cruz@imp.edu.mx (C. S. Cruz-Fuentes).

**Abbreviations:** BDNF, brain-derived neurotrophic factor; BDNFI, BDNF transcript with exon I; BDNFII, BDNF transcript with exon II; BDNFIV, BDNF transcript with exon IV; BDNFVI, BDNF transcript with exon VI; CT, cycle threshold; DEPC, diethyl-pyrocabonate; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; HDACi, histone deacetylase inhibitor; HS, hippocampal sclerosis; MRI, magnetic resonance image; PMIs, post-mortem intervals; SE, status epilepticus; SRT, Sertraline; TBP, TATA-binding protein; TLE, temporal lobe epilepsy; TPM, Topiramate; VPA, valproate.

BDNF is highly expressed in brain areas implicated in limbic seizures (Lee et al., 1997). Studies in animal models have related seizure activity with changes in the expression of this neurotrophin (Mudò et al., 1996; Lee et al., 1997). For example, transgenic mice overexpressing BDNF, display an increased seizure severity in response to kainic acid, associated to a hippocampal hyperexcitability (Croll et al., 1999). Similarly, the administration of an intra-hippocampal bolus of BDNF, before electrical stimulation increases the susceptibility and duration of kindled seizures (Xu et al., 2004). Moreover, the chronic inhibition of TrkB receptor, initiated after the status epilepticus (SE), prevents the consequent generation of spontaneous and recurrent seizures, ameliorates the anxiety-like behavior and reduces the loss of hippocampal neurons (Liu et al., 2013). All these observations support a role of this neurotrophin in the development of epilepsy. On the contrary, the reports showing that viral vector-mediated of BDNF/Fibroblast Growth Factor 2 decrease the severity and number of spontaneous seizures in the pilocarpine-induced SE model (Paradiso et al., 2009, 2011) these have led to consider the use of BDNF as an alternative therapy for epilepsy treatment (Simonato et al., 2006).

The contrasting effects of this neurotrophin could be associated with the multiplicity of transcripts synthesized from the BDNF gene (for a review see Martínez-Levy and Cruz-Fuentes, 2014). Messenger RNAs containing exons I (BDNFI), II (BDNFII), IV (BDNFIV) and VI (BDNFVI) (Fig. 1) are the most commonly expressed in different tissues (Garzon and Fahnestock, 2007; Pruunsild et al. 2007). In the brain, they are localized in different neuronal compartments (Pattabiraman et al., 2005; Aliaga et al., 2009), affecting synaptic plasticity within each area (Baj et al., 2011). The specific cellular location of the transcripts raises the possibility that a distorted regulated trafficking and localized synthesis of BDNF might impact the delicate balance between excitatory and inhibitory transmission, with possible neuro-patho-physiological consequences, including epileptogenesis (Cunha et al., 2010). It is worth noting that no studies have evaluated the expression of these transcripts in human TLE.

Finally, a relation between DNA methylation and BDNF expression has been described for some

neurological and psychiatric disorders (Martínez-Levy and Cruz-Fuentes, 2014). De-methylation of BDNF promoter 1 induced by picrotoxin (a GABAA receptor chloride channel blocker) has been associated with a synaptic hyper-excitability in slices from mouse hippocampus (Nelson et al., 2008). Similarly, an increment in the expression of transcripts with exon IX has been associated with a decrease in the methylation levels on its specific promoter in the kainic acid model of TLE (Parrish et al., 2013). Additionally, valproic acid (VPA) and other antiepileptic drugs (D'Addario et al., 2013; Guidotti and Gryson, 2014) seem to affect DNA methylation at specific BDNF promoters. These data indicate that changes in BDNF gene methylation could be associated with epilepsy; however, its role in human tissue remains to be analyzed.

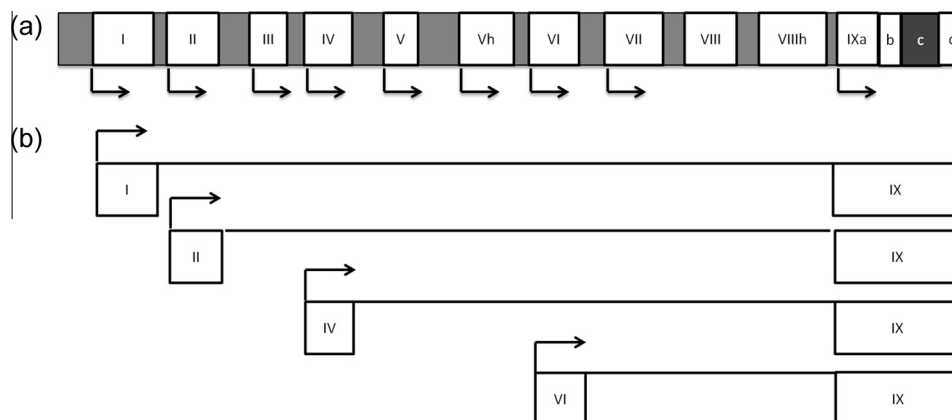
The aim of the present study is to assess the expression of four BDNF transcripts and methylation levels at specific BDNF promoters in the surgically resected hippocampi of patients with pharmacoresistant TLE, in comparison to the autopsy samples. Age, duration of epilepsy, HS and pre-surgical pharmacological treatment were considered as independent variables.

## EXPERIMENTAL PROCEDURES

### Subjects

All experiments were performed with the approval of the Ethics Committee of the National Institute of Neurology and Neurosurgery “Manuel Velasco Suárez” (INNNMVS) in Mexico City (project 70/12). All participants signed an informed consent. Patients ( $n = 24$ ) met diagnostic criteria for TLE and were refractory to pharmacological treatment. They underwent a detailed medical history, video electroencephalography (EEG) recordings and magnetic resonance image (MRI) studies. Hippocampus Sclerosis was defined by the presence of hippocampus atrophy and a hyper-intense signal on T2-weighted MRI (General Electric 3T unit).

The presence of depression and anxiety disorders was assessed by means of The Hospital and Anxiety Depression Scale (HADS; Spanish version) (Herrero et al., 2003; Gómez-Arias et al., 2012). Diagnoses were confirmed with the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) (First Michael et al.,



**Fig. 1.** BDNF gene structure and the four transcripts analyzed. White boxes indicate exons, gray spaces point to introns and arrows designate the alternative transcription start sites. Scheme showing the organization of the four alternative transcripts analyzed in this study.

2002), applied by a psychiatrist who was blind to the epilepsy diagnosis.

Relevant clinical data including: age of the subjects at the moment of surgery, gender, onset date year of epilepsy, epilepsy duration, epilepsy surgery date, precipitating factors of epilepsy, other diseases/diagnoses and treatments associated with epilepsy or with other comorbidities and the surgical outcome, were obtained from the clinical records.

Hippocampal samples from 18 patients displaying HS (10 men and 8 women) and from six patients without HS (3 men and 3 women) were collected. Clinical data from each patient and location of the epileptic focus are shown in Table 1.

Surgeries were performed between February 2012 and June 2013. The surgery procedure in patients with HS was carried out using the T2 or T3 trans-temporal approach guided by the electrocorticographic (ECoG) signals recorded on the brain surface (4 × 8 electrode-gird, Ad-Tech, Racine, WI, USA). The procedure consisted of the

unilateral amygdalo-hippocampectomy, which includes the microsurgical excision of the head of the anterior body of the hippocampus, amygdala, portions of the uncus, the parahippocampal gyrus and was followed by the tailored resection of the temporal neocortex (San Juan et al., 2011). Patients without HS were submitted to a similar surgical procedure; in this case amygdalo-hippocampectomy was practiced depending on the localization of the lesion. After a careful dissection, the hippocampus was stored for the following procedures. Biopsies from tumors, malformations or any other alteration were not included in the present study; in those cases where tumors or lesions were restricted to the hippocampus; lesion's margins were used for the present study.

All samples from TLE patients were stored immediately after resection in a Ringer solution (pH 7) containing diethyl-pyrocabonate (DEPC) and were transported on ice to the laboratory within a maximum of 2 h, where RNA extraction was performed.

**Table 1.** Demographic and clinical data of TLE patients and autopsies

PX	Gender	Age	Seizure onset age	Duration of epilepsy	Precipitating factors	AED before surgery	Other drugs	Side focus	Surgical out-come Engel's	Final diagnosis
367	F	22	12	10	FH, CET, FS	CBZ, LMG	–	Right	NA	HS
355	M	60	8	52	CET	VPA, CBZ, CNZ	RSP, BP	Left	1	HS
392	M	41	2	39	FH, FS	PHE, LMG, CBZ	SRT	Left	1	HS
359	F	33	27	6	FS	PHE, TPM, CBZ	–	Left	1	HS
215	M	19	1	18	FH, HYPOXIA	VPA, LMG, TPM, LVT	–	Right	1	HS
275	F	23	13	10	FH	CBZ, LVT	–	Left	3	HS
289	M	37	5	32	FS, CET	CBZ	–	Right	1	HS
290	F	24	1	23	CET	VPA, CBZ	–	Right	1	HS
295	M	43	27	16	–	VPA, PMD	–	Right	1	HS
301	M	56	11	45	–	PHE	–	Right	1	HS
283	M	30	14	16	–	VPA, CBZ	–	Left	1	HS
296	F	38	30	8	–	VPA	–	Left	1	HS
286	F	33	12	21	FH	VPA, PHB	–	Right	1	HS
305	F	21	14	7	FH, CET	PHE	–	Left	1	HS
276	M	54	13	41	FH, CET	CBZ, LMG	–	Right	1	HS
293	M	23	10	13	CET	PHE, VPA	–	Right	1	HS
292	F	42	4	38	–	CBZ, LMG	–	Left	1	HS
282	M	37	9	28	HYPOXIA	VPA, CBZ	SRT	Right	1	HS
67	F	27	27	0.2	CET	TPM	–	Right	1	CM
106	F	29	12	17	FS, CET	PHE, CBZ	–	Left	1	OA
207	M	46	45	1	–	PHE	DEX	Right	NA	GBM
279	F	18	14	4	–	PHE	–	Left	1	GII A
309	M	49	14	35	–	CBZ, LC	SRT	Left	1	Cysticercosis
291	M	34	30	4	–	CBZ	SRT	Left	1	CD
CTRL	Gender	Age	Side	Post-mortem internal (hours)		Cause of death				
C1	M	33	Right	8–10		Fire arm missile				
C2	M	37	Left	2–5		Fire arm missile				
C3	M	30	Left	10–12		Fire arm missile				
C4	M	73	Left	10–12		Pneumonia				
C5	M	36	Left	10–12		Fire arm missile				
C6	F	22	Left	12–14		Suffocation				
C7	F	18	Left	16–18		Suffocation				
C8	M	45	Left	16–18		Suffocation				

CET: cranoencephalic trauma; FH: family history; FS: febrile seizures; PHE: phenytoin sodium; VPA: valproic acid; LMG: lamotrigine; TPM: Topiramate; LVT: levetiracetam; CBZ: carbamazepine; GBP: gabapentin; PMD: pyrimidone; CNZ: clonazepam; PHB: phenobarbital; LC: lacosamide; DEX: dexamethasone; SRT: Sertraline; RSP: risperidone; BP: biperiden; HS: hippocampal sclerosis; CM: cavernous malformation; OA: oligoastrocytoma; GBM: glioblastoma; GII A: grade II astrocytoma; CD: cortical dysplasia; ENGEL 1: free of disabling seizure; 2: rare disabling seizures; 3: worthwhile improvement; NA: not available.



Hippocampus of eight individuals with a cause of death unrelated to a neurological or psychiatric disorder was obtained (Table 1); these samples were immediately frozen at  $-70^{\circ}\text{C}$ .

### Demographic data of the sample

The autopsy control group included six males and two females (mean age at the moment of death:  $35 \pm 12$  years old). In all but one of the samples, the left hippocampus was analyzed. In all cases, the post-mortem interval (PMI) was  $< 18$  h.

Patients with TLE were 13 males and 11 females with a mean age of  $35 \pm 12$  years. The male/female ratio ( $\chi^2 = 1.08$ ,  $df = 1$ ,  $p \geq 0.05$ ) and mean age ( $t = 0.37$ ,  $p \geq 0.05$ ) were statistically similar with respect to the autopsy group. Clinical features: age at seizure onset,  $14.6 \pm 10.7$ ; years with epilepsy,  $20.2 \pm 14.8$  and number of seizures the month before surgery  $2.2 \pm 4.5$ . The anti-epileptic treatment most commonly used was Carbamazepine (58.3%), followed by valproate (VPA) (40.9%), Phenytoin (33.3%), Lamotrigine (20.8%), Topiramate (TPM) (12.5%), Levetiracetam (8.3%), Clonazepam (4.2%) and Pyrimidine (4.2%). From six cases that received adjuvant pharmacological treatment for treating comorbidities, four used Sertraline (SRT) for ameliorating anxiety and depression symptoms. The post-operative outcome evaluated 1 year after surgery, indicated that in most cases (87.5%) the seizure freedom was achieved (Engel class I); in a single case a worthwhile improvement was demonstrated (Engel class III), while in two cases no data were available as they did not return to post-operative studies.

### Extraction of nucleic acids

Brain tissue (100 mg) was homogenized in 1 ml of guanidinium isothiocyanate solution; the half of the volume was used for RNA extraction as described in Chomczynski and Sacchi (1987); high-molecular weight DNA was extracted from the remaining amount using the chloroform–phenol technique.

In order to assure appropriate comparisons between-groups, every DNA/RNA sample was checked for parameters of quantity and purity (NanoDrop spectrophotometer); the integrity was evaluated by gel electrophoresis. No differences were detected in these variables between hippocampus of cases and autopsies.

Since samples of hippocampus from patients with TLE and those of autopsies were not obtained under strictly similar conditions, the following assay was performed: 12 rats were euthanized by decapitation and their skulls left at room temperature for different PMIs: 0 h ( $N = 4$ ), 4 h ( $N = 2$ ), 8 h ( $N = 2$ ), 16 h ( $N = 2$ ) and 24 h ( $N = 2$ ). At the end of each period, the cerebrum was removed and the brain cortex dissected. Two samples of the experimental condition 0 h were stored for 2 h on ice in a Ringer solution (pH 7) containing DEPC while the remaining were stored at  $-70^{\circ}\text{C}$  for 24 h. Total RNA was then extracted as described above.

Even though the amount of RNA recovered was inversely related to the postmortem interval; this

variable did not affect the integrity and purity, as well as the expression of the rat BDNFIV. Similarly, no significant differences were detected between the samples stored in Ringer/DEPC solution as compared to those from PMIs of 16 h. These results imply that different collecting methods did not exert important effects on the results obtained in the analysis of the human BDNF gene.

### Expression analysis of BDNF transcripts

Reverse transcription reactions were set up using 4  $\mu\text{g}$  of total RNA, and 200 U of the Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT; Invitrogen, Carlsbad, CA, USA; Catalog: 28025-013). The cDNA was stored until further use at  $-20^{\circ}\text{C}$ .

BDNF expression was assayed using the following ABI probes: BDNFI (Catalog: hs00538277-m1), BDNFII (Catalog: hs00538278-m1), BDNFIV (Catalog: hs00380947-m1), BDNFVI (Catalog: hs00156058-m1) (Wong et al., 2010).

TATA-binding protein (TBP) (Catalog: 4333769F) was chosen as the endogenous control probe. The rationale for its employment instead of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) used in previous reports (Mathern et al., 1997; Wang et al., 2011) was based on our experimental observation of an important delay in the cycle threshold (CT) values when using GAPDH (CT for epilepsy samples =  $34.7 \pm 3.7$  and CT for autopsies =  $31.2 \pm 1.7$ ,  $t = 2.4$ ,  $p \leq 0.05$ ) as opposed to TBP (CT for epilepsy:  $31.7 \pm 2.7$ ; CT for autopsies =  $31.6 \pm 1.2$ ,  $U = 93$ ,  $p \geq 0.05$ ).

Similar results have been previously reported in neocortical tissue of patients with TLE as compared to autopsy control groups (Wierschke et al., 2010). Moreover, it has also been reported that *in vitro* administration of BDNF changes GAPDH mRNA levels while TBP expression shows more stability (Santos and Duarte, 2008).

Each 25- $\mu\text{l}$  PCR reaction contained 250 ng of cDNA, 1.25  $\mu\text{l}$  of probe and 12.5  $\mu\text{l}$  of Taqman Universal PCR master mix, No-Amperase UNG (Catalog 4324018). No-template controls were included in each experiment. Samples were run in duplicate. Sds software (version 2.0; Applied Biosystems) was used for obtaining the Ct values.

### Methylation analysis of BDNF promoters

One microgram of DNA was treated with sodium bisulfite using an EpiTect 96 Bisulfite Kit, according to the manufacturer's instructions. (Qiagen, USA).

After bisulfite modification, a promoter IV and promoter VI PCR was performed, employing a biotinylated primer. In the first case, the PCR reaction was carried out in a 25  $\mu\text{l}$  solution containing 1  $\mu\text{l}$  of bisulfite-modified genomic DNA, 1X PyroMark PCR Master Mix, 1X CoralLoad Concentrate, 1X Q-Solution,  $\text{MgCl}_2$  2.5 mM (PyroMark PCR Kit, Qiagen, USA) and 0.5  $\mu\text{M}$  of the forward (F: 5'-[Bt]TTTGTGGGGTTGGAAGTGAAAAT-3') and reverse primers (R: 5'-CCCATCAACRAAAACTCCATTTAATCTC-3') (Keller et al., 2010).

PCR conditions included an initial denaturation step for 5 min at 95 °C, followed by 49 cycles of 15 s at 95 °C, 1 min at 59 °C, and 1 min at 72 °C and a final extension step of 10 min at 72 °C. For BDNF promoter VI, the PCR reaction was carried out in a 25 µl solution containing 1 µl of bisulfite-modified genomic DNA, 1× of KAPA2G Robust HotStart ReadyMix (Kapa Biosystems, USA) and 0.5 µM of the forward (F: 5'-GAGGAGGTGAG GATAGGTTT-3') and reverse primers (R: 5'-[Btñ]CTCAT TAAAACCCCCRAACAAAAAATAA-3'). PCR conditions included an initial denaturation step for 5 min at 95 °C; 40 cycles of 15 s at 95 °C, 15 s at 60 °C and 45 s at 72 °C and a final extension step of 10 min at 72 °C.

For pyrosequencing analyses, 10 µl of bisulfite-PCR products (previously cleaned with 2 µl of exo-sap; Affimetrix, USA) were processed according to the manufacturer's standard protocol (Qiagen, USA) with the following sequencing primers (Promoter IV S: 5'-AC AAAAAATTTTCATACTAA-3'; Promoter VI S: 5'-AGTT TAATYAGAAGAGTTAAATAATG-3').

Standard DNA samples (EpigenDX, USA) with methylation levels of 0%, 5%, 10%, 25%, 50%, 75%, and 100%, were also bisulfite-treated and analyzed for each group of primers.

### Statistical analyses

Expression levels are represented in graphs as  $2^{-\Delta\Delta CT}$  values. The  $2^{-\Delta\Delta CT}$  formula, as described in Schmittgen and Livak (2008) was used to calculate the fold of change between experimental groups. Percentage of methylation levels was obtained with the Pyromark CpG software. The results are presented in the graphs as mean  $\pm$  2 standard deviation (SD) and were considered significant when  $p < 0.05$ .

Statistical analyses were carried out with IBM SPSS Statistic v.20. *T* test for independent samples was used to evaluate differences between cases and autopsies, while the effect of specific clinical variables was evaluated using ANOVA and Dunnett's T3 post hoc test. A non-Parametric Mann–Whitney *U* Statistic test or the Kruskal–Wallis One-Way Analysis of Variance on Ranks was performed when necessary.

For the premixed methylation samples of DNA, a regression analysis was carried out for each group of primers (exons IV and VI) and linearity was evaluated.

## RESULTS

### Changes in BDNF mRNA transcripts expression

The four assayed BDNF transcripts were detected in all samples, with higher expression levels for those BDNFI and BDNFIV (Fig. 2) in the autopsy sample, as previously reported (Pruunsild et al. 2007).

Compared to autopsies, the fold-of-change for the different BDNF transcripts analyzed in patients with TLE were: (BDNFI: 2.3, BDNFII: 1.7; BDNFIV: 1.0; BDNFVI: 2.0) but transcript with exon VI ( $t = 2.1$ ,  $p \leq 0.05$ ) was the only one that shows a statistical significant increase (Fig. 2). At the individual patient level the expression of the four BDNF transcripts changed in a similar way (i.e.

if a patient showed high levels of transcript with exon I, high levels of transcript with exon II, IV and VI were also observed). Nonetheless the degree of increment in epileptic tissue was higher for transcript VI.

*Changes of BDNF expression with respect to other clinical variables.* Age, gender, age at seizure onset, years with epilepsy, number of monthly seizures and laterality of resection did not correlate with the expression of BDNF transcripts. Likewise, BDNF transcript expression in TLE-HS cases was similar to that associated with other types of lesions (TLE non-HS) (Fig. 3).

Among the different antiepileptic drugs used by patients, only TPM ( $n = 3$ ) was associated with a significantly lower expression ( $F = 5.28$ ,  $p \leq 0.05$ ) of BDNFVI as compared to the remaining group of the patients (0.2 times) and to the autopsy sample (0.5 times) (Table 2, Fig. 4). Moreover, a negative but not statistically significant correlation between the expression levels and the dose of TPM used was found ( $F = 63.5$ ,  $R^2 = 0.97$ ,  $p = 0.07$ ).

In contrast, an increase in BDNF transcript with exon VI was observed in TLE patients that used SRT ( $n = 4$ ,  $F = 6.42$ ,  $p \leq 0.05$ ) as compared to the remaining affected individuals (2.5 times), and to autopsies (4.5 times) (Table 2, Fig. 5).

### Methylation levels and BDNF promoters 4 and 6

For the premixed methylation samples of DNA, the analyses showed highly significant results ( $p \leq 0.01$ ) with  $R^2$  ranging from 0.84 to 0.92, indicating that our experiments adequately resemble the methylation levels of the samples.

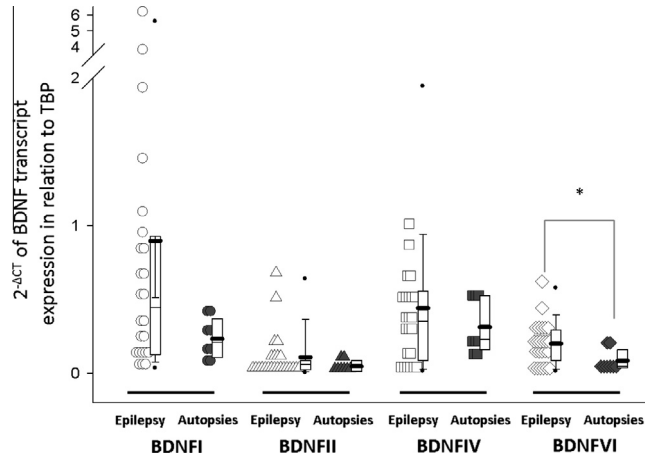
Low levels of methylation ( $\geq 15\%$ ) were detected in the first five CpGs located after the start site of transcription in promoters 4 and 6, but non-significant differences between TLE and autopsies were observed (Fig. 6).

No changes on DNA methylation was observed either with the presence of HS or with the use of specific pharmacological treatments such as TPM and SRT.

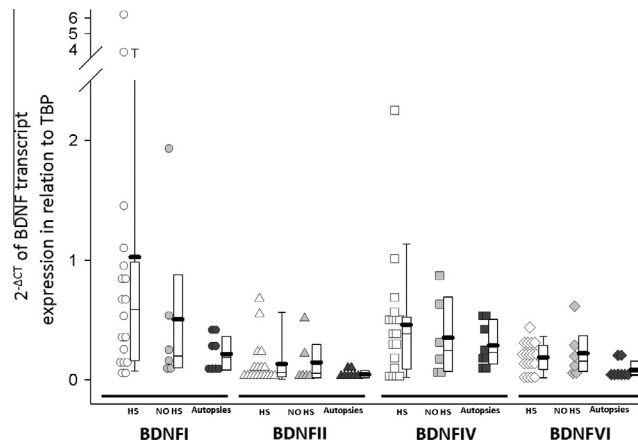
## DISCUSSION

We report here an increased expression of BDNF exon VI in pharmaco-resistant TLE patients' hippocampus compared to an autopsy control group. The heightened expression of a specific transcript is in agreement with previous studies in human epileptic tissue where a rise in BDNF protein and mRNA levels was reported (Mathern et al., 1997; Murray et al., 2000; Hou et al., 2010).

An increase in the expression of BDNF has been observed in the acute phase of the pilocarpine- or kainate-induced SE (Shetty et al., 2003; Tongiorgi et al., 2004). Some researchers (Heinrich et al., 2011; Parrish et al., 2013), but not others (Shetty et al., 2003; Tongiorgi et al., 2004) report that this increment is maintained during the late chronic phase (i.e. after at least 1 month of SE) when spontaneous recurrent seizures are already present, suggesting that BDNF could be impli-



**Fig. 2.** Expression of four BDNF transcripts in hippocampus of TLE patients and autopsies. Each symbol corresponds to individual data. Circles represent the expression of BDNFI, triangles BDNFII, square BDNFIV and diamonds BDNFVI. Open symbols represent TLE cases while dark gray symbols correspond to autopsies. The mean is described as a line in between the box-plot. Statistical significance was determined using an unpaired, two-tailed *t*-test, \**p* < 0.05.



**Fig. 3.** Expression of four BDNF transcripts in relation to hippocampal sclerosis (HS). HS: presence of hippocampal sclerosis; no-HS: absence of hippocampal sclerosis. Each symbol corresponds to individual data. Circles represent the expression of BDNFI, triangles BDNFII, square BDNFIV and diamonds BDNFVI. Open symbols represent HS-TLE cases; light gray symbols indicate non-HS-TLE cases and dark gray symbols correspond to autopsies. The mean is described as a line in between the box-plot. Statistical significance was determined using a one-way ANOVA followed by T3 Dunnett's multiple comparison test.

**Table 2.** Effect of pharmacological treatment in the expression of BDNF transcripts

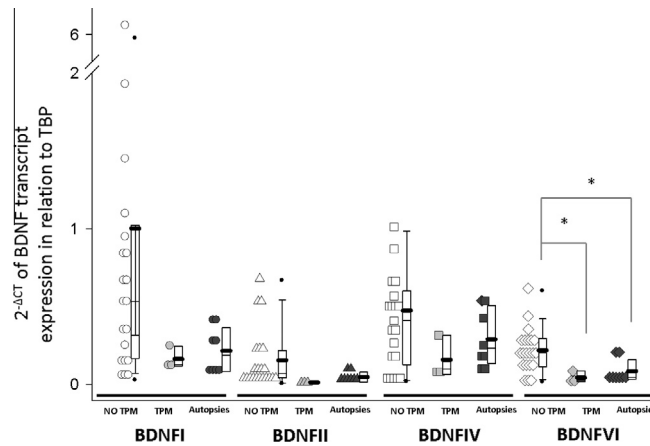
Drug	Sample	BDNFI	BDNFII	BDNFIV	BDNFVI
PHE	8	<i>F</i> = 0.86 <i>p</i> ≥ 0.05	<i>F</i> = 0.95 <i>p</i> ≥ 0.05	<i>F</i> = 0.49 <i>p</i> ≥ 0.05	<i>F</i> = 2.14 <i>p</i> ≥ 0.05
VPA	9	<i>F</i> = 1.74 <i>p</i> ≥ 0.05	<i>F</i> = 2.11 <i>p</i> ≥ 0.05	<i>F</i> = 0.79 <i>p</i> ≥ 0.05	<i>F</i> = 2.32 <i>p</i> ≥ 0.05
CBZ	14	<i>F</i> = 1.19 <i>p</i> ≥ 0.05	<i>F</i> = 0.62 <i>p</i> ≥ 0.05	<i>F</i> = 0.32 <i>p</i> ≥ 0.05	<i>F</i> = 2.14 <i>p</i> ≥ 0.05
LMG	5	<i>F</i> = 2.62 <i>p</i> ≥ 0.05	<i>F</i> = 0.66 <i>p</i> ≥ 0.05	<i>F</i> = 0.31 <i>p</i> ≥ 0.05	<i>F</i> = 2.19 <i>p</i> ≥ 0.05
TPM	3	<i>F</i> = 1.52 <i>p</i> ≥ 0.05	<i>F</i> = 1.30 <i>p</i> ≥ 0.05	<i>F</i> = 1.06 <i>p</i> ≥ 0.05	<b><i>F</i> = 5.28 <i>p</i> ≤ 0.05*</b>
LVT	2	<i>F</i> = 1.11 <i>p</i> ≥ 0.05	<i>F</i> = 0.64 <i>p</i> ≥ 0.05	<i>F</i> = 0.41 <i>p</i> ≥ 0.05	<i>F</i> = 2.14 <i>p</i> ≥ 0.05
SRT	4	<i>F</i> = 0.89 <i>p</i> ≥ 0.05	<i>F</i> = 0.80 <i>p</i> ≥ 0.05	<i>F</i> = 0.33 <i>p</i> ≥ 0.05	<b><i>F</i> = 6.28 <i>p</i> ≤ 0.05*</b>

Statistical significance was determined using a one-way ANOVA. Text in bold indicates statistical significant effect with *p* < 0.05\*.

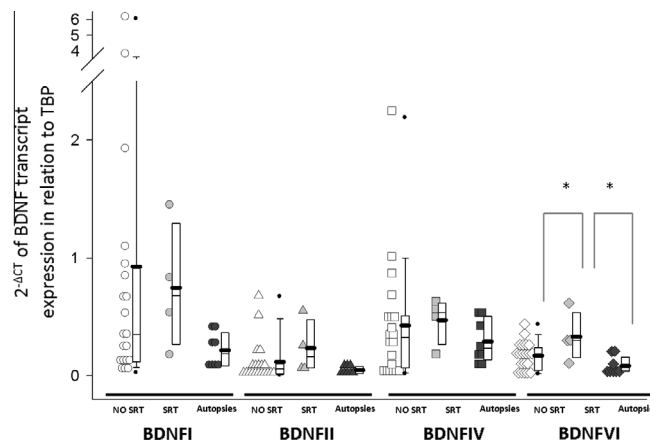
cated in the development of epilepsy while involvement in maintenance of the epileptic state remains uncertain.

The analysis of the expression of different BDNF transcripts is of interest since they show distinct spatial segregation within the neuron (Pattabiraman et al., 2005; Aliaga et al., 2009). It has been proposed that differential sorting of BDNF mRNAs to specific sub-cellular

compartments may modify the synaptic architecture, with possible implications for epilepsy (Tongiorgi et al., 2006). Remarkably, in animal model studies, BDNFVI increases in the distal dendrites a few hours after SE (Pattabiraman et al., 2005; Chiaruttini et al., 2008; Baj et al., 2013). It has been speculated that in this cellular location BDNF mRNAs may increase the production of pro-BDNF,



**Fig. 4.** Effect of the use of Topiramate (TPM) as an anti-epileptic drug previous to surgery on the hippocampal expression of the four BDNF transcripts analyzed. No TPM: no use of Topiramate; TPM: use of Topiramate. Each symbol corresponds to individual data. Circles represent the expression of BDNFI, triangles BDNFII, square BDNFIV and diamonds BDNFVI. Those cases that used (or not) TPM are represented by light gray or open symbols respectively. Dark gray symbols correspond to autopsies. The mean is described as a line in between the box-plot. Statistical significance was determined using a one-way ANOVA followed by T3 Dunnett's multiple comparison test. \* $p < 0.05$ .



**Fig. 5.** Effect of the use of Sertraline (SRT) as an antidepressant previous to surgery, on the expression of the four BDNF transcripts analyzed. No SRT: no use of Sertraline; SRT: use of Sertraline. Each symbol corresponds to individual data. Circles represent the expression of BDNFI, triangles BDNFII, square BDNFIV and diamonds BDNFVI. Those cases that used (or not) SRT are represented by light gray or open symbols respectively. Dark gray symbols correspond to autopsies. The mean is described as a line in between the box-plot. Statistical significance was determined using a one-way ANOVA followed by T3 Dunnett's multiple comparison test. \* $p < 0.05$ .

activate pro-apoptotic signaling cascades, potentiate excitatory synapses as well as increase synaptogenesis and dendrite arborization, eventually promoting epileptic seizures (Chiaruttini et al., 2008).

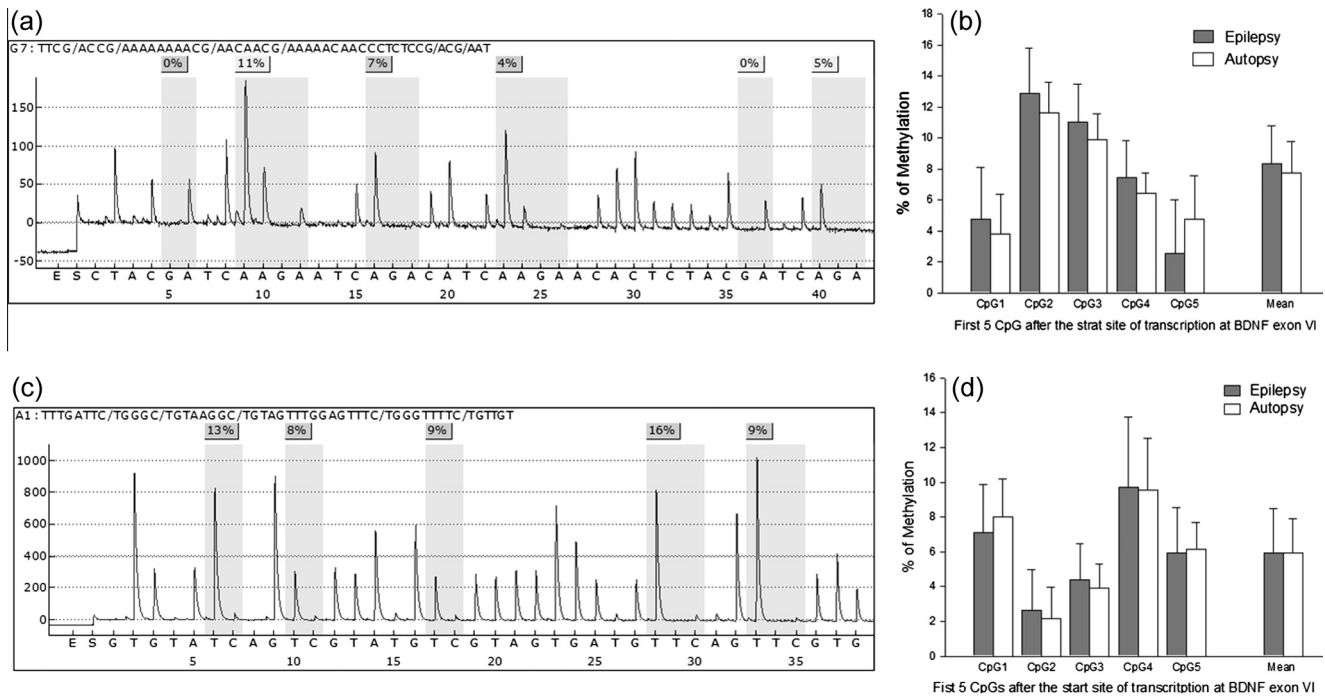
Even though Wang et al. (2011) previously reported an increase in BDNF exons II, IV and VI in TLE-HS patients as compared to those non-HS; in the present study HS did not affect the expression of the different BDNF transcripts analyzed; in contrast lack of association between BDNF protein levels and HS was also previously reported (Hou et al., 2010). Use of different endogenous controls in quantitative PCR (as described in the methodology) and/or the clinical heterogeneity of patients could also explain the lack of agreement among studies.

Variables such as age, gender, years with epilepsy, seizure onset age, number of seizures per day previous to surgery and the laterality of resection did not modify the results of BDNF expression. However, a decrease in BDNFVI was detected in three patients using TPM, a

drug with histone deacetylase inhibitor (HDACi) properties. An analogous dose-dependent decline of the expression of this neurotrophin was previously reported in rats' hippocampus chronically treated with TPM (Shi et al., 2010). Moreover, BDNF protein levels also decrease in hippocampal slices obtained from patients with pharmaco-resistant TLE after *in vitro* infusion of another HDACi: VPA (Hou et al., 2010). Lack of effect of VPA on BDNF mRNA levels observed in the present study could be due to the suspension of this drug two weeks before surgery, in order to prevent the risk of hemorrhage during operation.

We also found an increment in the expression of BDNFVI in hippocampus of four patients that used the serotonin reuptake inhibitor (SRI) SRT to treat anxiety and depression symptoms. A similar result was observed when the immunoreactivity for this neurotrophin was measured in TLE patients' hippocampus using fluoxetine (Kandratavicius et al.,





**Fig. 6.** Comparison of the methylation levels at the first five CpGs located after the transcription start site at exons IV and VI, between TLE cases and autopsies. Left panel: representative pyrograms of methylation levels at the BDNF promoter IV (a) and VI (c). Right panel: methylation levels at the BDNF promoter IV (b) and VI (d), between TLE cases (gray bars) and autopsies (white bars). Statistical significance was determined using an unpaired, two-tailed *t*-test.

2013). In addition, in depressive patients a rise in serum BDNF protein and mRNA BDNF has been associated to the use of a variety of SRI's (Molendijk et al., 2014; Hong et al., 2014). Interestingly, in animal models it has been reported that anti-depressive treatment as well as exercise specifically increase BDNFVI, promoting the formation of new distal dendrites in the CA3 neurons (Baj et al., 2012). This effect could enhance the communication with the dentate gyrus promoting the hippocampus' functional recovery (Baj et al., 2011, 2012).

The effect of the pharmacological treatments on the expression of BDNF transcripts has barely been studied, especially in epileptic tissue, and our results suggest that they probably affect its mRNA levels. Nonetheless, due to the discrete number of individuals taking these drugs, caution should be taken when interpreting the association related to the use of TPM or SRT. Future research including higher number of samples is required in order to further discriminate between the effects dependent on pharmacological treatment from those associated with epilepsy.

Finally, different studies have reported an inverse relationship between BDNF expression and methylation levels in specific promoters (Martínez-Levy and Cruz-Fuentes, 2014). Nonetheless, under our experimental conditions, the analyzed regions near the start sites of transcription at exons IV and VI were completely unmethylated in autopsies and cases. Other mechanisms like the intra-nuclear position of BDNF gene (Walczak et al., 2013), histone modification (Tian et al., 2009) or changes in the expression of specific transcription factors, such as CREB (Guo et al., 2014) must be considered in

order to understand BDNF expression regulation and its role in epilepsy.

## CONCLUSION

The role of the different BDNF transcripts in epilepsy is still controversial. Even though, previous studies have analyzed the steady-state expression of BDNF gene in relation to SE, reports in epileptic tissue are scarce. To our knowledge this is the first study that analyzed these variants in epileptic tissue of TLE patients in comparison to autopsies. Changes in the expression of specific BDNF transcripts, as reported here could have functional implication in epilepsy or could be related to the use of other pharmacological treatments. Future *in vitro* and animal model studies are warranted to further investigate these possibilities.

*Acknowledgments*—This article constitutes a partial requirement to obtain the PhD grade in the Postgraduate program of Biological Sciences at the National Autonomous University of Mexico (UNAM) for GAM-L. This study was supported the Research fund of the National Institute of Psychiatry “Ramón de la Fuente Muñiz” Project IC142040.0. We would like to thank Georgina Guerrero Avendaño, Gianelli Cortes Gonzalez and Jose Perez Luna for their technical help.

## REFERENCES

- Aliaga EE, Mendoza I, Tapia-Arancibia L (2009) Distinct subcellular localization of BDNF transcripts in cultured hypothalamic neurons and modification by neuronal activation. *J Neural Transm* 116 (1):23–32.

- Baj G, Leone E, Chao MV, Tongiorgi E (2011) Spatial segregation of BDNF transcripts enables BDNF to differentially shape distinct dendritic compartments. *Proc Natl Acad Sci USA* 108 (40):16813–16818.
- Baj G, D'Alessandro V, Musazzi L, Mallei A, Sartori CR, Sciancalepore M, Tardito D, Langone F, Popoli M, Tongiorgi E (2012) Physical exercise and antidepressants enhance BDNF targeting in hippocampal CA3 dendrites: further evidence of a spatial code for BDNF splice variants. *Neuropsychopharmacology* 37(7):1600–1611.
- Baj G, Del Turco D, Schlaudraff J, Torelli L, Deller T, Tongiorgi E (2013) Regulation of the spatial code for BDNF mRNA isoforms in the rat hippocampus following pilocarpine-treatment: a systematic analysis using laser microdissection and quantitative real-time PCR. *Hippocampus* 23(5):413–423.
- Casillas-Espinosa PM, Powell KL, O'Brien TJ (2012) Regulators of synaptic transmission: roles in the pathogenesis and treatment of epilepsy. *Epilepsia* 53:41–58.
- Chiaruttini C, Sonogo M, Baj G, Simonato M, Tongiorgi E (2008) BDNF mRNA splice variants display activity-dependent targeting to distinct hippocampal laminae. *Mol Cell Neurosci* 37(1):11–19.
- Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162(1):156–159.
- Croll SD, Suri C, Compton DL, Simmons MV, Yancopoulos GD, Lindsay RM, Wiegand SJ, Rudge JS, Scharfman HE (1999) Brain-derived neurotrophic factor transgenic mice exhibit passive avoidance deficits, increased seizure severity and in vitro hyperexcitability in the hippocampus and entorhinal cortex. *Neuroscience* 93(4):1491–1506.
- Cunha C, Brambilla R, Thomas KL (2010) A simple role for BDNF in learning and memory? *Front Mol Neurosci* 3:1–14.
- D'Addario C, Caputi FF, Ekström TJ, Di Benedetto M, Maccarrone M, Romualdi P, Candeletti S (2013) Ethanol induces epigenetic modulation of prodynorphin and pronociceptin gene expression in the rat amygdala complex. *J Mol Neurosci* 49:312–319.
- Engel J, Williamson PD, Wieser H-G (1997). In: Engel J, Pedley TA, editors. Chapter 231: mesial temporal lobe epilepsy in epilepsy: a comprehensive text book. Philadelphia: Lippincott-Raven Publishers. p. 2417–2424.
- First Michael B, Spitzer Robert L, Gibbon Miriam, Williams Janet BW (2002) Structured clinical interview for DSM-IV-TR axis I disorders, research version, patient edition. (SCID-I/P) New York: Biometrics Research, New York State Psychiatric Institute.
- Garzon DJ, Fahnstock M (2007) Oligomeric amyloid decreases basal levels of brain-derived neurotrophic factor (BDNF) mRNA via specific downregulation of BDNF transcripts IV and V in differentiated human neuroblastoma cells. *J Neurosci* 27 (10):2628–2635.
- Goldenberg MM (2010) Overview of drugs used for epilepsy and seizures. *P&T* 35(7):392–415.
- Gómez-Arias B, Crail-Meléndez D, López-Zapata R, Martínez-Juárez IE (2012) Severity of anxiety and depression are related to a higher perception of adverse effects of antiepileptic drugs. *Seizure* 21:588–594.
- Guidotti A, Grayson DR (2014) DNA methylation and demethylation as targets for antipsychotic therapy. *Dialogues Clin Neurosci* 16:419–429.
- Guo J, Wang H, Wang Q, Chen Y, Chen S (2014) Expression of p-CREB and activity-dependent miR-132 in temporal lobe epilepsy. *Int J Clin Exp Med* 7:1297–1306.
- Heinrich C, Lähtinen S, Suzuki F, Anne-Marie L, Huber S, Häussler U, Haas C, Larmet Y, Castren E, Depaulis A (2011) Increase in BDNF-mediated TrkB signaling promotes epileptogenesis in a mouse model of mesial temporal lobe epilepsy. *Neurobiol Dis* 42 (1):35–47.
- Herrero MJ, Blanch J, Peri JM, De Pablo J, Pintor L, Bulbena A (2003) A validation study of the hospital anxiety and depression scale (HADS) in a Spanish population. *Gen Hosp Psychiatry* 25:277–283.
- Hong W, Fan J, Yuan C, Zhang C, Hu Y, Peng D, Wang Y, Huang J, Li Z, Yu S, Liu X, Wu Z, Chen J, Yi Z, Xu L, Fang Y (2014) Significantly decreased mRNA levels of BDNF and MEK1 genes in treatment-resistant depression. *Neuroreport* 25(10):753–755.
- Hou X, Wang X, Zhang L (2010) Conditional downregulation of brain derived neurotrophic factor and tyrosine kinase receptor B blocks epileptogenesis in the human temporal lobe epilepsy hippocampus. *Neurol India* 58(1):29–34.
- Jardim AP, Neves RS, Caboclo LO, Lancellotti CL, Marinho MM, Centeno RS, Cavalheiro EA, Scorza CA, Yacubian EM (2012) Temporal lobe epilepsy with mesial temporal sclerosis: hippocampal neuronal loss as a predictor of surgical outcome. *Arq Neuropsiquiatr* 70(5):319–324.
- Kandravicius L, Monteiro MR, Assirati Jr JA, Carlotti Jr CG, Hallak JE, Leite JP (2013) Neurotrophins in mesial temporal lobe epilepsy with and without psychiatric comorbidities. *J Neuropathol Exp Neurol* 72(11):1029–1042.
- Keller S, Sarchiapone M, Zarrilli F, Videtic A, Ferraro A, Carli V, Sacchetti S, Lembo F, Angiolillo A, Jovanovic N, Pisanti F, Tomaiuolo R, Monticelli A, Balazic J, Roy A, Marusic A, Cocozza S, Fusco A, Bruni CB, Castaldo G, Chiariotti L (2010) Increased BDNF promoter methylation in the Wernicke area of suicide subjects. *Arch Gen Psychiatry* 67(3):258–267.
- Koyama R, Ikegaya Y (2005) To BDNF or not to BDNF: that is the epileptic hippocampus. *Neuroscientist* 11(4):282–287.
- Lee S, Williamson J, Lothman EW, Szele FG, Chesselet MF, Von Hagen S, Sapolsky RM, Mattson MP, Christakos S (1997) Early induction of mRNA for calbindin-D28k and BDNF but not NT-3 in rat hippocampus after kainic acid treatment. *Brain Res Mol Brain Res* 47(1–2):183–194.
- Liu G, Gu B, He XP, Joshi RB, Wackerle HD, Rodriguiz RM, Wetsel WC, McNamara JO (2013) Transient inhibition of TrkB kinase after status epilepticus prevents development of temporal lobe epilepsy. *Neuron* 79(1):31–38.
- Martínez-Levy GA, Cruz-Fuentes CS (2014) Genetic and epigenetic regulation of the brain-derived neurotrophic factor in the central nervous system. *Yale J Biol Med* 87:173–186.
- Mather GW, Babb TL, Micevych PE, Blanco CE, Pretorius JK (1997) Granule Cell mRNA levels for BDNF, NGF, and NT-3 correlate with neuron losses or supragranular mossy fiber sprouting in the chronically damaged and epileptic human hippocampus. *Mol Chem Neuropathol* 30:53–76.
- Molendijk ML, Spinhoven P, Polak M, Bus BA, Penninx BW, Elzinga BM (2014) Serum BDNF concentrations as peripheral manifestations of depression: evidence from a systematic review and meta-analyses on 179 associations (N = 9484). *Mol Psychiatry* 19:791–800.
- Mudò G, Jiang XH, Timmusk T, Bindoni M, Belluardo N (1996) Change in neurotrophins and their receptor mRNAs in the rat forebrain after status epilepticus induced by pilocarpine. *Epilepsia* 37(2):198–207.
- Murray KD, Isackson PJ, Eskin TA, King MA, Montesinos SP, Abraham LA, Roper SN (2000) Altered mRNA expression for brain-derived neurotrophic factor and type II calcium/calmodulin-dependent protein kinase in the hippocampus of patients with intractable temporal lobe epilepsy. *J Comp Neurol* 418(4):411–422.
- Nelson ED, Kavalali ET, Monteggia LM (2008) Activity-dependent suppression of miniature neurotransmission through the regulation of DNA methylation. *J Neurosci* 28:395–406.
- Paradiso B, Marconi P, Zucchini S, Berto E, Binaschi A, Bozac A, Buzzi A, Mazzuferi M, Magri E, Navarro Mora G, Rodi D, Su T, Volpi I, Zanetti L, Marzola A, Manservigi R, Fabene PF, Simonato M (2009) Localized delivery of fibroblast growth factor-2 and brain-derived neurotrophic factor reduces spontaneous seizures in an epilepsy model. *Proc Natl Acad Sci USA* 106 (17):7191–7196.
- Paradiso B, Zucchini S, Su T, Bovolenta R, Berto E, Marconi P, Marzola A, Navarro Mora G, Fabene PF, Simonato M (2011) Localized overexpression of FGF-2 and BDNF in hippocampus reduces mossy fiber sprouting and spontaneous seizures up to 4

- weeks after pilocarpine-induced status epilepticus. *Epilepsia* 52:572–578.
- Parrish R, Albertson AJ, Buckingham SC, Hablitz JJ, Mascia KL, Davis Haselden W, Lubin FD (2013) Status epilepticus triggers early and late alterations in brain-derived neurotrophic factor and NMDA glutamate receptor Grin2b DNA methylation levels in the hippocampus. *Neuroscience* 17(248):602–619.
- Pattabiraman PP, Tropea D, Chiaruttini C, Tongiorgi E, Cattaneo A, Domenici L (2005) Neuronal activity regulates the developmental expression and subcellular localization of cortical BDNF mRNA isoforms in vivo. *Mol Cell Neurosci* 28(3):556–570.
- Pruunsild P, Kazantseva A, Aid T, Palm K (2007) Dissecting the human BDNF locus: bidirectional transcription, complex splicing and multiple promoters. *Genomics* 90(3):397–406.
- San-Juan D, Tapia CA, González-Aragón MF, Martínez Mayorga A, Staba RJ, Alonso-Vanegas M (2011) The prognostic role of electrocorticography in tailored temporal lobe surgery. *Seizure* 20:564–569.
- Santos AR, Duarte CB (2008) Validation of internal control genes for expression studies: effects of the neurotrophin BDNF on hippocampal neurons. *J Neurosci Res* 86:3684–3692.
- Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* 3(6):1101–1108.
- Shetty AK, Zaman V, Shetty GA (2003) Hippocampal neurotrophin levels in a kainate model of temporal lobe epilepsy: a lack of correlation between brain-derived neurotrophic factor content and progression of aberrant dentate mossy fiber sprouting. *J Neurochem* 87(1):147–159.
- Shi XY, Wang JW, Cui H, Li BM, Lei GF, Sun RP (2010) Effects of antiepileptic drugs on mRNA levels of BDNF and NT-3 and cell neogenesis in the developing rat brain. *Brain Dev* 32(3):229–235.
- Simonato M, Tongiorgi E, Kokaia M (2006) Angels and demons: neurotrophic factors and epilepsy. *Trends Pharmacol Sci* 27(12):631–638.
- Takahashi M, Hayashi S, Kakita A, Wakabayashi K, Fukuda M, Kameyama S, Tanaka R, Takahashi H, Nawa H (1999) Patients with temporal lobe epilepsy show an increase in brain-derived neurotrophic factor protein and its correlation with neuropeptide Y. *Brain Res* 818:579–582.
- Tian F, Hu XZ, Wu X, Jiang H, Pan H, Marini AM, Lipsky RH (2009) Dynamic chromatin remodeling events in hippocampal neurons are associated with NMDA receptor-mediated activation of BDNF gene promoter 1. *J Neurochem* 109:1375–1388.
- Tongiorgi E, Armellini M, Giulianini PG, Bregola G, Zucchini S, Paradiso B, Steward O, Cattaneo A, Simonato M (2004) Brain-derived neurotrophic factor mRNA and protein are targeted to discrete dendritic laminae by events that trigger epileptogenesis. *J Neurosci* 24(30):6842–6852.
- Tongiorgi E, Domenici L, Simonato M (2006) What is the biological significance of BDNF mRNA targeting in the dendrites? Clues from epilepsy and cortical development. *Mol Neurobiol* 33(1):17–32.
- Walczak A, Szczepankiewicz AA, Ruszczycki B, Magalska A, Zamlynska K, Dzwonek J, Wilczek E, Zybura-Broda K, Ryłski M, Malinowska M, Dabrowski M, Szczepinska T, Pawłowski K, Pyskaty M, Włodarczyk J, Szczerbal I, Switonski M, Cremer M, Wilczynski GM (2013) Novel higher-order epigenetic regulation of the BDNF gene upon seizures. *J Neurosci* 33:2507–2511.
- Wang FJ, Li CM, Hou XH, Wang XR, Zhang LM (2011) Selective upregulation of brain-derived neurotrophic factor (BDNF) transcripts and BDNF direct induction of activity independent N-methyl-D-aspartate currents in temporal lobe epilepsy patients with hippocampal sclerosis. *J Int Med Res* 39(4):1358–1368.
- Wierschke S, Gigout S, Horn P, Lehmann TN, Dehnicke C, Bräuer AU, Deisz RA (2010) Evaluating reference genes to normalize gene expression in human epileptogenic brain tissues. *Biochem Biophys Res Commun* 403:385–390.
- Wong J, Hyde TM, Cassano HL, Deep-Soboslay A, Kleinman JE, Weickert CS (2010) Promoter specific alterations of brain-derived neurotrophic factor mRNA in schizophrenia. *Neurosci* 169:1071–1084.
- Xu B, Michalski B, Racine RJ, Fahnestock M (2004) The effects of brain-derived neurotrophic factor (BDNF) administration on kindling induction, Trk expression and seizure-related morphological changes. *Neuroscience* 126(3):521–531.

(Accepted 19 November 2015)  
(Available online 24 November 2015)



# A novel chromatin insulator regulates the chicken folate receptor gene from the influence of nearby constitutive heterochromatin and the $\beta$ -globin locus

Edgar González-Buendía<sup>a</sup>, Martín Escamilla-Del-Arenal<sup>a,c</sup>, Rosario Pérez-Molina<sup>a</sup>, Juan J. Tena<sup>b</sup>, Georgina Guerrero<sup>a</sup>, Fernando Suaste-Olmos<sup>a</sup>, Erandi Ayala-Ortega<sup>a</sup>, José Luis Gómez-Skarmeta<sup>b</sup>, Félix Recillas-Targa<sup>a,\*</sup>

<sup>a</sup> Instituto de Fisiología Celular, Departamento de Genética Molecular, Universidad Nacional Autónoma de México, Ciudad de México, D.F., Mexico

<sup>b</sup> Centro Andaluz de Biología del Desarrollo, CSIC-Universidad Pablo de Olavide, Campus UPO, Sevilla, Spain

<sup>c</sup> Department of Biochemistry and Molecular Biophysics, Mortimer B. Zuckerman Mind, Brain, and Behavior Institute, Columbia University, NY, USA

## ARTICLE INFO

### Article history:

Received 6 March 2015

Received in revised form 22 May 2015

Accepted 25 May 2015

Available online 14 June 2015

### Keywords:

Folate receptor and globin genes

Transcription

Heterochromatin

Chromatin insulator

DNA methylation

CTCF

## ABSTRACT

The three-dimensional architecture of genomes provides new insights about genome organization and function, but many aspects remain unsolved at the local genomic scale. Here we investigate the regulation of two erythroid-specific loci, a folate receptor gene (*FOLR1*) and the  $\beta$ -globin gene cluster, which are separated by 16 kb of constitutive heterochromatin. We found that in early erythroid differentiation the *FOLR1* gene presents a permissive chromatin configuration that allows its expression. Once the transition to the next differentiation state occurs, the heterochromatin spreads into the *FOLR1* domain, concomitant with the dissociation of CTCF from a novel binding site, thereby resulting in irreversible silencing of the *FOLR1* gene. We demonstrate that the sequences surrounding the CTCF-binding site possess classical insulator properties *in vitro* and *in vivo*. In contrast, the chicken *CHS4*  $\beta$ -globin insulator present on the other side of the heterochromatic segment is in a constitutive open chromatin configuration, with CTCF constantly bound from the early stages of erythroid differentiation. Therefore, this study demonstrates that the 16 kb of constitutive heterochromatin contributes to silencing of the *FOLR1* gene during erythroid differentiation.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

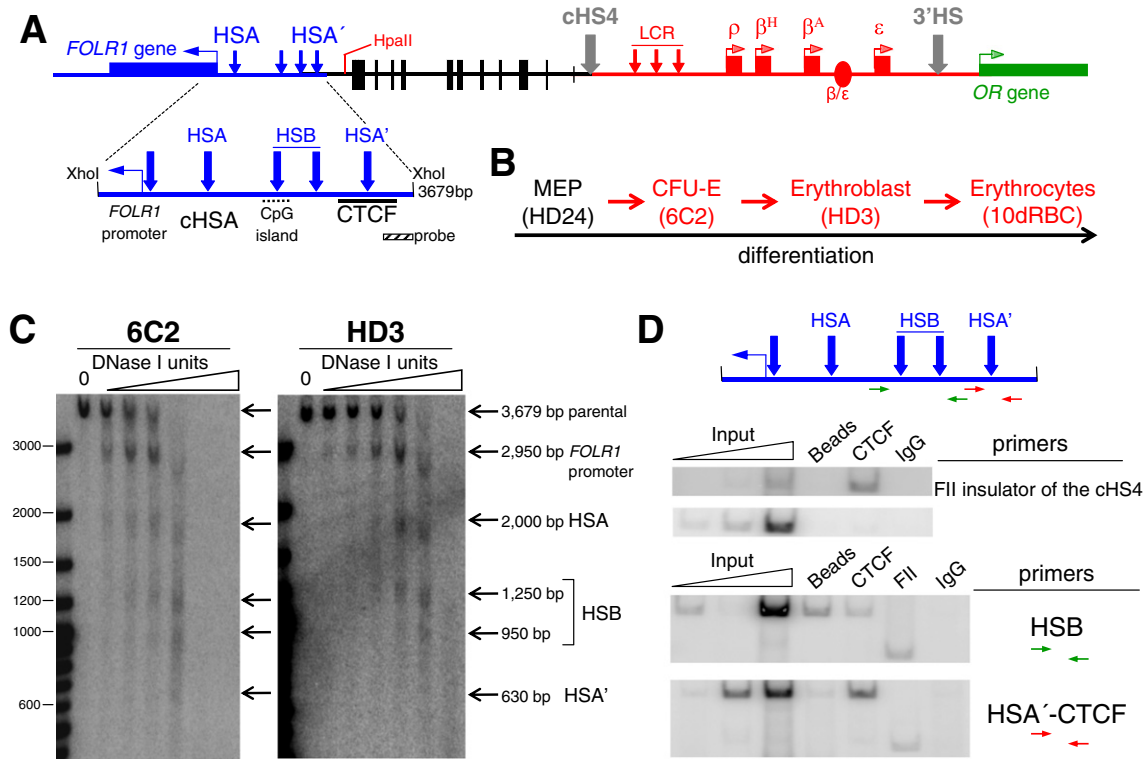
Chromatin domains are defined by an alternation of genomic regions enriched in repressive or active chromatin structures [1,2]. Based on the chromosome conformation capture technique (also known as 3C) and its derivatives, many large-scale genomic features have been proposed, including Topological Associated Domains or TADs [3,4]. These genome-wide studies have also contributed to a better understanding of the three-dimensional organization of the genome inside the nucleus but they involve considerably more distance than the local scale. In such lower-scale genomic resolution studies there are still many aspects to be understood, particularly from the perspective of how the genome is partitioned and regulated into transcriptionally active domains. Here we analyze the organization and dynamics of a paradigmatic group of chromatin domains associated with the chicken  $\beta$ -globin locus.

Historically, this locus attracted the attention of several research groups due to its differential regulation of gene expression but, more recently, by the early description of a transition between DNase I-resistant and hypoacetylated histones [5]. We now know that such a transition coincides with the location of the chicken *CHS4*  $\beta$ -globin insulator [6–8]. Based on those seminal works, the early concept of transcriptionally active domains was initially exemplified by the description of the chicken folate receptor (*FOLR1*) gene, followed by approximately 16 kilobases (kb) of condensed chromatin and the  $\beta$ -globin locus separated by the *CHS4* boundary element [9] (Fig. 1A). Interestingly, this compacted 16 kb genomic region possesses a significant amount of CR1-type repetitive sequences, histone repressive marks and notably, the 16 kb region is DNA-hypermethylated, which allows it to be recovered by sucrose gradient sedimentation when digested *in situ* with the *HpaII* restriction enzyme [9–11]. Furthermore, as additional evidence of its heterochromatic features, this region is transcribed and acquires a closed conformation in a Dicer-dependent manner [12]. Together, these observations strongly suggest that the 16 kb region of heterochromatin, which is located between the erythroid-specific folate receptor gene and the  $\beta$ -globin locus, represents a compact and very well-defined unit of constitutive heterochromatin (Fig. 1A).

\* Corresponding author at: Instituto de Fisiología Celular, Departamento de Genética Molecular, Universidad Nacional Autónoma de México, Apartado Postal 70-242, México, D.F. 04510, Mexico.

E-mail address: [frecilla@ifc.unam.mx](mailto:frecilla@ifc.unam.mx) (F. Recillas-Targa).





**Fig. 1.** Folate receptor gene expression and associated chromatin structure is restricted to an early erythroid differentiation stage. (A) Scheme of the genomic region encompassing the chicken folate receptor (*FOLR1*) gene and regulatory-related DNase I hypersensitive sites (HSA, HSB and HSA'); the 16 kb of constitutive heterochromatin with vertical black lines representing CR1-type repetitive sequences and the DNA methylated HpaII site (shown in red) correspond to the 5' margin to the heterochromatic region [11]; the  $\beta$ -globin locus shielded by two chromatin insulators, cHS4 and 3'HS bound by CTCF [8,11], and the 3'-most domain represented by the brain-specific olfactory receptor gene [43]. A detailed map of the intergenic region (with the distribution of the DHSs) located between the *FOLR1* gene and the 16 kb of heterochromatin is shown. (B) Erythroid cell-types used in this study and their differentiation stages are shown. In black (HD24 cells) correspond to a myeloid/erythroid progenitor cell line and in red, erythroid cell lines at different stages of differentiation. (C) Comparative DNase I hypersensitive assays using nuclei isolated from 6C2 and HD3 cells. The DNase I units used for nuclei digestion are: 0.3; 0.7; 1.25; 2.5; 5 and 10. The probe location used for hybridization is shown in A. (D) Chromatin immunoprecipitation (ChIP) assay showing the *in vivo* binding of CTCF using, as a positive control, primers from the cHS4  $\beta$ -globin insulator centered at the Fil CTCF-binding site. Primers used to define the binding of CTCF at the HSA' site are described in Supplementary Fig. S3. Note the lack of enrichment at the HSB site, a result that is in concordance with previously published data [29]. This is a representative set of results from four independent ChIP assays.

From the literature several lines of evidence have shown that constitutive heterochromatin can have a negative effect on the expression of genes that are translocated in its proximity, through a phenomenon known as position effect variegation [13]. It has been mechanistically demonstrated that certain histone repressive marks, like histone H3 lysine 9 trimethylation (H3K9me3) or even DNA methylation, can spread from a heterochromatic source over several kilo bases [14]. Therefore, it has been proposed that the chicken cHS4 insulator possesses a barrier function to counteract the spreading of the heterochromatin over the  $\beta$ -globin locus [15–17]. Here we asked whether, as in the case of the  $\beta$ -globin locus (Fig. 1A), the folate receptor gene, as an independent domain, needs to be protected by a barrier element against the silencing effects of the nearby heterochromatin.

This locus encodes an erythroid-specific folate receptor with a different and early program of gene expression compared to the chicken  $\beta$ -globin group of genes [9]. The chicken folate receptor gene is transcribed in early erythroid precursor cells and its expression is repressed during terminal erythroid differentiation [9]. In other words, the folate receptor pattern of gene expression precedes globin gene activation, and when the globin genes are transcriptionally activated, the folate receptor gene is already silenced [9,10]. A strong DNase I-hypersensitive site (DHS) designated HSA shows an apparent enhancer activity over the *FOLR1* gene, with an activity over a reporter gene that is copy-number dependent and independent of the integration site in stably transformed erythroid cells [9]. With all this in mind we asked if a chromatin insulator is present between the HSA enhancer and the delimiting HpaII site which represents the 5' boundary of the 16 kb of heterochromatin (Fig. 1A). At this point it is important to outline that

the previously reported work has described the existence of a boundary at the 5' end of the condensed chromatin region, protecting against the encroachment of histone modifications in particular histone monoubiquitination [18]. We hypothesized that such an insulator may allow the *FOLR1* gene to be expressed based on an optimal chromatin configuration, and in an erythroid-specific stage of cellular differentiation.

To address this aim, a comparative DNase I hypersensitivity assay showed different and unobserved sites between the *FOLR1* HSA and the boundary of the 16 kb of heterochromatin. We found a novel CTCF binding site with chromatin insulator properties. Interestingly, when we compared the two erythroid stages represented by 6C2 and HD3 cells, we found a differential pattern of DNA methylation and binding of CTCF. Together, these results support the presence of CTCF and regulated boundary activity when the folate receptor gene needs to be expressed. Therefore, we have characterized a novel boundary element that acts in response to different stages of erythroid differentiation.

## 2. Materials and methods

### 2.1. Plasmid constructs

DNA fragments containing the HSB–HSA' insulator was amplified by PCR from genomic DNA from 10-day-old chicken embryonic erythrocytes with the following primers: Forward: 5'-CCAGACACACTGCTCCCAC-3' and Reverse: 5'-GGCATCCATGGAAAAGGCTGC-3', then cloned into pG $\alpha^D$ 3 at the EcoRI 5' site and NheI–MluI at the 3' site flanking the  $\alpha^D$  promoter and EGFP cDNA, respectively. A DNA subfragment containing 6.7 kb of the 16 kb heterochromatin region was

subcloned from pSKHetero8.7 (kindly provided by M.N. Prioleau and G. Felsenfeld) into pG $\alpha^D3$  in the BglIII–EcoRI sites.

## 2.2. Cell lines and cell culture

HD3 cells represent a *Gallus gallus* erythroblast AEV virus-transformed cell line that was grown at 37°C, 5% CO<sub>2</sub>, in Dulbecco's Modified Eagle's Medium (DMEM, GIBCO) supplemented with 10% Fetal Bovine Serum, 2% chicken serum and antibiotics (100 U Penicillin/0.1 mg Streptomycin/ml) (Fig. 1B and Supplementary Fig. S2). 6C2 cells represent a *G. gallus* pre-erythroblast (CFU-E) AEV virus-transformed cell line that was grown at 37°C, 5% CO<sub>2</sub>, in  $\alpha$ -Minimum Essential Medium ( $\alpha$ -MEM, GIBCO) supplemented with 10% Fetal Bovine Serum, 2% chicken serum, 1 mM HEPES (pH 7.2, SIGMA), 50 mM  $\beta$ -mercaptoethanol and antibiotics (100 U Penicillin/0.1 mg Streptomycin/ml). HD24 cells represent a *G. gallus* multipotent erythroid–myeloid E26 virus-transformed cell line that was grown in Blastoderm medium [19]. DT40 cells represent a *G. gallus* leukosis ALV virus-induced bursal lymphoma cell line that was grown at 37°C, 5% CO<sub>2</sub>, in Dulbecco's Modified Eagle's Medium (DMEM, GIBCO), with 2 mM L-glutamine and 50 mM  $\beta$ -mercaptoethanol, 10% tryptose phosphate broth, 10% Fetal Bovine Serum, 5% chicken serum and antibiotics (100 U Penicillin/0.1 mg Streptomycin/ml). Five- and 10-day erythrocytes (5dRBCs and 10dRBCs) were obtained by bleeding embryos from fertilized chicken eggs (ALPES Farm, Tehuacán, Puebla). The K562 cells used in the colony assay represent a human erythroleukemic cell line which was grown at 37°C, 5% CO<sub>2</sub>, in RPMI medium with 2 mM glutamine, 10% Fetal Bovine Serum and antibiotics (100 U Penicillin/0.1 mg Streptomycin/ml).

## 2.3. Transient and stable transfections

$1 \times 10^6$  cells were resuspended in 800  $\mu$ l of medium without antibiotics and Fetal Bovine Serum, then plated in a 6-well plate, then transfected with the plasmid and lipofectamine 2000 according to the manufacturer's instructions. After 48 h of cell recuperation, cells were collected: a) transient transfections were analyzed for *Luciferase* activity using *Renilla* as a normalizer, or b) selected by antibiotics for stable integration.

## 2.4. Reverse transcription (RT)-quantitative PCR

Total RNA was extracted from indicated cells with TRIzol Reagent (Invitrogen) according to the manufacturer's instructions, and treated with DNase I (RQ1, Promega). cDNA was generated from 5  $\mu$ g DNase I-treated RNA (Reverse Transcription System, Promega) using oligo-dT. This was followed by PCR amplification using radioactively labeled [ $\alpha$ -<sup>32</sup>P]dCTP-specific primers for each indicated product (*FOLR1* Forward: 5'-CCAGTTCCTTTGCATGTGACG-3' and Reverse: 5'-GCAAGCGTCACTTCATCCAGG-3') and  $\beta$ -actin as an endogenous normalization control (Forward: 5'-CCAGACATCAGGGTGTGATG-3' and Reverse: 5'-GAACACGGTATTGTACCAACTGG-3'). For quantitative PCR primers were intronspanning, Platinum Taq DNA polymerase (Invitrogen) was used mixed with SYBR-Green I (Sigma) dissolved in DMSO. The qPCR reactions were performed in a StepOne™ detection system (Applied Biosystems). The relative expression levels of gene expression were determined by using actin mRNA levels as an endogenous reference and the  $\Delta\Delta$ CT method was used for normalization.

## 2.5. Chromatin immunoprecipitation (ChIP)-quantitative PCR

The ChIP assay was performed as previously reported [19]. Briefly,  $5 \times 10^7$  cells were cross-linked with 1% formaldehyde, sonicated and 4  $\mu$ g of the new and purified stock antibody against chicken CTCF was used (own-generated Antibody) [20]. Immunoprecipitated DNA was evaluated by duplex PCR radioactively labeled with

[ $\alpha$ -<sup>32</sup>P]dCTP with specific primers for each indicated product (HSA'-CTCF, Forward: 5'-ACGCCTTTGCTTCGTGGTATAG-3' and Reverse: 5'-GCCCCCTCCAATATACTGCAGG-3'; FII-CTCF Forward: 5'-GGTTGAAGAAAGAAGCAGGC-3' and Reverse: 5'-CGTTCAGACGAAAGCGATCC-3') and using H1 (constitutive heterochromatin region, Forward: 5'-GGAAGTGTGGCAAGGTCCTCT-3' and Reverse: 5'-TCTTCTGCCCTGCCGAT-3') as an endogenous normalization control. For quantitative real time PCR (qPCR), Platinum Taq DNA polymerase (Invitrogen) was used mixed with SYBR-Green I (Sigma) dissolved in DMSO. The qPCR reactions were performed in a StepOne™ detection system (Applied Biosystems) and the  $\Delta\Delta$ CT method was used for normalization. Enrichment levels = (exp / control)Ab / (exp / control)IgG.

## 2.6. DNA sodium bisulfite conversion

Genomic DNA was extracted from indicated cells by the phenol-chloroform technique, and 1.5  $\mu$ g was cut with EcoRI prior to bisulfite conversion. Bisulfite conversion was performed as described previously [21]. Specific primers for the converted promoter region were used to generate PCR products (*FOLR1* gene promoter, Forward: 5'-GGGTTTTGGTAGGTATTATTA-3' and Reverse: 5'-CTCTAAAAATAAAATAACTCTC-3'; HSB, Forward: 5'-GGTTTTGGGTAGTTGGGTTG-3' and Reverse: 5'-CCAAAAACAACTATCACAACCT-3'; HSA', Forward: 5'-GGGTATTGTGTATAGGATAGT-3' and Reverse: 5'-ATAAAACAAACACAAATAATCCC-3'; Hetero1, Forward: 5'-GGGATTATTTGTGTTTTAT-3' and Reverse: 5'-CCACACTTAATTAATCTACTAC-3'; Hetero3, Forward: 5'-GTAGGTTTGGTATAATTTTAGTTA-3' and Reverse: 5'-AAAAATAACCTTAAAAACATCC-3'). PCR products were cloned in pGEMT-easy (Promega) followed by sequencing with the Sp6 primer.

## 2.7. The cCTCF shRNAi and Lentivirus generation

shRNAi-mediated knockdowns that were designed in <http://sirna.wi.mit.edu/> [22] target sequence 5'-GCCAGCAGGGATACCTTACA-3' based on the NCBI mRNA reference sequence for chicken CTCF, NM\_205332.4, then cloned into the HpaI–XhoI sites in an LL3.7 expression vector and sequenced. shRNAi-mediated knockdowns were performed essentially as described previously (kindly provided by Joaquín M. Espinosa) [23]. HEK293FT cells were used to produce the pLL3.7 control and chicken CTCF shRNAi-cCTCF1: 5'-GCCAGCAGGGATACCTTACA-3', lentiviruses with generation III packaging vectors (pLL3.7 system was kindly provided by J. Espinosa). 6C2 cells were transduced with virus for 8 h in the presence of polybrene (8  $\mu$ g/ml; Sigma). Cultures were then selected for 3–4 days with puromycin (3  $\mu$ g/ml; Sigma) and then harvested for the experiments detailed in this article. The USF1 knockdown was performed using a mixture of five different USF1-specific RNAi-expressing vectors in the same conditions as described above (kindly provided by Suming Huang, University of Florida, Gainesville, FL) [24].

## 2.8. DNase I hypersensitivity assay

DNase I hypersensitivity assays in 6C2 and HD3 cell lines were performed as previously reported [25].

## 2.9. Protection against chromosomal position effects (CPE) assay

CPE assays were performed as previously described [26]. The stably transfected plasmids were: pG $\alpha^D3$  containing EGFP under the control of the  $\alpha^D$  promoter, pG $\alpha^D3$ HSA4 containing 5'cHS4 (2xcHS4 on each side of  $\alpha^D$ -EGFP), and pG $\alpha^D3$ HSA' (2xHSA' on each side of  $\alpha^D$ -EGFP).

## 2.10. Colony assay and in vivo Enhancer Blocking Assay in zebrafish

The enhancer blocking colony assay was performed essentially as in [6,7]. Briefly,  $1 \times 10^7$  K562 cells were electroporated, pooled and plated

in soft agar in presence of geneticin at 900 mg/ml. Colonies were counted after 3 to 4 weeks of drug selection and the colony number was normalized to the number of colonies obtained with the pJC3–4 plasmids (2.3 kb- $\lambda$ ). The *in vivo* Enhancer Blocking Assay was performed as described [27]. The Tol2 transposon/transposase method of transgenesis was used with minor modifications. Two nanoliters were injected in the cell of one-cell stage embryos containing 50 ng/ $\mu$ l of transposase mRNA, 40 ng/ $\mu$ l of phenol/chloroform-purified DNA, and 0.05% phenol red. Embryos were raised until 30 hpf, manually dechorionated and treated with tricaine to take pictures at the fluorescence stereoscope. Brightness at the midbrain and muscle regions of each embryo was measured using ImageJ software (Rasband, W.S., ImageJ, U.S. National Institutes of Health, Bethesda, Maryland, USA, <http://imagej.nih.gov/ij/>, 1997–2014). A Zebrafish Enhancer Detection vector was used to test insulator capacity of HSA' *in vivo* by microinjection in one-cell zebrafish embryos as previously described [27,28]. About 10–30 individual zebrafish were analyzed and quantified for each condition.

### 2.11. Generation of chicken CTCF antibody

A new batch of the previously obtained chicken CTCF antibody was generated [20]. The cDNA fragment of chicken CTCF coding for amino acids 82–233 was cloned in the BamHI–XhoI sites of pET28(b), expressed in *E. coli* strain BL21(DE3) and the recombinant protein was isolated and purified. This was used to immunize a New Zealand White rabbit female and three booster doses of 250  $\mu$ g of a recombinant fragment of chCTCF were performed at days 14, 28 and 42. Serum was collected and analyzed, and validated for a new batch of anti-chCTCF<sub>82–233</sub> immunoreactivity and purified using an AminoLink Plus Immobilization Kit (Thermo Scientific). This new production of the chicken anti-chCTCF<sub>82–233</sub> was validated by western blot and immunofluorescence (Supplementary Fig. S1).

## 3. Results

### 3.1. A differential pattern of folate receptor gene expression correlates with differences in chromatin structure during early erythroid differentiation

Previous studies demonstrated that the chicken folate receptor gene is expressed in 6C2 cells, which are derived from a colony forming unit erythrocyte stage cell line (CFU-E), but it is not expressed in terminally differentiated erythroid cells [9] (Fig. 1B). Based on this observation we asked whether the folate receptor gene is expressed in HD3 cells, which are AEV-transformed chicken pro-erythroblast cells corresponding to chicken adult erythroid cells arrested at early stages of differentiation [29] (Supplementary Fig. S2). HD3 cells represent a stage of differentiation intermediate between 6C2 cells and early embryonic erythrocytes. In agreement with previous reports, we found that the *FOLR1* gene is expressed in 6C2 cells but low levels of expression are observed in HD3 cells (Supplementary Fig. S2). As a control, RNA from primary chicken fibroblasts and 10-day chicken embryo erythrocytes (10dRBCs) was isolated, and in primary fibroblasts no significant transcription of the *FOLR1* gene was detected.

From a chromatin structural perspective, several DHSs were previously mapped in erythroid cells, one corresponding to the *FOLR1* gene promoter, the HSA site corresponding to the *FOLR1* enhancer, and the HSA' site that is less prevalent and has undesignated function [9]. Based on our previous expression analysis we decided to perform a comparative DHS assay using nuclei from 6C2 and HD3 cells (Fig. 1C). Using a different genomic restriction enzyme digestion, with the aim to increase the resolution of the assay, we found five instead of three DHSs in 6C2 cells, in particular, the three DHSs that were previously observed, designated as HSA, HSB and HSA' [9,18] (Fig. 1C). To define the location of additional DHSs 3' of the HSA' region we performed DNase I genomic digestion on 6C2 nuclei at a higher resolution using the XhoI restriction enzyme. Of note, in HD3 cells we observed the same DHSs

but an increased range of DNase I enzyme units was needed (0.3 vs 3 units), suggesting a progressive compaction of the chromatin in this region (Fig. 1C).

Together, these results suggest that folate receptor gene transcription is restricted to an early differentiation stage represented here by the 6C2 cells, and that in later stages there is a progressive silencing, already insinuated in HD3 cells that culminates with a repressive chromatin structure in terminally differentiated adult erythrocytes [9].

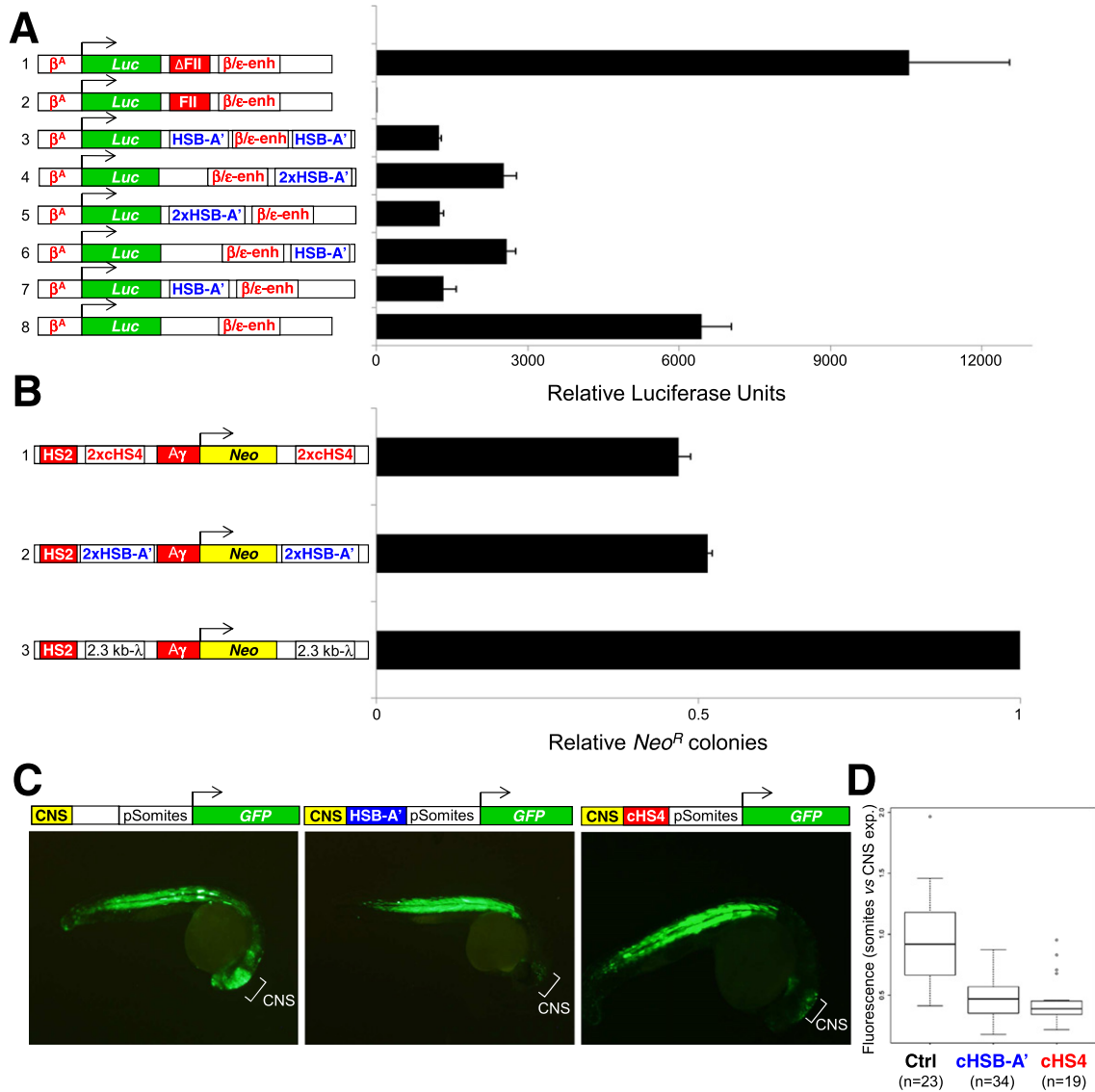
### 3.2. Identification of CTCF binding *in vitro* and *in vivo* at the 5' boundary of the 16 kb of heterochromatin

Since we hypothesized that a chromatin insulator is required to protect the folate receptor gene against epigenetic silencing from the adjacent heterochromatin, we searched for the presence of the multifunctional and insulator-related CTCF factor [30]. As an initial approximation we took advantage of our previously published ChIP-seq series of experiments and found the presence in 10dRBCs of a CTCF located in the genomic region comprising HSA' [28]. Detailed analysis of the HSA' sequence revealed the presence of a 20 base pair (bp) motif similar to the consensus sequence derived from the ChIP-seq experiments [28] (Supplementary Fig. S3).

To further characterize this binding we performed chromatin immunoprecipitation assays using 6C2 cells (Fig. 1D). Based on the sequence motif and the location of the DHS in the HSA' region we choose two sets of primers to demonstrate the *in vivo* binding of CTCF (Fig. 1D and Supplementary Fig. S3). CTCF binding was seen using both pairs of primers, even though primers located in the proximity of the 5' boundary of the heterochromatin revealed more pronounced *in vivo* binding of CTCF to HSA' (Fig. 1D and Supplementary Fig. S3). Importantly, we repeated the same experiments with primers from HSB and HSA' (Fig. 1D and Supplementary Fig. S3) and chromatin isolated from erythroid and non-erythroid cells (Supplementary Fig. S4). As for the non-erythroid cells we tested the chicken lymphoid DT40 cell line, and for the erythroid cells the HD24, 6C2, HD3 and primary cells from 5- and 10-day chicken embryo erythrocytes (5dRBC and 10dRBC). We found once again that CTCF is enriched at HSA' in 6C2 cells and no binding was observed in the other cell types (Supplementary Fig. S4). Therefore, CTCF binds to HSA' in 6C2 cells, in an early erythroid differentiation stage in concordance with transcription of the folate receptor gene and a permissive chromatin configuration of the locus.

### 3.3. Enhancer blocking activity of the HSB–HSA' boundary element delimits the folate receptor intergenic region and the 16 kb of heterochromatin

Based on all our previous evidence we asked whether the HSA' element behaves like a *bona fide* chromatin insulator. We began by testing its capacity to block an enhancer in a position-dependent manner. We performed a transient transfection enhancer-blocking assay in 6C2 cells using *Luciferase* as a reporter gene and the chicken  $\beta^A$  promoter and  $\beta/\epsilon$  enhancer as regulatory elements [26,31]. We tested an 870 bp DNA fragment that includes HSB and HSA' "between" the  $\beta^A$  promoter and the  $\beta/\epsilon$  enhancer (Fig. 2A; construct 7) and downstream to the  $\beta/\epsilon$  enhancer (Fig. 2A; construct 6), in one and two copies (Fig. 2A; constructs 5 and 7; constructs 4 and 6, respectively). We found that HSA' possesses an enhancer-blocking activity that is not increased when two copies of the 870 bp DNA fragment (HSB–HSA') were tested. When HSA' was located downstream to the  $\beta/\epsilon$  enhancer we observed a partial reduction in *Luciferase* activity (Fig. 2A; constructs 4 and 6). This is not surprising since we are transfecting circular plasmids in which it has been demonstrated that the enhancer's activity is bidirectional, so some enhancer-blocking activity is also expected when the insulator element is placed downstream to the enhancer [31]. As a positive control for our transfection enhancer-blocking assay we used both normal and mutated versions of the cHS4 FII CTCF-binding site



**Fig. 2.** *In vitro* and *in vivo* enhancer blocking activity of the HSB–HSA' element. (A) Enhancer-blocking in 6C2 cells that were transiently transfected with circular plasmids containing the chicken  $\beta^A$  gene promoter and the  $\beta/\epsilon$  enhancer driving expression of the *Luciferase* (*Luc*) reporter gene. Binding site FII of the chicken  $\beta$ -globin cHS4 insulator and its mutant version ( $\Delta$ FII) were used as controls. (B) Colony assay performed in the human erythroleukemic K562 cell line. Neomycin-resistant colonies were counted in agar plates from three independent stable transfections; the two copies of the 1.2 kb  $\beta$ -globin cHS4 insulator were used as a positive control and 2.3 kb of  $\lambda$ -DNA as a distance control as described previously [6]. (C) *In vivo* enhancer-blocking assay using transgenic zebrafish in which the HSB–HSA' insulator is placed between the midbrain-specific enhancer and the somite-specific promoter. The vector comprises a strong and midbrain-specific enhancer (CNS) and a cardiac-actin promoter (pSomites) driving a robust expression of the EGFP reporter in somites and heart [27]. As in B, the  $\beta$ -globin cHS4 was used as a positive control. (D) Graph showing the normalized (somite-vs-CNS) fluorescence intensities. The number of injected embryos is shown in the graph.

(Fig. 2A; FII–BamHI (between; construct 2) and  $\Delta$ FII–BamHI (between; construct 1), respectively).

To further confirm the enhancer-blocking activity of the HSB–HSA' element we decided to survey such activity in an integrated chromatin context by performing a colony assay in human erythroleukemic K562 cells [6]. We took advantage of the pJ3–4 and pJ5–4 plasmids and inserted the HSB–HSA' element between the human  $\beta^A$  gene promoter driving the neomycin resistance gene, and the HS2 enhancer of the human  $\beta$ -globin LCR. For comparison, we also stably transfected two copies of the original 1.2 kb DNA fragment of the chicken  $\beta$ -globin cHS4 insulator [6,7]. Our results confirmed a reduction in the number of drug-resistant colonies, suggesting an enhancer-blocking activity of the HSB–HSA' element to an extent equivalent to the cHS4 insulator (Fig. 2B).

Taking advantage of a recently developed *in vivo* enhancer-blocking assay in zebrafish, we asked if the HSB–HSA' element can retain this insulator function *in vivo* [27,28] (Fig. 2C and D). The reporter system is based on the zebrafish *irx* Z48 enhancer which promotes strong expression in

the midbrain, and the cardiac actin promoter isolated from *Xenopus laevis*, which drives the *EGFP* gene as a reporter [27]. Thus, these regulatory elements drive strong expression in somites and heart. We incorporated the 870 bp DNA fragment corresponding to the HSB–HSA' element and used the cHS4 insulator as a positive control (Fig. 2C). We were able to confirm the enhancer-blocking activity of the HSB–HSA' element with high confidence based on a reduced ratio between the midbrain and somite *EGFP* signals (Fig. 2C and D). In summary, we conclude that the HSB–HSA' element possesses one of the two functional properties of a chromatin insulator, and therefore, we next asked if the HSB–HSA' can protect a transgene against chromosomal position effects.

#### 3.4. The HSB–HSA' element protects a transgene against chromosomal position effects

To investigate if the HSB–HSA' sequence is able to protect a transgene against chromosomal position effects (CPEs), we inserted two



copies of the HSB–HSA' 870 bp DNA fragment on each side of the reporter vector represented by the *GFP* gene under the transcriptional control of the chicken  $\alpha^D$  gene promoter, a vector that has been previously tested as being sensitive to CPEs [26]. It is worth mentioning that the HSA' sequence was not tested because knockdown of CTCF had no effect on the ability of HSB–A' to protect against silencing (see Supplementary Fig. S5). Then, with the intention to incorporate a comparative parameter as a positive control, we analyzed the expression of the  $\alpha^D$ G transgene shielded by two copies of the 1.2 kb cHS4 insulator on both sides of the vector [16,17,26]. On average we generated 10 independent monoclonal 6C2 cell lines, and surveyed them by Southern blotting for their integrity and copy-number. Individual cell lines were maintained in continuous cell culture in the absence of drug selection for more than 90 days (Fig. 3). FACS analysis was performed periodically to determine the percentage of GFP-positive cells as a measure of the transcriptional activity of the transgene. From previous work we know that in the cell populations in which expression of the transgenes has been progressively extinguished the effect is not due to the inactivation of transcription factors [15,16]. Our results showed that the  $\alpha^D$ G transgene is, as expected, subject to large variability in expression and a progressive epigenetic silencing [15,26]. In contrast, the cHS4 insulator, and even more pronouncedly the HSB–HSA' insulator, maintained relatively constant and elevated levels of GFP-positive cells over long periods of continuous cell culture (Fig. 4). Of note, we cannot discard that the elevated transgene expression observed in the presence of the HSB–HSA' sequences could come from an enhancer activity (see the Discussion section).

In summary, we have shown that the HSB–HSA' sequence binds the insulator factor CTCF, possesses enhancer-blocking activity and is capable of protecting a transgene against CPEs. Therefore, we can conclude that the HSB–HSA' element can be considered a chromatin insulator

that is strategically located in the transition between the 5' limit of the 16 kb of heterochromatin and the folate receptor intergenic region.

### 3.5. *In vivo* CTCF binding coincides with the *FOLR1* gene expression pattern and is dissociated in later erythroid differentiation stages

Previous chromatin immunoprecipitation assays suggested the *in vivo* binding of CTCF to this element in 6C2 cells but not in other cell-types in a non-quantitative way (Fig. 1D). Due to the apparent relevance of CTCF to insulator function, but most importantly the erythroid-differentiation stage expression of the *FOLR1* gene, we investigated the pattern of CTCF binding to HSA' in different erythroid and non-erythroid cell-types (Fig. 4A). We confirmed that CTCF binds preferentially to the HSA' region in the CFU-E stage corresponding to 6C2 cells by quantitative ChIP assay (Fig. 4A). This result may be relevant for stage-specific transcriptional regulation of the *FOLR1* gene.

Next, and based on the fact that CTCF is bound to HSA' and cHS4 *in vivo* and may demarcate the 16 kb of heterochromatin we investigated the *in vivo* association of CTCF with these two boundary elements in different cell-types (Fig. 4B). Importantly, for the HSA' element and *FOLR1* gene expression, CTCF binding is restricted to the early stage represented by pro-erythroblast 6C2 cells (Fig. 4B). In contrast, CTCF is constitutively bound to the FII motif within the cHS4 insulator, including in a non-erythroid cell line, like DT40 cells (Fig. 4B).

These results may suggest a more tight and stage-restricted role of CTCF and the HSA' element in early differentiation stages to allow the expression of the *FOLR1* gene and to keep it irreversibly silenced in subsequent stages. In contrast, the cHS4 insulator and the  $\beta$ -globin locus are constitutively structured in a poised chromatin configuration that may facilitate the initial steps needed for  $\beta$ -globin gene activation (see the Discussion section).

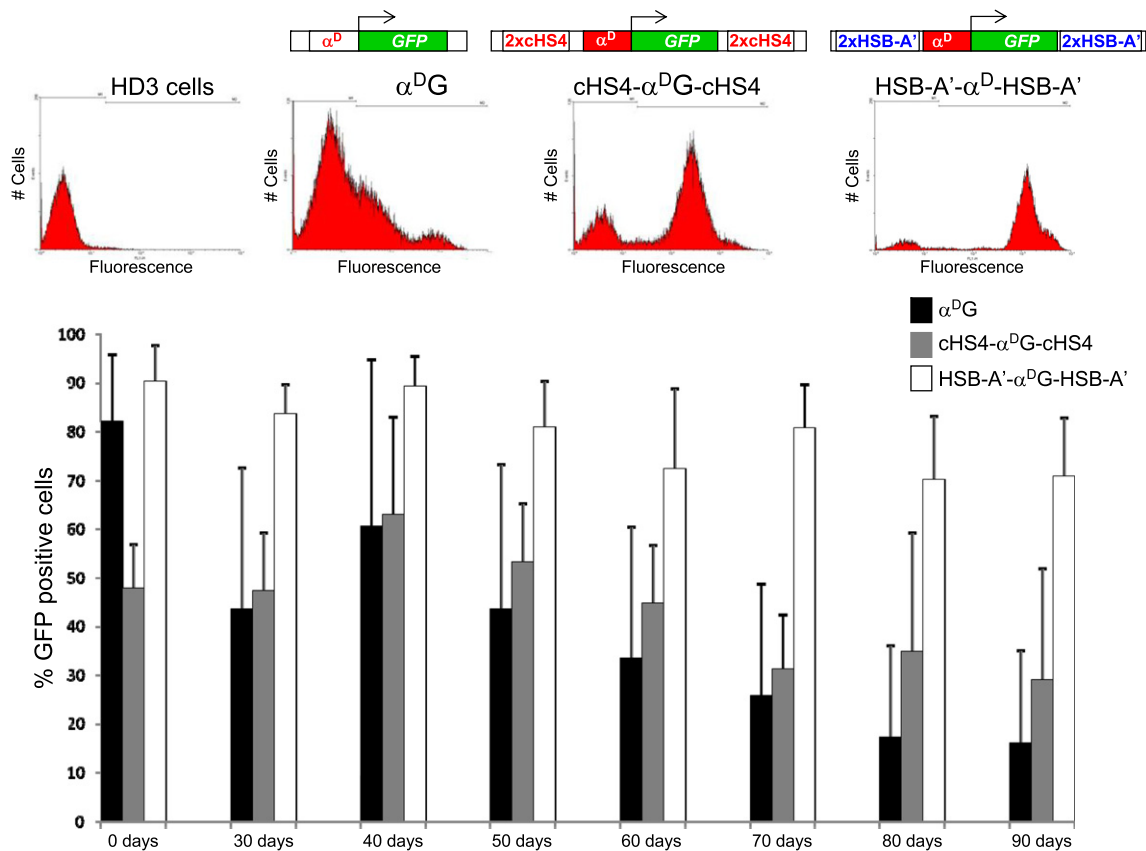
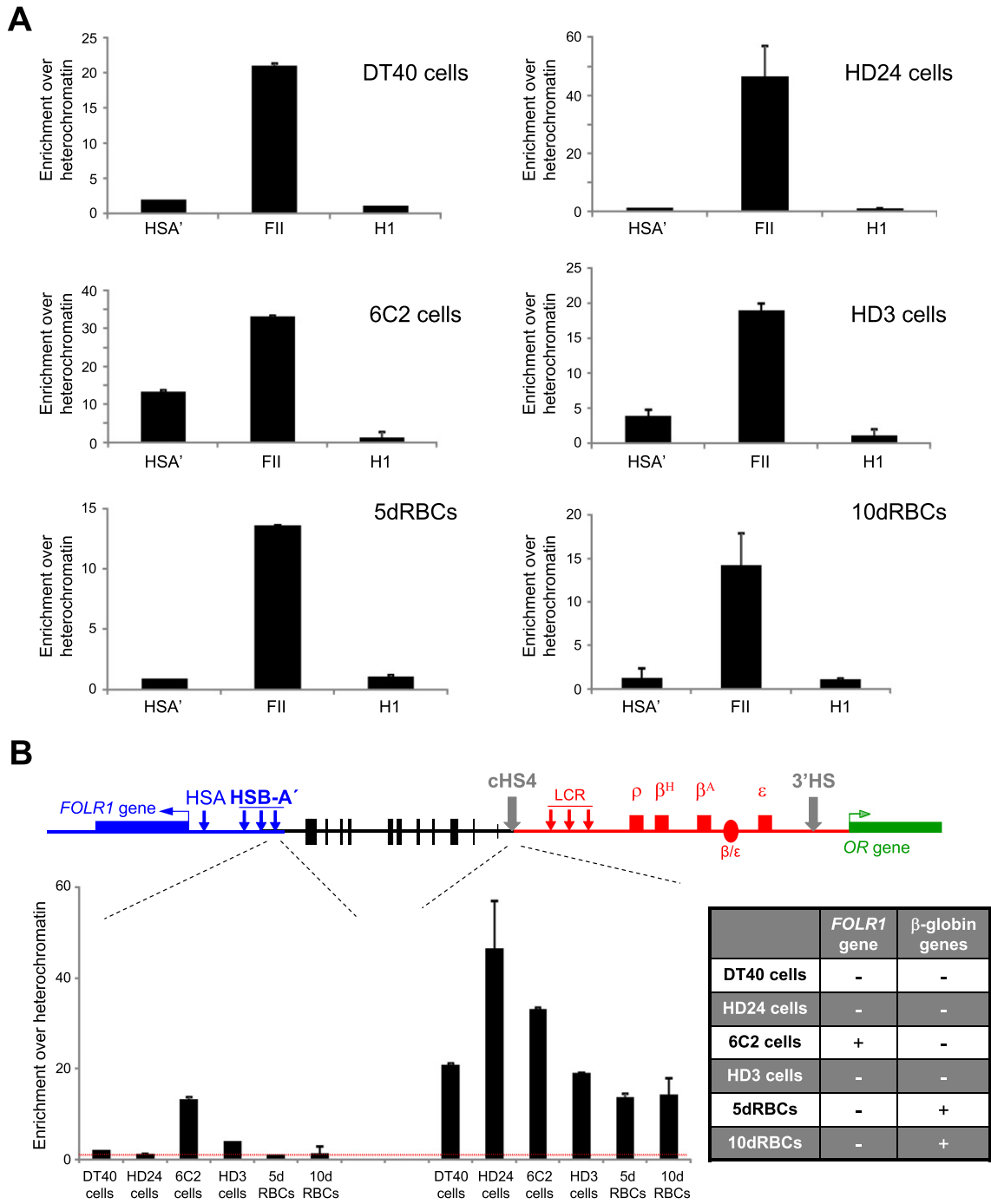


Fig. 3. Protection against chromosomal position effects by the HSB–HSA' element. At least 10 independent monoclonal and single-copy transgenic HD3 cells lines were generated. Each transgenic cell line was maintained in continuous cell culture and flow cytometry profiles were obtained periodically. The constructs and representative flow cytometry profiles from 80 days of continuous cell culture are shown on top. The graph at the bottom corresponds to the mean of GFP-positive cells and the error bars are the SD.

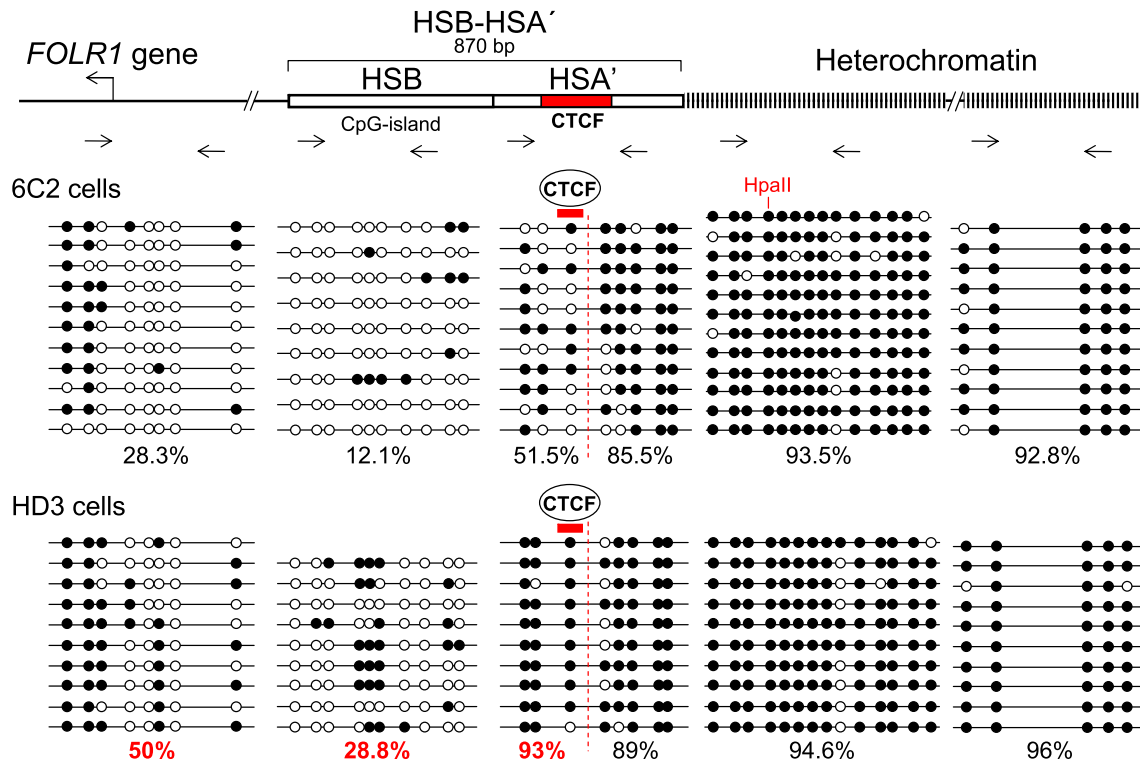


**Fig. 4.** *In vivo* binding of CTCF to the HSA' element in different cell types representing progressive erythroid differentiation stages. (A) For comparative purposes we performed a ChIP assay using primers (shown in Supplementary Fig. S3) that reveal *in vivo* binding to the CTCF site known as FII and located at the  $\beta$ -globin cHS4 insulator. Note that DT40 cells are from a lymphoid cell lineage and the different differentiation stages are shown in Fig. 1B. (B) A large-scale view of the genomic region under study with a comparative analysis of the *in vivo* binding of CTCF to the HSA' and FII sites is shown. A table recapitulating the different patterns of gene expression between the *FOLR1* gene and the  $\beta$ -globin group of genes is also shown. Note that this comparative study contributes to the functional model proposed later in Fig. 8.

**3.6. Gain of DNA methylation in the intergenic sequences between the *FOLR1* gene and the heterochromatin region correlates with *FOLR1* gene silencing**

On the basis of the CTCF cell type-specific pattern of binding, together with the restricted expression of the *FOLR1* gene to the CFU-E stage (6C2 cells), we asked whether there is a gain of DNA methylation due to hypothetical spreading from the heterochromatin in an advanced erythroid stage (pro-erythroblast HD3 cells). Previous published work

established that 15.5 kb of heterochromatin are highly methylated and can be confined to a unique 15.5 kb DNA fragment after digestion by the methylation-sensitive HpaII restriction enzyme [9,11]. In addition, the same region is abundant in CR1-type repetitive sequences that represent a source of heterochromatin formation. Therefore, we performed a sodium bisulfite genomic DNA transformation using 6C2 and HD3 cells followed by DNA sequencing of five genomic DNA subfragments comprising the intergenic region between the *FOLR1* gene and the heterochromatin region (Fig. 5A). The promoter region of the



**Fig. 5.** Comparative analysis of the DNA methylation status of the intergenic region located between the *FOLR1* gene and the 16 kb of constitutive heterochromatin. Open circles represent unmethylated CpGs and filled circles methylated  $C^m$ pGs. The primers used for the analysis of each of the five different sub-fragments are shown as horizontal arrows under the genomic region scheme. The genomic location of the DNA methylated *HpaII* site representing the 5' boundary of the 15.5 kb of heterochromatin is shown in red.

*FOLR1* gene and an internal segment of the 5'-most region of heterochromatin were included in the analysis. The results show that in 6C2 cells, where the *FOLR1* gene is expressed and acquires its permissive chromatin configuration, there are low levels of DNA methylation (Fig. 5). Of note, a transition is seen from the site of CTCF binding. Upstream of this site, towards the heterochromatin, the region analyzed reaches between 85% and 93% of DNA methylation (Fig. 5). In contrast, downstream of the CTCF-binding site the DNA methylation levels are reduced and vary between 12% and 51%. Importantly, when we compared these DNA methylation patterns with those obtained from HD3 cells, we found that there is a gain in DNA methylation ranging from 29% to 50%, reaching 93% in the vicinity of the CTCF-binding site (Fig. 5). It is worth mentioning that this stage-specific gain in DNA methylation is coincident with a loss in *FOLR1* gene expression, CTCF dissociation and a more repressive chromatin structure of the region (Figs. 1, 4 and 5). Together, these results support a model in which HSA' and CTCF are part of a barrier element that allows a permissive chromatin structure to facilitate *FOLR1* gene transcription in 6C2 cells.

### 3.7. CTCF loss of function affects the expression of the chicken folate receptor gene

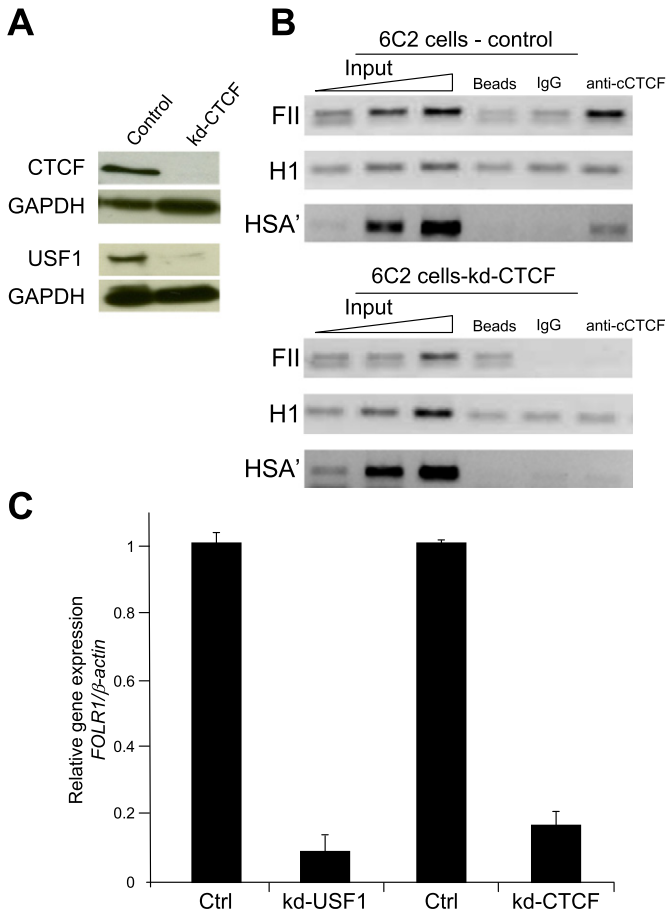
Due to the apparent relevance of CTCF to the timing in which the folate receptor gene is transcribed, we decided to perform a loss of function assay. Then, we decided to re-analyze the sequence and generate two new small-hairpin RNA interference (shRNAi) sequences. The shRNAi-cCTCF1 sequence was the most efficient, much more so than those tested previously (Fig. 6A). We then generated lentivirus particles expressing shRNAi-cCTCF1 in infected 6C2 cells, and confirmed CTCF absence by ChIP using primers over the HSA' element (Fig. 6B). We next questioned if the absence of CTCF would have any effect on *FOLR1* gene expression. We therefore performed quantitative RT-PCR and found a drastic reduction in *FOLR1* gene transcription (Fig. 6C). Based on the functional dependence of the cHS4  $\beta$ -globin insulator on

the USF1–USF2 transcription factors, we stably transfected five shRNAi plasmids against USF1 [24,32] (Fig. 6A). The contribution of the USF1/2 factors was further supported by the recent demonstration of the *in vivo* binding of the USF1 factor to the *FOLR1* gene promoter, and to the HSA and the HSB regions in 6C2 cells [18]. As for the CTCF knock-down, we observed a significant reduction in *FOLR1* gene expression (Fig. 6C).

In conclusion, our results suggest that CTCF, as well as the USF1 and USF2 factors, can participate in the epigenetic regulation of the folate receptor gene, notwithstanding the role of other regulatory components (see the Discussion section).

### 3.8. The HSB–HSA' insulator counteracts the silencing effect of the heterochromatin region

Since we hypothesized that the HSB–HSA' element protects the *FOLR1* gene against heterochromatin silencing in 6C2 cells, we decided to assess the capacity of the same transgenes to counteract chromosomal position effects (Fig. 3) by incorporating 6.7 kb of the heterochromatin region in those transgenes (Fig. 7). As predicted and as a confirmation of the heterochromatin effect, the  $\alpha^D$ G transgene was rapidly and strongly subjected to CPEs. Instead, two copies of the core cHS4  $\beta$ -globin insulator [16] and two copies of the 870 bp HSB–HSA' element were able to protect the transgene against chromosomal position effects and progressive silencing after long periods of continuous cell culture (Fig. 7). In addition, we selected three independent lines for the Hetero[cHS4– $\alpha^D$ G–cHS4] and Hetero[HSB–HSA'– $\alpha^D$ G–HSB–HSA'] constructs and infected each cell line with a lentivirus expressing the shRNAi-cCTCF1 sequence targeting chicken CTCF (Supplementary Fig. S5). After more than 40 days of continuous cell culture we were not able to identify any significant loss in expression for each transgene. These results suggest that, as has been previously demonstrated for the cHS4 insulator [16], CTCF is important for HSB–HSA' activity and influences *FOLR1* gene expression, but other factors (like USF1/2 and VEZF1) [24,32–34] and chromatin



**Fig. 6.** Effect of knockdown of CTCF and USF1 factor on folate receptor gene expression in 6C2 cells. (A) Western blot demonstrating the depletion of CTCF and USF1. (B) ChIP assay showing the drastically reduced binding of CTCF in 6C2 knockdown cells. The chicken  $\beta$ -globin cHS4 insulator FII site is used as a positive control for CTCF binding, and H1 primers, located around the central part of the 16 kb of constitutive heterochromatin, are used as a negative control. This is a representative assay of two independent sets of experiments. (C) The effect of the USF1 (kd-USF1) and CTCF (kd-CTCF) depletion evaluated by quantitative RT-PCR is shown.

remodelers are contributing to the overall activity of the HSB–HSA' insulator (see below).

In summary, these observations are in concordance with a barrier function being associated with the HSB–HSA' element, and this property is not only restricted to the activity of CTCF.

#### 4. Discussion

Regulated changes in chromatin structure that have consequences on gene expression during cell differentiation have recently attracted our attention, not only at the genome-wide scale, but also mechanistically. In the present study we addressed three related chromatin domains contained within an ~50 kb genomic region from an integrative point of view; two independent and erythroid-specific gene domains which are separated by ~16 kb of constitutive heterochromatin (Fig. 1A). Here we report that the erythroid-specific *FOLR1* gene is expressed earliest in a primitive stage of cell differentiation, with the adoption of a characteristic chromatin configuration. Following a restricted window of time in which *FOLR1* gene expression ceases, there is a regulated epigenetic silencing of this gene through heterochromatinization. This transition period is coincident with the release of CTCF binding to the 5' boundary of the heterochromatin region, the gain of DNA methylation and the acquisition of histone repressive marks over the *FOLR1* gene [18]. This apparent CTCF- and chromatin-associated regulation of the *FOLR1* gene during early erythroid differentiation is needed to allow its expression

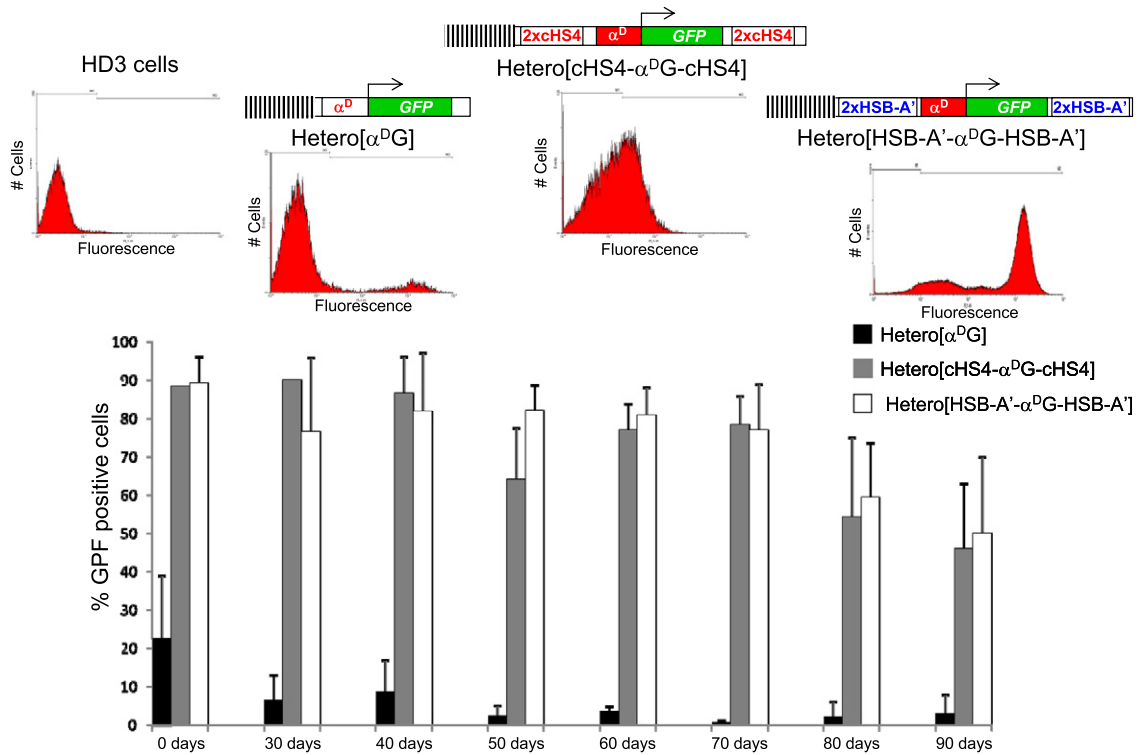
and subsequent irreversible silencing in the transition to later differentiation stages. Unexpectedly, at the 3' side of the heterochromatin region the cHS4 insulator and associated factors, including CTCF, are not subjected to this kind of restricted regulation since a constant open configuration of the cHS4 boundary region and binding of CTCF is observed in different erythroid cell types irrespective of the expression order of adjacent genes (Fig. 4B). Therefore, we demonstrated that the heterochromatin region is shielded by two chromatin insulators with apparent similar properties, and that, as occurs in embryonic stem cells, heterochromatin formation represents one of the mechanisms to induce cell differentiation (Fig. 8).

A large combination of active histone marks have previously been characterized over the cHS4 insulator irrespective of the transcriptional status of the *FOLR1* or  $\beta$ -globin genes in distinct erythroid cell types [10, 18]. In a recent study, the great majority of these open chromatin features were described at the HSA and HSB regions in 6C2 cells [29]. Furthermore, the E3 ligase RNF20/BRE1A, which mediates histone H2B mono-ubiquitination, was found not only at cHS4 but also at HSA and HSB [18]. Loss of function experiments demonstrated that RNF20 depletion and loss of H2Bub1 induces the loss of active histone modifications and the spreading of adjacent heterochromatin over the *FOLR1* gene and  $\beta$ -globin locus [18]. It is also important to note that USF1/USF2, which recruits the histone H3 methyltransferase Set1 complex and the histone H4 arginine 3 methyltransferase PRMT1, are also found at the HSA and HSB sites [24,33]. In addition, the BGP1/VEZF1 protein that protects the cHS4 insulator, transgenes and CpG-islands against DNA methylation has also been found around the HSA and HSB regions [18,34].

All these observations argue in favor of the need for a boundary element, at least at the CFU-E stage of 6C2 cells to allow *FOLR1* gene expression and to avoid its epigenetic silencing by heterochromatin formation (Fig. 7). We propose that HSB together with the HSA' element plays this boundary role by recruiting chromatin remodeling complexes. However, in contrast to the mechanism of action of the  $\beta$ -globin cHS4 element, the HSB–HSA' element is regulated during a precise period of time to firstly allow *FOLR1* gene expression, and secondly, to facilitate its irreversible heterochromatinization during later differentiation. Based on our assays in HD3 cells and on DNase I hypersensitivity and DNA methylation experiments, there is no rapid epigenetic silencing of the *FOLR1* locus, which is consistent with the progressive chromatin repression model [35,36] (Fig. 8). As a non-exclusive model and based on our CPE results we should take into consideration that the HSB–HSA' elements can behave as an enhancer (Fig. 3). This is not surprising since it has recently been demonstrated by *in situ* Hi-C experiments that a high percentage of CTCF sites in the human genome, with convergent orientation sequence, correspond to anchor sites contributing to chromatin loop formation [37]. Among these interactions, 30% are formed between promoters and enhancers [37]. Therefore, the HSB–HSA' can have a dual function in its influence over the *FOLR1* gene expression during erythroid differentiation.

Another model for the action of the HSB–HSA' element comes from recently described non-coding transcription of the heterochromatin region, which is mainly Dicer-dependent [12]. It is known that Dicer processes bidirectional intergenic transcripts to generate siRNAs which function in the establishment and maintenance of heterochromatin [38,39]. From another non-coding RNA perspective, and in an attempt to link HSA' function and the role of CTCF, it has previously been shown that the DEAD Box RNA helicase p68, which binds the RNA component SRA (Steroid Receptor RNA activator), forms a complex that interacts with CTCF to help stabilize the binding of cohesin [12]. More recently, we have shown that CTCF presents an RNA-binding region that interacts with a large amount of transcripts along the genome [40]. Interestingly, CTCF–RNA interaction seems to contribute to CTCF multimerization. Together, p68/SRA, cohesin and RNA interactions with CTCF may contribute to the establishment of higher-order chromatin structure in 6C2 cells that allows *FOLR1* gene expression at this particular differentiation stage. A hypothetical prediction for the dissociation of such a higher-order organization that is prone to *FOLR1* gene

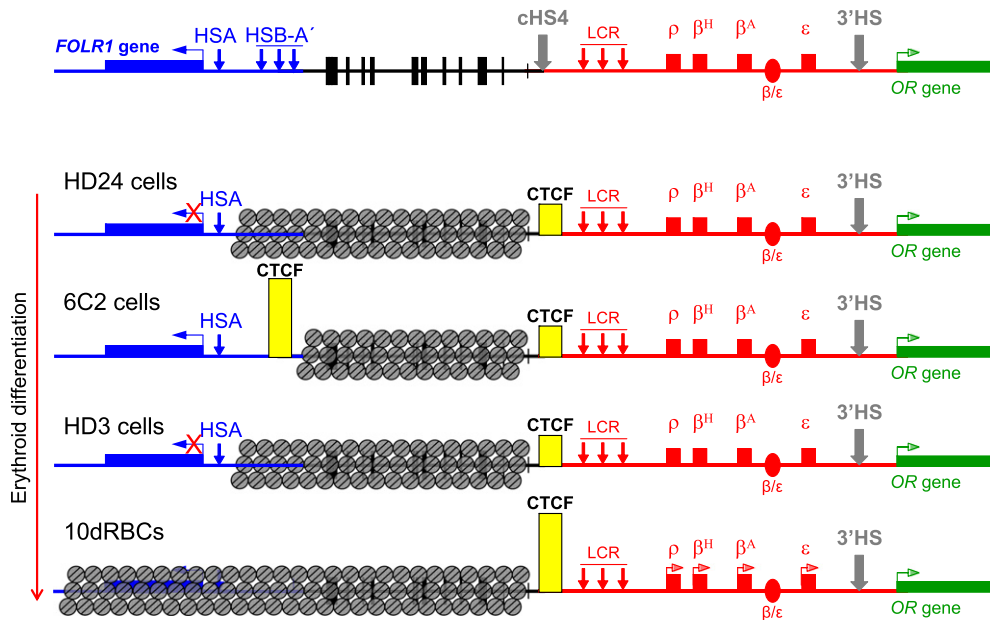




**Fig. 7.** Protection by the cHS4 and HSB-HSA' insulators against the presence of constitutive heterochromatin. The assay is basically the same as in Fig. 3 except that here 6.7 kb of constitutive heterochromatin is incorporated (vertical dashed lines). Due to the high abundance of repetitive sequences, it was technically not possible to test the entire 16 kb of constitutive heterochromatin. Notice how, in the absence of either the cHS4 or HSB-HSA' insulators, the 6.7 kb of heterochromatin exerts a strong silencing effect over the transgene (see Hetero[α<sup>D</sup>] constructs).

expression could come from the CDK phosphorylation of CTCF, which disrupts its association with RNA. This is an attractive possibility that deserves to be addressed experimentally with many regulatory implications.

Several reports have shown the potential role of CTCF in chromatin loop formation and genome organization through topologically associated domains or TADs [4,41,42]. Based on a series of high-resolution Hi-C experiments it has been demonstrated that in humans, TADs are



**Fig. 8.** Model summarizing our results and showing how the 16 kb of heterochromatin participates in the erythroid differentiation process through the regulation of the folate receptor gene chromatin configuration. The yellow bars schematically represent the relative binding frequency of CTCF at the HSA' and cHS4 sites. It is important to insist that the insulator function is not only restricted to CTCF binding, since USF1/2 and probably other remodelers expanding the HSB-HSA' element may contribute to the overall activity. Interestingly, CTCF is constitutively bound to the cHS4 β-globin insulator, probably to avoid the irreversible spreading of the constitutive heterochromatin, thereby permitting the proper chromatin configuration of the chicken β-globin locus and the regulated expression of the genes located therein.

composed by sub-compartments formed by chromatin loops and, that in the majority of cases (92%) CTCF binding sites are found at loop anchors in a convergent orientation [37]. Importantly, CTCF binding sites found in the genome in a non-convergent orientation are not prone to contribute to loop formation [37]. Thus, based in such findings CTCF sites located on each end of the 16 kb of heterochromatin seems unlikely to form a chromatin loop since they are found in divergent orientation (data not shown). One possibility is that the CTCF site associated to the HSB-HSA' element may form a regulatory loop with a distal site located 5' containing at least the *FOLR1* gene locus in the 6C2 cell stage, which will be dissociated in later erythroid stages. In contrast, the loop formed by the cHS4 CTCF site could form a constitutive loop with a distant downstream CTCF site. It is evident that we are reaching a point in which the study of local and chromatin domain scale is relevant but not sufficient. The three-dimensional organization of the genome is also needed to be addressed to better understand differential gene expression.

### Acknowledgements

We thank Catherine Farrell, Karin Meier and Rodrigo Arzate-Mejía for critical reading of the manuscript. This work was supported by the DGAPA-PAPIIT, UNAM (IN209403, IN203811 and IN201114), and CONACyT (42653-Q, 128464 and 220503) to FR-T, and by a Ph.D. fellowship from CONACyT (CVU 207989) and Programa de Apoyo a los Estudios del Posgrado (PAEP), UNAM to EG-B. Additional support was provided by the PhD Graduate Program, “Doctorado en Ciencias Bioquímicas”, to the Instituto de Fisiología Celular and the Universidad Nacional Autónoma de México.

JLG-S was funded by grants from Ministerio de Economía y Competitividad (BFU2013-41322-P) and the Andalusian Government (BIO-396). A postdoctoral grant from the Universidad Pablo de Olavide supported JJT.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.bbagr.2015.05.011>.

### References

- [1] G. Cavalli, T. Misteli, Functional implications of genome topology, *Nat. Struct. Mol. Biol.* 20 (2013) 290–299.
- [2] C.T. Ong, V.G. Corces, CTCF: an architectural protein bridging genome topology and function, *Nat. Rev. Genet.* 15 (2012) 234–246.
- [3] E. de Wit, W. de Laat, A decade of 3C technologies: insights into nuclear organization, *Genes Dev.* 26 (2012) 11–24.
- [4] E.P. Nora, B.R. Lajoie, E.G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, et al., Spatial partitioning of the regulatory landscape of the X-inactivation centre, *Nature* 485 (2012) 381–385.
- [5] T.R. Hebbes, A.L. Clayton, A.W. Thorne, C. Crane-Robinson, Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken  $\beta$ -globin chromosomal domain, *EMBO J.* 13 (1994) 1823–1830.
- [6] J.H. Chung, M. Whiteley, G. Felsenfeld, A 5' element of the chicken  $\beta$ -globin domain serves as an insulator in human erythroid cells and protects against position effects in *Drosophila*, *Cell* 74 (1993) 505–514.
- [7] A.C. Bell, A.G. West, G. Felsenfeld, The protein CTCF is required for the enhancer blocking activity of vertebrate insulators, *Cell* 98 (1999) 387–396.
- [8] B. Burgess-Beusse, C. Farrell, M. Gaszner, M. Litt, V. Mutskov, F. Recillas-Targa, et al., The insulation of genes from external enhancers and silencing chromatin, *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002) 16433–16437.
- [9] M.N. Prioleau, P. Nony, M. Simpson, G. Felsenfeld, An insulator element and condensed chromatin region separate the chicken  $\beta$ -globin locus from an independently regulated erythroid-specific folate receptor gene, *EMBO J.* 18 (1999) 4035–4048.
- [10] M.D. Litt, M. Simpson, F. Recillas-Targa, M.N. Prioleau, G. Felsenfeld, Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci, *EMBO J.* 20 (2001) 2224–2235.
- [11] R. Ghirlando, M.D. Litt, M.N. Prioleau, F. Recillas-Targa, G. Felsenfeld, Physical properties of a genomic condensed chromatin fragment, *J. Mol. Biol.* 336 (2004) 597–605.
- [12] K.E. Giles, R. Ghirlando, G. Felsenfeld, Maintenance of a constitutive heterochromatin domain in vertebrates by a Dicer-dependent mechanism, *Nat. Cell Biol.* 12 (2010) 94–99.
- [13] F. Recillas-Targa, V. Valadez-Graham, C. Farrell, Prospects and implications of using chromatin insulators in gene therapy and transgenesis, *BioEssays* 26 (2004) 796–807.
- [14] S.C.R. Elgin, G. Reuter, Position-effect variegation, heterochromatin formation and gene silencing in *Drosophila*, *Cold Spring Harb. Perspect. Biol.* 5 (2013) a017780.
- [15] M.J. Pikaart, F. Recillas-Targa, G. Felsenfeld, Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators, *Genes Dev.* 12 (1998) 2852–2862.
- [16] F. Recillas-Targa, M.J. Pikaart, B. Burgess-Beusse, A.C. Bell, M.D. Litt, A.G. West, et al., Position-effect protection and enhancer blocking by the chicken  $\beta$ -globin insulator are separable activities, *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002) 6883–6888.
- [17] H. Rincón-Arango, M. Furlan-Magaril, F. Recillas-Targa, Protection against telomeric position effects by the chicken cHS4  $\beta$ -globin insulator, *Proc. Natl. Acad. Sci. U. S. A.* 104 (2007) 14044–14049.
- [18] M.K.W. Ma, C. Heath, A. Hair, A.G. West, Histone crosstalk directed by H2B ubiquitination is required for chromatin boundary integrity, *PLoS Genet.* 7 (2011) e1002175.
- [19] T. Metz, T. Graf, v-myb and v-ets transform chicken erythroid cells and cooperate both in *trans* and in *cis* to induce distinct differentiation phenotypes, *Genes Dev.* 5 (1991) 369–380.
- [20] V. Valadez-Graham, S.V. Razin, F. Recillas-Targa, CTCF-dependent enhancer blockers at the upstream region of the chicken  $\alpha$ -globin gene domain, *Nucleic Acids Res.* 2 (2004) 1354–1362.
- [21] M. Davalos-Salas, M. Furlan-Magaril, E. González-Buendía, C. Valdes-Quezada, E. Ayala-Ortega, F. Recillas-Targa, Gain of DNA methylation is enhanced in the absence of CTCF at the human retinoblastoma gene promoter, *BMC Cancer* 11 (2011) 232.
- [22] Y. Bingbing, L. Robert, H. Markus, T. Thomas, L. Fran, siRNA Selection Server: an automated siRNA oligonucleotide prediction server, *Nucleic Acids Res.* 32 (2004) W130–W134.
- [23] N.P. Gomes, J.M. Espinosa, Gene-specific repression of the p53 target gene PUMA via intragenic CTCF-cohesin binding, *Genes Dev.* 24 (2010) 1022–1034.
- [24] A.G. West, M. Huang, M. Gaszner, M.D. Litt, G. Felsenfeld, Recruitment of histone modifications by USF proteins at a vertebrate barrier element, *Mol. Cell* 16 (2004) 453–463.
- [25] M. Escamilla-Del-Arenal, F. Recillas-Targa, GATA-1 modulate the chromatin structure and activity of the chicken  $\alpha$ -globin 3' enhancer, *Mol. Cell Biol.* 2 (2008) 575–586.
- [26] M. Furlan-Magaril, E. Rebollar, G. Guerrero, A. Fernández, E. Moltó, E. González-Buendía, et al., An insulator embedded in the chicken  $\alpha$ -globin locus regulates chromatin domain configuration and differential gene expression, *Nucleic Acids Res.* 39 (2011) 89–103.
- [27] J. Bessa, J.J. Tena, E. de la Calle-Mustienes, A. Fernández-Miñán, S. Narango, A. Fernández, et al., Zebrafish enhancer detection (ZED) vector: a new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish, *Dev. Dyn.* 238 (2009) 2409–2417.
- [28] D. Martin, C. Pantoja, A. Fernández Miñán, C. Valdes-Quezada, E. Moltó, F. Matesanz, et al., Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes, *Nat. Struct. Mol. Biol.* 18 (2011) 708–714.
- [29] H. Beug, A. von Kirchbach, G. Doderlin, J.F. Conscience, T. Graf, Chicken haematopoietic cells transformed by serum strains of defective avian leukemia viruses display three distinct phenotypes of differentiation, *Cell* 18 (1979) 375–390.
- [30] J.E. Phillips, V.G. Corces, CTCF: master weaver of the genome, *Cell* 137 (2009) 1194–1211.
- [31] F. Recillas-Targa, A.C. Bell, G. Felsenfeld, Positional enhancer-blocking activity of the chicken  $\beta$ -globin insulator in transiently transfected cells, *Proc. Natl. Acad. Sci. U. S. A.* 96 (1999) 14354–14359.
- [32] S. Huang, X. Li, T.M. Yusufzai, Y. Qiu, G. Felsenfeld, USF1 recruits histone modification complexes and is critical for maintenance of a chromatin barriers, *Mol. Cell Biol.* 27 (2007) 7991–8002.
- [33] S. Huang, M. Litt, G. Felsenfeld, Methylation of histone H4 by arginine methyltransferase PRMT1 is essential *in vivo* for many subsequent histone modifications, *Genes Dev.* 19 (2005) 1885–1893.
- [34] J. Dickson, H. Gowher, R. Strogantsev, M. Gaszner, A. Hair, G. Felsenfeld, et al., VEZF1 elements mediate protection from DNA methylation, *PLoS Genet.* 6 (2010) e1000804.
- [35] V. Mutskov, G. Felsenfeld, Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9, *EMBO J.* 23 (2004) 138–149.
- [36] J. Wang, S.T. Lawry, A.L. Cohen, S. Jia, Chromosome boundary elements and regulation of heterochromatin spreading, *Cell. Mol. Life Sci.* 71 (2014) 4841–4852.
- [37] S.S.P. Rao, M.H. Huntley, N.C. Duran, E.K. Stamenova, I.D. Bochkov, J.T. Robinson, et al., A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping, *Cell* 159 (2014) 1665–1680.
- [38] S.I. Grewal, RNAi-dependent formation of heterochromatin and its diverse functions, *Curr. Opin. Genet. Dev.* 20 (2010) 134–141.
- [39] S.E. Castel, R.A. Martienssen, RNA interference in the nucleus: roles for small RNAs in transcription and beyond, *Nat. Rev. Genet.* 14 (2013) 100–112.
- [40] R. Saldaña-Meyer, E. González-Buendía, G. Guerrero, V. Narendra, R. Bonasio, F. Recillas-Targa, et al., CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, *Wrap53*, *Genes Dev.* 28 (2014) 723–734.
- [41] J.R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, et al., Topological domains in mammalian genomes identified by analysis of chromatin interactions, *Nature* 485 (2012) 376–380.
- [42] T. Sexton, G. Cavalli, The role of chromosome domains in shaping the functional genome, *Cell* 160 (2015) 1049–1059.
- [43] N. Saitoh, A.C. Bell, F. Recillas-Targa, A.G. West, M. Simpson, M. Pikaart, et al., Structural and functional conservation at the boundaries of the chicken  $\beta$ -globin domain, *EMBO J.* 19 (2000) 2315–2322.

## Experimental Strategies to Manipulate the Cellular Levels of the Multifunctional Factor CTCF

Edgar González-Buendía, Rosario Pérez-Molina, Erandi Ayala-Ortega, Georgina Guerrero, and Félix Recillas-Targa

### Abstract

Cellular homeostasis is the result of an intricate and coordinated combinatorial of biochemical and molecular processes. Among them is the control of gene expression in the context of the chromatin structure which is central for cell survival. Interdependent action of transcription factors, cofactors, chromatin remodeling activities, and three-dimensional organization of the genome are responsible to reach exquisite levels of gene expression. Among such transcription factors there is a subset of highly specialized nuclear factors with features resembling master regulators with a large variety of functions. This is turning to be the case of the multifunctional nuclear factor CCCTC-binding protein (CTCF) which is involved in gene regulation, chromatin organization, and three-dimensional conformation of the genome inside the cell nucleus. Technically its study has turned to be challenging, in particular its posttranscriptional interference by small interference RNAs. Here we describe three main strategies to downregulate the overall abundance of CTCF in culture cell lines.

**Key words** CTCF, Lentivirus, Interference RNA, Inducible vectors, Lipofection, Electroporation, Fluorescent cytometry

---

### 1 Introduction

Over the years the progress in the capacity to manipulate the levels of a particular messenger RNA and/or protein synthesis in cultured cells or even in organisms has contributed to the understanding in more detail many cellular processes [1]. From time to time, the RNA interference strategy is inefficient or even deleterious for the cell. One example is the knockdown of the multifunctional 11-zinc-finger protein CTCF [2, 3]. CTCF functions are so varied; it is not surprising that its manipulation can dramatically alter many cellular processes rendering its study difficult.

CTCF was initially discovered as a transcriptional repressor of the *c-myc* gene in chicken embryonic fibroblasts [4]. CTCF versatility has also been documented since it can also act as transcriptional

activator, in particular, based on its capacity to employ distinct combinatorial of the zinc fingers to bind numerous and highly divergent sequences [5–7]. Furthermore, CTCF can interact with itself and with a large variety of cofactors [6, 7]. Another interesting aspect is that the binding of CTCF to a subset of recognition sequences is sensitive to DNA methylation [6, 8, 9].

In addition to the functional consequences of the use of different zinc fingers, CTCF factor can be posttranslationally modified. CTCF can be phosphorylated in the C-terminal domain of the protein and poly-ADP-ribosylated in the N-terminal domain [10, 11]. Interestingly, CTCF poly-ADP-ribosylation is associated with the 180 kDa form that is present in normal breast tissues, in contrast to the 130 kDa protein, which is present in breast tumors [11]. More recently, it has been demonstrated that CTCF can be sumoylated promoting relaxation of the chromatin structure and transcriptional activation [12, 13].

The contribution of CTCF to cell homeostasis and organism development has also been documented. For example, *ctcf*<sup>-/-</sup> knockout mice show a lethal phenotype in early stages of development (before implantation). In contrast, its over-expression in breast cancer cell lines and tumors causes resistance to apoptosis [3, 14]. Conditional knockout mice have shown that the absence of CTCF in mouse limb affects its development increasing apoptosis of limb cells [15]. CTCF also participates in cell differentiation, as in the case of myogenic differentiation by regulating MyoD myogenic potential [16]. Another related aspect has to do with the interaction between CTCF and the TBP-associated factor TAF3, an interaction that confers differentiation capabilities to stem cells and gives rise to different somatic cell types [17].

Together, CTCF has a central role in the control of many processes that converge in the control of the cell differentiation, organism development, and, furthermore, cell cycle since CTCF has also been involved in cancer cells epigenetically regulating tumor-suppressor genes like *p53*, *Rb*, and *p16<sup>NK4a</sup>* and the proto-oncogene *c-myc*, among others [18, 19].

Furthermore, CTCF has been linked to the transition between open chromatin and highly compacted genomic regions. Its location in transition zones of open and repressive chromatin led to the idea that CTCF represents a chromatin insulator or boundary element [20]. Based in this view, two functional definitions emerge for the insulator regulatory elements. The first one has to do with the ability of an insulator to block promoter–enhancer communication in a position-dependent manner [21]. The second one defines an insulator as a sequence together with its associated factors (among them CTCF) that can protect a transgene against chromosomal position effects [22–24]. Derived from this second functional definition, CTCF and its insulator properties have been proposed as a barrier element that prevents the spreading of repressive

heterochromatin [20]. Therefore, CTCF-dependent insulators contribute to the chromatin organization of the genome.

Nowadays, this view has been reinforced by the experimental evidence showing CTCF as one of the most relevant factors involved in long-range chromatin–fiber interactions. These interactions were initially described in *Drosophila* and now observed in many genomic loci including the *Igf2/H19*, the mouse  $\beta$ -globin locus, and the immunoglobulin heavy-chain locus, among others [25–27]. This alternative activity of CTCF is apparently relevant for the genome organization in the nuclear space through the formation of domains and large chromosomal regions associated with the nuclear lamina (also known as LADs) [28].

Here we present a broad overview of CTCF and its varied functions. For several years now, our research group and many others have been working in the generation of different CTCF knockdowns to manipulate the cellular levels of this protein. This useful experimental strategy has allowed us to understand and reveal new functions for CTCF. Unfortunately, we have been confronted with several technical difficulties in addition to the pleiotropic effects due to a reduction in the cellular levels of the CTCF factor. Thus, in our experience, we observed that the transduction of interference RNAs against CTCF requires several considerations that we discuss here in detail. We address these aspects from the perspective of the generation of stable cell lines expressing small hairpin interference RNAs (shRNAi), their transduction taking advantage of lentiviral particles, and, alternatively, the use of an inducible system. These three different strategies are described considering their advantages and limitations.

### **1.1 CTCF Knockdown Stable Transfection**

Stable transfection can provide long-term expression of the integrated DNA, in contrast to transiently transfected DNA. Stably transfected cells segregate the foreign DNA to their progeny, because the transfected DNA is apparently incorporated randomly into the genome. This kind of transfection is useful for production of recombinant proteins and analysis of short- and long-term effects of exogenous transgene expression [29].

Nowadays, the most common way to deliver DNA into a cell is via the lipofectamine reagent. Lipofectamine is a polycationic synthetic lipid mixed with a fusogenic lipid with an amine group. This lipid mixture forms DNA–lipid complexes due to ionic interactions between the head group of the lipid with a strong positive charge, which neutralizes the negative charge of phosphate groups of DNA. Lipid–DNA mixture results in the formation of structures that fuse and pass across the plasma membrane to deliver DNA into the cell [30].

Electroporation is another strategy to introduce DNA in a cell. It is a mechanical method that uses an electrical pulse to create temporary pores in cell membranes through which substances like



nucleic acids can be introduced into cells. Host cells and selected molecules must be suspended in a conductive solution, where an electrical circuit is closed around the mixture, creating an electrical pulse during a short period of time. This disturbs the phospholipid bilayer of the membrane and causes the formation of temporary pores through which charged molecules, such as DNA, are driven across the membrane. This method is mainly recommended for tissue culture cells [30]. It is worth mentioning that electroporation can cause significant cell death due to the electrical pulse.

Whichever the method, the next key step in a stable transfection is to select cells in which the transgene is stably integrated. Usually, transfected DNA includes an antibiotic resistance gene as part of the construct in order to perform drug selection of the cells after a short time of recovery. Cells that were not transfected will die, and those that express the antibiotic resistance gene at sufficient levels will survive. G418 (Geneticin), hygromycin B, puromycin, blasticidin, and Zeocin compounds are the most commonly used antibiotics to select stably transfected cells.

## **1.2 Manipulating the Levels of the Nuclear Factor CTCF**

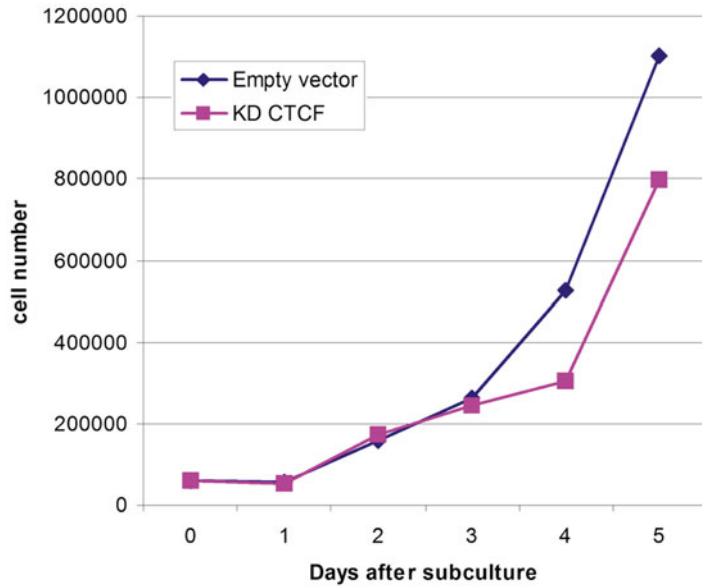
Different study groups have performed stable transfection of shRNAi against CTCF in several cell lines. The effect of these transfections appears to be quite variable depending on the system used as well as the cell type. For example, we have found that the stable transfection of shRNAi against CTCF is for example very unstable in HeLa cells. In order to investigate the role of CTCF in the maintenance of expression of latency genes in Epstein–Barr virus-infected B-cell lines (Kem I and Mutu I cell lines), Sample and collaborators carried out a knockdown of CTCF in Kem I and Mutu I cells using SureSilencing shRNA expression plasmids (SABiosciences) that also encodes for hygromycin or puromycin resistance [31]. The SureSilencing shRNA plasmids are designed to knock down individual genes under stable transfection conditions after appropriate selection procedures. Each vector contains the shRNA under the control of the UI promoter and a neomycin or a puromycin resistance gene (<http://www.sabiosciences.com/RNAi>). They transfected  $3 \times 10^6$  cells with a plasmid via Amaxa Nucleofection, which is an electroporation commercial method. Cells were transfected with four different shRNA plasmids, using 5 µg of each, and control cells received 20 µg of control shRNA plasmid. Two days after Nucleofection, cells were selected with puromycin (Kem I 200 ng/ml and Mutu I 2 µg/ml). In both cell lines the knockdown was confirmed by immunoblotting. According to the manufacturer of Amaxa Nucleofection, this method has little effect on the viability of the cells, although there are no comments regarding it.

MCF-7 breast cancer cell line has also been transfected with siRNA against CTCF. Since CTCF is distributed differentially in a cell type-specific manner along the *kcnq5* locus, this research group

wanted to know if the distribution of CTCF determined the interactions occurring in the gene locus [32]. The RNA interference experiments were performed according to the siRNA strategy designed by Thermo Scientific Dharmacon (ON-TARGET plus siRNA Reagents). This siRNA technology is different from others because it has a dual-strand modification that allows the reduction of nonspecific targets. First, the sense strand is modified to prevent uptake by the RISC complex, which favors the antisense strand loading into the complex [33, 34]. Next, the antisense strand is modified to destabilize off-target activity and enhance target specificity. This is mainly achieved by excluding siRNA with known microRNA seed region motifs and preferentially choosing lower frequency seed regions [35, 36]. They used the ON-TARGET plus non-targeting siRNA as a control. The siRNA transfection of MCF-7 cells was performed using DharmaFECT 1 in 6-well plates. The DharmaFECT 1 transfection protocol is based on the lipofectamine reagent (Dharmacon Inc. 2006).

This system was also used in primary cells that were transfected with siRNA against CTCF. Such is the case of erythroid cells extracted from spleen of transgenic mice that were genetically modified with an extra CTCF site between their endogenous CTCF insulator (HS5) and the  $\alpha$ -globin locus [37]. In order to probe the CTCF dependence of the insulator loop, they isolated erythroid cells from mice and cultured them. These cells were also transfected with the mouse CTCF ON-TARGET plus siRNA system with the Amaxa Biosystems Nucleofector, as described previously. In this case, the cells were harvested 72 h after the transfection.

The knockdown of CTCF by stable transfection of a siRNA has also been carried out in embryonic stem cells (ESCs). It is well known that CTCF has a cell type-specific distribution, including human ESCs. Single cells were collected and transfected with a predesigned Silencer Select siRNA (Ambion) against CTCF [38]. The Silencer Select siRNA system presumably incorporated improvements in siRNA design, such as prediction algorithms that allow the siRNA to recognize its target in a more specific manner and reduce off-targets, increasing their efficiency up to a 100-fold. In this system siRNAs also have a locked nucleic acid (LNA) chemical modification that also influences the off-target reduction ([www.invitrogen.com](http://www.invitrogen.com)). For the transfection they employed RNAiMAX (Invitrogen)-based lipofectamine. Importantly, a second round of transfection was performed 24 h later. They proved that proliferation of hESCs was affected by the ablation of CTCF since they incorporated 60 % less BrdU, and it also decreases the expression of pluripotency and self-renewal genes. This data suggests that CTCF may have different roles in the regulation of stem cells, which is no wonder given its varied distribution along the genome and the amount of genes and nuclear organization it may be regulating.



**Fig. 1** Growth curves for a control HeLa cell line and a CTCF knockdown HeLa cell line grown in culture. There is a decrease in the population doubling time in the CTCF knockdown cell line compared with the control cell line

In conclusion, there are numerous ways to carry out a stable transfection to perform a CTCF knockdown. But, it is important to mention that an effective knockdown generated by a stable transfection often depends on the transfection capacity of a cell type and its own phenotype. For example we have reproducibly observed differences in the timing of cell division in stable CTCF knockdown context versus a control (Fig. 1). It is also important to design a highly specific siRNA with a strong ability to decrease its target gene without affecting other genes. The site of integration of the transgene has also shown to be quite influential in the knockdown process. Next, we present the detailed protocol for the generation of a stable knockdown against CTCF.

## 2 Materials

### 2.1 Stable Transfection of Interference RNA Against CTCF

#### 2.1.1 Cell Culture

1. HeLa cell culture is performed in a fresh Dulbecco MEM medium (with 10 % fetal bovine serum (FBS) and 1 % penicillin/streptomycin).
2. Cells are incubated at 37 °C and 5 % CO<sub>2</sub>.

#### 2.1.2 Plasmid Preparation

1. Linearize the plasmid with the convenient enzyme (*see Note 1*).



## 2.2 Pseudo-Lentiviral Particle Production

### 2.2.1 Plasmids for Lentiviral Particle Production (See **Note 2**)

### 2.2.2 Cell Culture

### 2.2.3 Solutions and Materials

1. Transfer, envelope, and packaging plasmids must have high quality for transfection. Use commercial kits to purify DNA and avoid phenol–chloroform extractions.

1. HEK293T (high titer of infectious virus) or HEK293 (parental cell line) are human embryonic kidney cell lines. Both are commonly used for virus production, and their low passing number in culture is critical to obtain high viral titers.
2. HEK293T are cultured in DMEM: Dulbecco's Modified Eagle Medium (GIBCO), supplemented with 10 % FBS, pyruvate, glutamine, and penicillin/streptomycin 1× antibiotics at 37 °C, 5 % CO<sub>2</sub>, in a 150 mm dish.

0.1 % Sodium hypochlorite solution.

1.25 M calcium chloride solution in H<sub>2</sub>O, filter sterilized and aliquoted.

0.25 % Trypsin–EDTA.

15 ml centrifuge tube.

HeBS 2×.

Ultra-clear centrifuge tubes.

Swinging bucket rotor (Beckman SW28 or SW32).

Polybrene (hexadimethrine bromide).

Puromycin dihydrochloride or the corresponding antibiotic.

Epifluorescence microscope.

## 2.3 Inducible System

In many fields of basic research, temporal expression of a gene can be needed depending on the experimental setting or, in worse cases, large exposures or high amounts of a transcript can be toxic to the cell, so it would be necessary to turn on and off their expression [41]. The tetracycline-dependent transcriptional regulatory system is one of the best studied inducible systems with proven efficacy in vitro and in vivo [42]. This system is based on the *E. coli* Tn10 tetracycline resistance operator, and its components were improved by mutagenesis and codon optimization to obtain variants with reduced basal activity and greater sensitivity to doxycycline [43].

Currently, lentiviral vectors were developed to incorporate all elements of the doxycycline-inducible system in one or two vectors, and choosing between them will depend on individual applications. One method that is frequently used in this inducible lentiviral vector is the silencing of a gene of interest by interference RNA. The latest generation of this kind of vectors uses microRNA sequences that have been adapted to incorporate siRNA sequences [44]. One example of this kind of vectors is the pTRIPZ lentiviral inducible vector that combines the design advantages of microRNA-adapted shRNA (shRNAmir) with a lentiviral induc-

ible vector to create a powerful RNAi trigger capable of producing RNAi in most cell types (Open Biosystems). The vector is engineered to be Tet-On and produces tightly regulated induction of shRNA expression in the presence of doxycycline. With the aim to generate a knockdown of CTCF under controlled conditions, it seems to be convenient to use an inducible system to turn on/off the interference RNA expression whenever the inductor is present or not (Fig. 1). Furthermore, when CTCF is knocked down in a stable context it is harder to sustain shRNA expression for several days and even more difficult to recover the transgenic line after freezing the cells. In conclusion it seems that this way to perform the knockdown of an essential protein is more effective and offers many advantages over a constitutive knockdown.

### 2.3.1 Solutions

1 mg/ml Doxycycline solution in H<sub>2</sub>O: Sterilize the solution by filtration and store aliquoted at -20 °C.

---

## 3 Methods

### 3.1 Stable Transfection of an Interference RNA Against CTCF

#### Day 1

1. For each transfection  $3 \times 10^5$  cells are placed in 3 ml of complete DMEM; we recommend to perform it in triplicates (*see Note 3*).

#### Day 2

1. Remove the medium of the cells, and add 800  $\mu$ l of DMEM lacking FBS and antibiotic (DMEM (-)).
2. Mix 2  $\mu$ l of lipofectamine and 98  $\mu$ l of DMEM (-) and incubate for 15 min (*see Note 4*).
3. Next add the desired amount of DNA in a volume of 100  $\mu$ l of DMEM (-) to the lipofectamine mix and incubate for 30 min (*see Note 5*).
4. Add the final mix to the cells to reach a final volume of 1 ml and incubate at 37 °C and 5 % CO<sub>2</sub> for at least 6 h.
5. Add 3 ml of complete medium.

#### Day 4

1. Change the cells to a selective medium containing antibiotic at the previously established concentration (*see Note 6*).

#### Day 5

1. Harvest the cells, and perform a total protein extraction to verify the abundance of CTCF by Western blot (*see Note 7*).

## 3.2 **Lentiviral Particle Production**

Extreme caution is recommended when working with lentivirus.

([http://oba.od.nih.gov/oba/rac/Guidance/LentiVirus\\_Containment/pdf/Lenti\\_Containment\\_Guidance.pdf](http://oba.od.nih.gov/oba/rac/Guidance/LentiVirus_Containment/pdf/Lenti_Containment_Guidance.pdf).)

### 3.2.1 *Cell Culture*

#### *Day 1*

1. Plate  $4 \times 10^6$  HEK293T cells in two 150 mm culture dishes each, at least 12 h before the viral plasmid co-transfection.
2. The confluence of cells must be 60–70 %; this is a critical aspect for the efficient production of lentivirus particles.

### 3.2.2 *Co-transfection of HEK293T Cells*

#### *Day 2*

1. Mix 15  $\mu\text{g}$  transfer DNA, 10  $\mu\text{g}$  gag/pol-expressing plasmid, 10  $\mu\text{g}$  rev/tat-expressing plasmid, and 5  $\mu\text{g}$  VSV-G envelope plasmid in a 15 ml centrifuge tube, and make up to 450  $\mu\text{l}$  with sterile water.
2. Add dropwise constantly 50  $\mu\text{l}$  of 2.5 M  $\text{CaCl}_2$ .
3. Add 500  $\mu\text{l}$  HeBS 2 $\times$  in a dropwise fashion.
4. Incubate at room temperature for 20 min.
5. Add dropwise the transfection mixture across the plate of HEK293T.
6. Incubate the cells at 37 °C, 5 %  $\text{CO}_2$ , for 8 h.
7. Aspirate medium, wash with PBS 1 $\times$ , and add 16 ml of fresh medium.

### 3.2.3 *Collect Viral Supernatants*

#### *Day 4*

1. Centrifuge supernatant at  $778 \times g$  for 5 min to clear debris (*see Note 8*).
2. Supernatant can be stored at 4 °C for a couple of days prior to concentration.

#### *Day 5*

1. Collect and centrifuge supernatant at  $778 \times g$  for 5 min to remove debris.
2. The total volume of collected supernatant along this time must be cleared with a 0.45  $\mu\text{m}$  filter to remove viral aggregates and debris (*see Note 9*).
3. At this point the supernatant can be used to infect target cells or concentrate the viral particles (*see Note 10*).
4. Preferably aliquot the supernatant before storage to avoid thawing several times.
5. Store at  $-80$  °C (*see Note 10*).

### 3.2.4 Lentivirus Concentration

#### Day 6

1. Pre-chill rotor and swinging buckets.
2. Add 32 ml of the total amount of lentiviral supernatant in each ultra-clear centrifuge tube.
3. Centrifuge at  $82,705 \times g$  for 2–2.5 h at 4 °C (*see Note 11*).
4. Discard the supernatant by decantation in a container with 10 % bleach.
5. Resuspend the lentiviral pellet (it may be translucent or invisible) with 60  $\mu$ l of chilled PBS 1 $\times$  pH 7.4 and let stand for 8–12 h at 4 °C.
6. Make 10  $\mu$ l aliquots in clearly labeled microcentrifuge tubes.
7. Froze in liquid nitrogen, and store at –80 °C.
8. Calculate the titer of the lentivirus by transducing target cells with serially diluted viral preparations.

### 3.2.5 Titration of Lentivirus Expressing Fluorescent Proteins by Flow Cytometry (FACS)

#### Days 7, 8

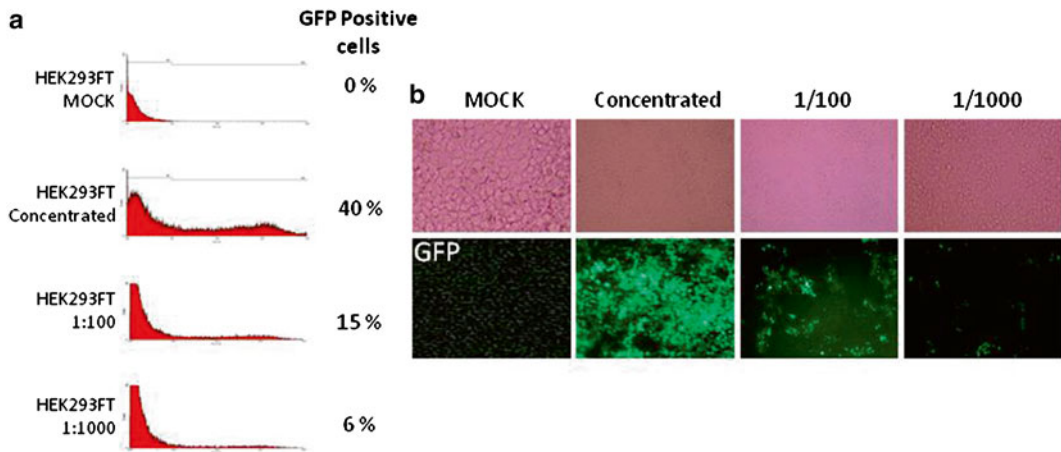
1. Plate HEK293T cell line at a density of  $1 \times 10^5$  cells per well in 6-well plate 12–24 h before the transduction.
2. Remove the medium from the wells and wash twice with PBS 1 $\times$  pH 7.4.
3. Add 0.5 ml of serum- and antibiotic-free fresh medium with 8  $\mu$ g/ml of polybrene.
4. Generate a tenfold serial dilution of the lentiviral preparation: Mock, undiluted, 10, 100, 1,000, and 10,000 (*see Note 12*).
5. Grow the cells for 48 h.
6. Collect the cells by trypsinization.
7. Determine the percentage of fluorescent cells by FACS (*see Figs. 2 and 3*).
8. Calculate lentiviral titer according of the following formula:

$$\text{TU}/\mu\text{l} = (\% \text{GFP-positive cells} \times 100,000) / (100 \times \text{volume added to each } \mu\text{l per well} \times \text{DF})$$

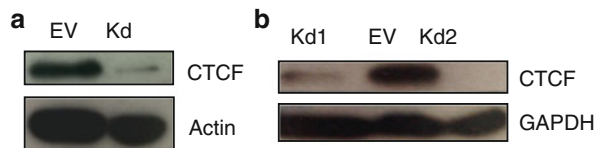
Where TU = transduced units (Fig. 3).

DF = Dilution factor.

Dilution	Dilution factor
Undiluted	1
1/10	$10^{-1}$
1/100	$10^{-2}$
1/1,000	$10^{-3}$
1/10,000	$10^{-4}$



**Fig. 2** HEK293-infected cell line with *GFP*-expressing lentivirus. **(a)** Flow cytometry profiles of the different dilutions used. The percentages of GFP-positive cells of each dilution are shown. **(b)** Epifluorescence microscope images of the expression of *GFP*, which correspond to each condition described in **(a)**



**Fig. 3** Western blot of knockdown of CTCF. **(a)** Infection with lentivirus in U87MG glioblastoma cell line. Actin antibody was used as a loading control. **(b)** Western blot of knockdown of CTCF by infection with lentivirus in U87MG glioblastoma cells. The first line corresponds to one round of infection, the second line is the control of empty vector, and the third line corresponds to U87MG cells with two rounds of infection with lentivirus

- To obtain an accurate titer, average values from at least two lentiviral dilutions are required.  $TU/\mu l$  is the theoretical quantity of virus particles capable of infecting the same number of cells. For example, to integrate one copy of lentivirus vector in the genome of 100,000 cells it is necessary to use the volume ( $\mu l$ ) corresponding to 100,000 TU. Multiplicity of infection (MOI) is the calculated number of lentiviral copies per host genome; thus, MOI depends on  $TU/\mu l$  and the total cell number used during infection. MOI values between 10 and 100 MOIs are the most frequently used.

### 3.2.6 Selection and Amplification of the Transduced Cell Line

#### Day 9

After transduction, cells must be drug selected. To accomplish this, many lentiviral systems contain the cDNA of fluorescent proteins and/or the resistance cassette for any antibiotic (i.e., puromycin,

ampicillin, G418). There are many ways to perform the selection of positive clones; we will briefly describe three of them.

### 3.2.7 Manual Selection

Taking advantage of the coupled expression of fluorescent proteins it is possible to enrich the infected population by picking the spot with the highest fluorescent cells identified by an epifluorescence microscope. A single cell with high fluorescence is indicative of a multiple infection event. This kind of selection allows the isolation of colonies with the best efficiency of infection.

1. Using an epifluorescence microscope, search for the highest fluorescent signal in the visual field and spot it in the plate with a marker.
2. Remove the medium, and wash the cells with PBS.
3. Using sterile tweezers place a cloning disc soaked in trypsin over the previously marked spot, and wait for around 5 min.
4. Remove the cloning disc, and rinse it in a new plate with fresh medium in order to expand the selected clone (*see Note 13*).

### 3.2.8 Sorting by Flow Cytometry (FACS)

Based on the same principle described previously, it is possible to use flow cytometry cell sorting for a quantitative selection of fluorescent colonies in the entire infected culture (Fig. 2a).

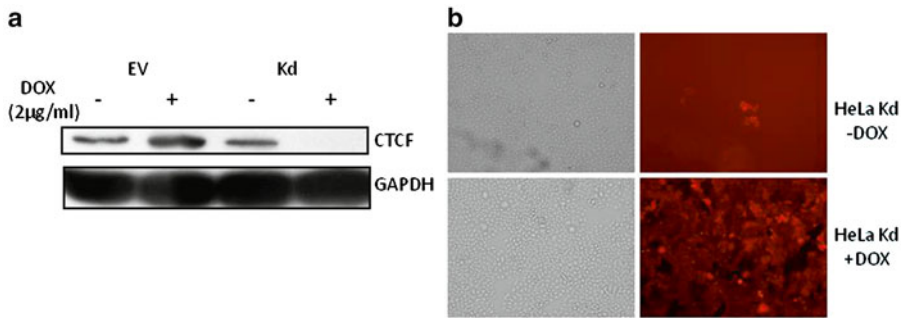
1. Harvest ~1 million infected cells in PBS.
2. Determine the profile for the positive fluorescence cells in the FACS.
3. Sort the cells, selecting those with the highest fluorescence emission.
4. Expand enriched cell culture.
5. Days later check again the selected culture in the FACS to corroborate that the levels of fluorescence are maintained (Fig. 2b).

### 3.2.9 Selection by Antibiotic Resistance (See Note 14)

1. After infection cells are recovered and cultured in the presence of 1 µg/ml of puromycin or the appropriate antibiotic (*see Note 15*).
2. In parallel, add the same concentration of antibiotic to a non-infected culture, as a control to define the time of death.
3. The infected cells will be selected when the control cell culture (without resistance gene) is death.

## 3.3 Inducible System

The methodology used for transduction of this lentiviral system is the same as described above. Depending on the resistance cassette of the vector, select the transduced culture with the proper antibiotic in the previously established concentration. After 3–5 days of drug selection, proceed with the following steps for induction.



**Fig. 4** Inducible system for the CTCF knockdown. (a) Western blot for CTCF of the different HeLa cell lines with or without induction of doxycycline with their respective loading control using a GAPDH antibody. (b) Induction of shRNAmir (tracked by TurboRFP) with doxycycline is tightly regulated. HeLa cells were transfected with the different plasmids (empty vector *upper panels* and the shRNAi against CTCF at the *bottom*) selected with puromycin (2 µl/ml) for 96 h after which 2 µg/ml of doxycycline was added to the cells and TurboRFP expression was assessed at 72 h

#### Day 1

1. Remove the medium with the selection antibiotic, and add fresh medium with the desired concentration of induction reagent, doxycycline (0.1–2 µg/ml) (*see Note 16*).
2. Repeat the process every 24 h for the next 48–72 h (*see Note 17*).

#### Days 3, 4

1. If the vector employed has a reporter gene (TurboRFP, GFP, etc.), it is possible to evaluate in a fluorescent microscope the expression of the transgene after induction, comparing with a non-induced infected cell context (Fig. 3).
2. Evaluate by the desired method (Western blot, RT-PCR) the knockdown after induction, again, comparing with non-induced infected cells (Fig. 4).

---

## 4 Notes

1. The plasmid we use for CTCF knockdown is a siRNA against CTCF cloned in a pSilencer vector (Ambion).
2. Viral vectors emerge as a method for modifying genetically target cells *in vitro* and *in vivo*. There are different types of virus used to transfer genetic information including adenoviruses, retroviruses, adeno-associated viruses, herpes simplex virus, and flaviviruses [39]. In particular, lentiviruses are members of the *Retroviridae* family that infect primarily vertebrates and are associated with chronic immune disease. Most of the lentiviral vectors used are derived from the human HIV-1. This family



replicates through the conversion of viral RNA genome into double-stranded DNA. Lentiviruses have the capability to integrate into the genome of dividing and nondividing cells and show stable long-term transgene expression. During the last decade, lentiviruses have provided a delivery platform for several gene therapy clinical approaches [39, 40].

3. For cells in suspension you can count them on the same day of the transfection, and culture them in their respective non-complemented medium.
4. To establish the amount of lipofectamine required, we begin by using a proportion of 1:2 (1  $\mu\text{g}$  of DNA:2  $\mu\text{g}$  of lipofectamine). Otherwise, different concentrations can be tested.
5. We recommend using 1  $\mu\text{g}$  of plasmid for each transfection.
6. Antibiotic concentration for selection depends on the cell type; we suggest using the highest possible avoiding massive cell death.
7. In the knockdown cell line the maximum decrease reached in this kind of assays is about 30–40 %. In our experience it seems that the expression of the plasmid is very unstable since in a couple of days the protein levels of CTCF become similar to those observed in the control line transfected with an empty vector, although the cells continue to survive in the selective medium. The same effect was observed when we froze the knockdown CTCF cell line and later tried to use it for different purposes.
8. Since the highest production of viral particles occurs between 48 and 72 h post-transfection, 16 ml should be sufficient to maintain the cell culture in optimal conditions. Sometimes the cell culture reaches a high confluence earlier; in that case collect virus supernatant and replace with fresh medium.
9. The supernatant contains viral particles, and you must be careful with handling.
10. Virus can be harvested up to 4 days post-transfection.
11. Transport the tubes carefully to avoid disturbing the viral pellet.
12. Avoid pipetting volumes  $<10 \mu\text{l}$  to minimize dilution errors.
13. Repeat the procedure with as many different individual colonies as you want to test.
14. This method is the most practical and direct, although the selected cell culture possesses a more variable phenotype due to the different efficiencies of infection for each cell type.
15. It is possible to carry out a concentration curve to determine the ideal selection condition, in which the cell viability is not greatly affected. Depending on the efficiency of infection and the cell type, the concentration of antibiotic can be increased. We suggest testing a range going from 1 to 10  $\mu\text{g}/\text{ml}$ .

16. The amounts of doxycycline necessary vary due to the infection efficiency of the particular cell line as well as the amounts of DNA used.
17. In the case of CTCF knockdown, by 48 h after induction it is possible to observe an effect, although the highest levels of expression of the transgene have been detected at 72 h post-induction.

---

## Acknowledgements

We acknowledge Ricardo Saldaña-Meyer for comments and critical reading of the manuscript. This work was supported by the Dirección General de Asuntos del Personal Académico-Universidad Nacional Autónoma de México (IN209403 and IN203811), Consejo Nacional de Ciencia y Tecnología, México (CONACyT; 42653-Q and 128464), and a Ph.D. fellowship from CONACyT and Dirección General de estudios de Posgrado-Universidad Nacional Autónoma de México (DGEP) (EG-B: 207989, RP-M: 173754, and EA-O: 256339). Additional support was provided by the Ph.D. Graduate Program, “Doctorado en Ciencias Biomédicas and Ciencias Bioquímicas.” We thank Fernando Suaste-Olmos for technical assistance.

## References

1. Boudreau RL, Davidson BL (2012) Generation of hairpin-based RNAi vectors for biological and therapeutic applications. *Methods Enzymol* 507:275–296
2. Phillips JE, Corces VG (2009) CTCF: master weaver of the genome. *Cell* 137:1194–1211
3. Herold M, Bartkuhn M, Renkawitz R (2012) CTCF: insights into insulator function during development. *Development* 139:1045–1057
4. Lobanenkov VV, Nicolas RH, Adler VV, Paterson H, Klenova EM, Polotskaja AV et al (1990) A novel sequence-specific DNA binding protein which interacts with three regulatory spaced direct repeats of the CCCTC-motif in the 5' flanking sequence of the chicken c-myc gene. *Oncogene* 5:1743–1753
5. Philippova GN, Qi CF, Ulmer JE, Moore JM, Wards MD, Hu YJ et al (2002) Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter its DNA-binding specificity. *Cancer Res* 62:48–52
6. Ohlsson R, Lobanenkov V, Klenova E (2010) Does CTCF mediate between nuclear organization and gene expression? *BioEssays* 32:37–50
7. Valadez-Graham V, Razin SV, Recillas-Targa F (2004) CTCF-dependent enhancer blocker at the upstream region of the chicken  $\alpha$ -globin gene domain. *Nucleic Acids Res* 32:1354–1362
8. Engel N, West AG, Felsenfeld G, Bartolomei MS (2004) Antagonism between DNA hypermethylation and enhancer-blocking activity at the H19 DMD is uncovered by CpG mutations. *Nat Genet* 36:883–888
9. De La Rosa-Velázquez IS, Rincón-Arano H, Benítez-Bribiesca L, Recillas-Targa F (2007) Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Res* 67:2577–2585
10. El-Kady A, Klenova E (2005) Regulation of the transcription factor, CTCF, by phosphorylation with protein kinase CK2. *FEBS Lett* 579:1424–1434
11. Docquier F, Jita GX, Farrar D, Jat P, O'Hare M, Chernukhin I et al (2009) Decreased poly(ADP-ribosyl)ation of CTCF, a transcription factor, is associated with breast cancer phenotype and cell proliferation. *Clin Cancer Res* 15:5762–5771

12. Kitchen NS, Schoenherr CJ (2010) Sumoylation modulates a domain in CTCF that activates transcription and decondenses chromatin. *J Cell Biochem* 11:665–675
13. MacPherson MJ, Beatty LG, Zhou W, Du M, Sdowski PD (2009) The CTCF insulator protein posttranslationally modified by SUMO. *Mol Cell Biol* 29:714–725
14. Docquier F, Farrar D, D'Arcy V, Chernukhin I, Robinsin AF, Loukinov D et al (2005) Heightened expression of CTCF in breast cancer cells is associated with resistance to apoptosis. *Cancer Res* 65:5112–5122
15. Soshnikova N, Montavon T, Leleu M, Galjart N, Duboule D (2010) Functional analysis of CTCF during mammalian limb development. *Dev Cell* 19:819–830
16. Delgado-Olguín P, Brand-Arzamendi K, Scott IC, Jungblut B, Stainier DY, Bruneau BG et al (2011) CTCF promotes muscle differentiation by modulating the activity of myogenic regulatory factors. *J Biol Chem* 286:12483–12484
17. Liu Z, Scannell DR, Eisen MB, Tjian R (2011) Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell* 146:720–731
18. Fiorentino FP, Giordano A (2011) The tumor suppressor role of CTCF. *J Cell Physiol* 227:479–492
19. Recillas-Targa F, De La Rosa-Velázquez IA, Soto-Reyes E (2011) Insulation of tumor suppressor genes by the nuclear factor CTCF. *Biochem Cell Biol* 89:479–488
20. Barkess G, West AG (2012) Chromatin insulator elements: establishing barriers to set heterochromatin boundaries. *Epigenomics* 4:67–80
21. Burgess-Beusse B, Farrell C, Gaszner M, Litt M, Mutskov V, Recillas-Targa F et al (2002) The insulation of genes from external enhancers and silencing chromatin. *Proc Natl Acad Sci USA* 4:16433–16437
22. Pikaart M, Recillas-Targa F, Felsenfeld G (1998) Loss of transcriptional activity of the transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes Dev* 12:2852–2862
23. Recillas-Targa F, Valadez-Graham V, Farrell CM (2004) Prospects and implications of using chromatin insulators in gene therapy and transgenesis. *BioEssays* 26:796–807
24. Furlan-Magaril M, Rebollar E, Guerrero G, Fernández A, Moltó E, González-Buendía E et al (2011) An insulator embedded in the Chicken  $\alpha$ -globin locus regulates chromatin domain configuration and differential gene expression. *Nucleic Acids Res* 39:89–103
25. Splinter E, Heath H, Kooren J, Palstra RJ, Klaus P, Grosveld F et al (2006) CTCF mediates long-range chromatin loop and local histone modification in the  $\beta$ -globin locus. *Genes Dev* 20:2349–2354
26. Kurukuti S, Tiwari VK, Tavosoidana G, Pugacheva E, Murrell A, Zhao Z et al (2006) CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc Natl Acad Sci USA* 103:10684–10689
27. van Bortle K, Corces VG (2013) The role of chromatin insulators in nuclear architecture and genome function. *Curr Opin Genet Dev* 23(2):212–218
28. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W et al (2008) Domain organization of human chromosomes revealed by mapping the nuclear lamina interactions. *Nature* 453:948–951
29. Bonetta L (2005) The inside scoop-evaluating gene delivery methods. *Nat Methods* 2:875–883
30. Sambrook J, Russell DW. *Molecular cloning, a laboratory manual*, 2(4):15.1–15.5
31. Hughes D, Marendy E, Dickerson C, Yetming K, Sample C, Sample J (2012) Contribution of CTCF and DNA methyltransferases DNMT1 and DNMT3B to Epstein-Barr Virus restricted latency. *J Virol* 86:1034–1045
32. Ren L, Wang Y, Shi M, Wang X, Yang Z, Zhao Z (2012) CTCF mediates the cell-type specific spatial organization of the *Kcnq5* locus and the local gene regulation. *PLoS One* 7:e31416
33. Jackson AL, Burchard J, Leake D, Reynolds A, Schelter J, Guo J et al (2006) Position-specific chemical modification increases specificity of siRNA-mediated gene silencing. *RNA* 12:1197–1205
34. Chen PY, Weinmann L, Gaidatzis D, Pei Y, Zavolan M, Tuschl T et al (2008) Strand-specific 5'-O-methylation of siRNA duplexes controls guide strand selection and target specificity. *RNA* 2:263–274
35. Birmingham A, Anderson EM, Reynolds A, Ilsley-Tyree D, Leake D, Fedorov Y et al (2006) 3'-UTR seed matches, but not overall identity, are associated with RNAi off-targets. *Nat Methods* 3:199–204
36. Anderson EM, Birmingham A, Baskerville S, Reynolds A, Maksimova E, Leake D et al (2008) Experimental validation of the importance of seed frequency to siRNA specificity. *RNA* 14:853–861
37. Hou C, Zhao H, Tanimoto K, Dean A (2008) CTCF-dependent enhancer-blocking by alter-

- native chromatin loop formation. *Proc Natl Acad Sci USA* 105:20398–20403
38. Balakrishnan S, Witcher M, Berggren T, Emerson B (2012) Functional and molecular characterization of the role of CTCF in human embryonic stem cell biology. *PLoS One* 7:e42424
  39. Klimatcheva E, Ronsenblatt JD, Planelles V (1999) Lentiviral vectors and gene therapy. *Front Biosci* 4:D481–96
  40. Biasco L, Baricordi C, Aiuti A (2012) Retroviral integrations in gene therapy trials. *Mol Ther* 20:709–716
  41. Markusic D, Seppen J (2010) Doxycycline regulated lentiviral vectors. *Methods Mol Biol* 614:69–76
  42. Gossen M, Bujard H (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci USA* 89:5547–5551
  43. Urlinger S, Baron U, Thellmann M, Hasan MT, Bujard H, Hillen W (2000) Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. *Proc Natl Acad Sci USA* 97:7963–7968
  44. Zeng Y, Wagner EJ, Cullen BR (2002) Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol Cell* 9:1327–1333

### Further Reading

DharmaFECT General Transfection Protocol, Dharmacon Inc. 2006, 00033–05-F-03-U  
Ambion by Life Technologies, Silencer Select Pre-designed, Validated and Custom Designed siRNA, Custom Select siRNA, Ambion In Vivo Pre-designed, Custom

Designed and Custom siRNA, 2011, Insert PN4457171 Rev B  
<http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/rnai/Synthetic-RNAi-Analysis/Ambion-Silencer-Select-siRNAs/silencer-select-sirna.html.html>