



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

**PAPEL DE CTCF EN LA TRANSICIÓN CROMATÍNICA ASOCIADA A LA
REGULACIÓN DE LOS GENES ADULTOS α -GLOBINA DE POLLO**

TESIS

**QUE PARA OPTAR POR EL GRADO DE:
Maestro en Ciencias**

**PRESENTA:
GUSTAVO TAPIA URZÚA**

DR. FÉLIX RECILLAS TARGA
[Instituto de Fisiología Celular](#)

MIEMBROS DEL COMITÉ TUTOR
DR. JAIME IVÁN VELASCO VELÁZQUEZ
[Instituto de Fisiología Celular](#)

DRA. VIVIANA DEL CARMEN VALADEZ GRAHAM
[Instituto de Biotecnología](#)

Ciudad de México. Enero de 2017



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.

Agradecimientos académicos

El presente trabajo fue elaborado en el Departamento de Genética Molecular del Instituto de Fisiología Celular de la Universidad Nacional Autónoma de México en el laboratorio del Dr. Félix Recillas Targa, con los apoyos de la Dirección General de Asuntos del Personal Académico (DGAPA-PAPIIT) IN201114, IN203917, del Consejo Nacional de Ciencia y Tecnología (CONACYT) 220503 y de Fronteras de la Ciencias 290.

Los estudios de maestría de Gustavo Tapia Urzúa, con número de CVU 630552, se realizaron con el apoyo de la Beca CONACYT Nacional para posgrados y con el Programa de Apoyo para Estudios del Posgrado (PAEP).

Se agradece y reconoce al comité tutor de este proyecto, constituido por:

Dr. Félix Recillas Targa	Instituto de Fisiología Celular
Dr. Jaime Iván Velasco Velázquez	Instituto de Fisiología Celular
Dra. Viviana del Carmen Valadez Graham	Instituto de Biotecnología

Se agradece el apoyo y la excelente asesoría técnica de la Biol. Georgina Guerrero Avendaño y la Dra. Karin Meier. Asimismo, se reconoce el trabajo de la laboratorista Gianelli Cortés González y el apoyo de la Unidad de Biología Molecular del Instituto de Fisiología Celular en las secuenciaciones de plásmidos realizadas en este proyecto.

Adicionalmente se reconoce la asesoría académica de todos y cada uno de los miembros del laboratorio 122N del Dr. Félix Recillas Targa, en particular del LCG Rodrigo Gacel Arzate Mejía, del Dr. Edgar Fernán González Buendía, del LCG Josué Daniel Cortés Fernández de Lara, del QBP Hober Nelson Nuñez Martínez, del Dr. Christian Gustavo Valdés Quezada y del Médico Carlos Alberto Peralta Álvarez.

Se reconoce también el apoyo de la M. en C. Norma Trejo Medina, de Leticia García Gutiérrez, de Rosa Lilia de la Rocha Viñas y de Julio Ignacio Palacios Ordoñez de la Coordinación del Programa de Posgrado Maestría y Doctorado en Ciencias Bioquímicas de la UNAM.

Finalmente, se agradece a los miembros del jurado, quienes enriquecieron este trabajo con sus observaciones:

Dra. Diana María Escalante Alcalde	Instituto de Fisiología Celular
Dr. Ángel Alfonso Zarain Herzberg	Facultad de Medicina
Dra. Ángeles Concepción Tecalco Cruz	Instituto de Investigaciones Biomédicas
Dr. Héctor Mayani Viveros	Centro Médico Nacional
Dra. Rosa Estela Navarro González	Instituto de Fisiología Celular

ÍNDICE

RESUMEN	1
ABSTRACT.....	2
1. INTRODUCCIÓN.....	3
1.1 La función del genoma	3
1.2 El genoma está organizado en cromatina.....	3
1.3 Papel de la estructura y organización de la cromatina en la expresión diferencial	5
1.3.1 Metilación del ADN.....	7
1.3.2 Modificaciones post-traduccionales en las histonas.....	9
1.3.3 Elementos distales de regulación: <i>Enhancers</i> , Silenciadores e <i>Insulators</i>	11
1.4 El factor nuclear CTCF y la organización tridimensional de la cromatina al interior del núcleo	17
1.5 El sistema de edición del genoma CRISPR-Cas9	22
2. ANTECEDENTES	25
2.1 El dominio α -globina de pollo durante el desarrollo embrionario	25
2.2 Regulación del dominio α -globina de pollo en la diferenciación eritroide	28
3. PLANTEAMIENTO DEL PROBLEMA	29
4. HIPÓTESIS.....	30
5. OBJETIVO GENERAL	30
6. OBJETIVOS PARTICULARES	30
7. MATERIAL Y MÉTODOS.....	31
8. RESULTADOS	34
8.1 CTCF se une <i>in vitro</i> al sitio CTCF-1 del promotor del gen adulto α D	34
8.2 Diseño de gRNAs y oligonucleótidos para la eliminación del sitio de unión CTCF-1 por CRISPR-Cas9	39
8.3 El sistema de edición genómica CRISPR-Cas9 permite eliminar la región correspondiente al motivo CTCF-1	40
a) Eliminación de 33 pb - CRISPR1.....	40
b) Eliminación de 91 pb - CRISPR2.....	45
8.4 La expresión del gen α D adulto disminuye en las células HD3 mutadas con el sistema CRISPR-Cas9	46
9. DISCUSIÓN	48
10. CONCLUSIONES.....	52
11. PERSPECTIVAS.....	52
12. BIBLIOGRAFÍA	53

RESUMEN

La expresión diferencial de genes que ocurre en las células de un organismo eucarionte depende de distintos factores epigenéticos. Entre ellos, la organización tridimensional de la cromatina al interior del núcleo permite regular la expresión de un gen o grupo de genes, esto a través de la formación de asas de cromatina específicas. Existen distintas proteínas que se han visto implicadas en la formación de asas de cromatina siendo CTCF la mejor caracterizada. En este proyecto, utilizamos la activación transcripcional de los genes adultos αD y αA del dominio α -globina de pollo como modelo para estudiar la importancia de la unión de CTCF en la regulación de la expresión génica durante la diferenciación celular. Los genes αD y αA se encuentran silenciados en la línea celular eritroblástica de pollo HD3, para activarlos es necesario inducir la diferenciación de las células HD3 bajo ciertas condiciones experimentales. En estudios previos de nuestro grupo de investigación, se comprobó la unión de CTCF en el promotor del gen αD al diferenciar las células HD3, y en un *enhancer* que se encuentra en la región 3' del dominio (*3'-enh*). Además, por medio de ensayos de captura conformacional de cromosomas se ha observado que al diferenciar las células HD3 se forma un asa de cromatina entre el promotor del gen αD y el *3'-enh*, por lo que se propone como un mecanismo para la activación transcripcional del *locus*. En el presente estudio demostramos que al eliminar total o parcialmente el sitio de unión a CTCF, utilizando el sistema de edición genómica CRISPR-Cas9 en células HD3, disminuyen los niveles de expresión del gen αD en las células mutantes HD3 diferenciadas. Por lo tanto, se favorece el mecanismo de acción en el que se propone que CTCF se une al promotor del gen αD al diferenciar las células HD3 para formar un asa de cromatina con el *3'-enh* del dominio y activar la transcripción de dicho gen.

ABSTRACT

Differential gene expression in an organism depends on epigenetic factors. Among them, chromatin tridimensional organization in the nucleus regulates expression of a single gene or a group of genes, this happens through the formation of specific chromatin loops. Many proteins have been associated with chromatin loops formation but CTCF is the best characterized. In this project, we took advantage of the transcriptional activation of adult chicken α -globin genes (α D and α A) as a model to study the role of CTCF binding in gene expression regulation during cellular differentiation. Genes α D and α A are silenced in the chicken erythroblastic cell line HD3, to activate their expression is necessary to induce HD3 differentiation in the laboratory. Previous reports in our research group showed that CTCF binds in α D gene promoter when HD3 cells differentiate. CTCF was also found in an enhancer located in the 3' region (3'-enh) of the chicken α -globin domain. In addition, it has been reported that HD3 cell differentiation induces the formation of a chromatin loop between α D gene promoter and 3'-enh. Therefore, this mechanism is proposed for the transcriptional activation of the locus. In this study, we prove that total or partial deletion of the CTCF binding motif in the α D gene promoter in HD3 cells reduces the expression levels of the α D gene in mutant differentiated HD3 cells. CTCF binding in α D gene promoter of differentiated HD3 cells is necessary for its interaction with 3'-enh and transcriptional activation of α D.

1. INTRODUCCIÓN

1.1 La función del genoma

En un organismo eucarionte todas las células poseen el mismo genoma y la expresión de éste; en otras palabras, la transcripción de sus genes y su posterior traducción para llevar a cabo la síntesis de proteínas, son necesarias para determinar el fenotipo y función de cada una de dichas células (Zhu *et al.* 2013). Sin embargo, si todas las células tienen la misma información genética, ¿cómo es que un organismo está compuesto por cientos de distintos tipos celulares? La respuesta a esta pregunta se encuentra en el conjunto de genes que se expresan en cada una de éstas células. Por lo tanto, todas las células tienen un genoma idéntico pero un transcriptoma distinto para cada tipo celular. En los primeros estados de desarrollo de un organismo sus células se encuentran en un estado no diferenciado y son capaces de dar origen a la mayoría de los tipos celulares, a ésta característica se le conoce como pluripotencia (Martin 1981). El establecimiento de un transcriptoma determinado ocurre gracias a una regulación coordinada de la expresión génica espacio-temporal durante el desarrollo del organismo y, una vez establecido un linaje celular, el perfil de expresión debe mantenerse el resto de la vida de la célula terminalmente diferenciada y debe también de transmitirse a su descendencia. Pero, ¿cuáles son los mecanismos que permiten una expresión diferencialmente regulada del genoma tanto en tiempo (desarrollo) como en espacio (tipo celular)?

1.2 El genoma está organizado en cromatina

Para tratar de abordar esta pregunta debemos recordar que el genoma de un eucarionte se encuentra duplicado dentro del núcleo celular. En humanos una sola copia de los 23 cromosomas que forman parte del genoma está compuesta por 3 mil millones de nucleótidos. Dado que ésta información se encuentra duplicada, en el núcleo de una célula humana hay 6 mil millones de bases. Como consecuencia de lo anterior, si se pudieran extender de manera lineal el genoma completo, éste tendría una longitud de aproximadamente dos metros. Por lo tanto, para poder contener una molécula tan grande en un espacio tan reducido como el núcleo celular, el ADN forma un complejo muy

ordenado alrededor de proteínas llamadas histonas, que a su vez forman a los nucleosomas, y que en su conjunto se conoce como cromatina (Annunziato, 2008).

La cromatina está estructurada a distintos niveles (Figura 1): el nivel primario está representado por la doble hélice de ADN; el siguiente nivel está formado por los nucleosomas, que son octámeros de histonas rodeados por ADN; la interacción de un nucleosoma con una histona llamada H1 se llama cromatosoma; el siguiente nivel de compactación del genoma está dado por la fibra de 30 nm, también conocida como solenoide; el cual a su vez forma asas de cromatina. El mayor grado de compactación de la cromatina ocurre cuando los cromosomas se arreglan en cromátidas durante la metafase (Felsenfeld & Groudine, 2003).

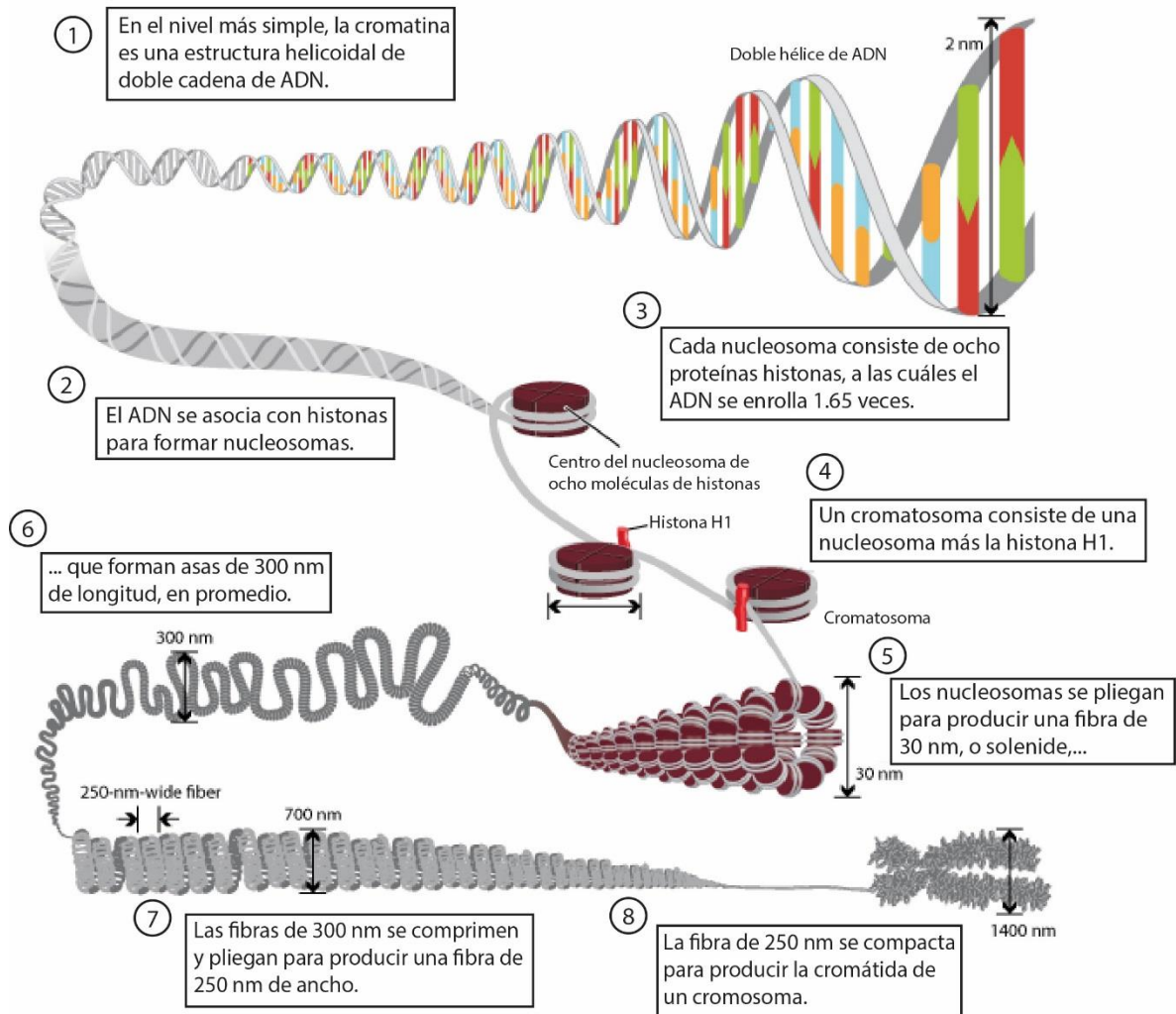


Figura 1. Niveles de organización de la cromatina (Modificada de Annunziato, A., 2008).

Además, la función del genoma está directamente relacionada con su organización tridimensional al interior del núcleo celular (Ong & Corces, 2014). Aunque el primer nivel de regulación de la expresión génica es dirigido por la unión de factores de transcripción en secuencias promotoras; existen modificaciones covalentes-heredables sobre el ADN y en particular sobre las histonas que contribuyen a la formación de dominios de transcripción activa o inactiva (Smith & Meissner, 2013). El establecimiento de dominios funcionales de expresión génica participa en la generación de patrones de expresión tipo celular-específicos en los eucariontes (van Bortle & Corces, 2012). Este aspecto es uno de los temas de estudio centrales en nuestro grupo de investigación y en esta tesis estudiamos la importancia de un elemento estructural de la cromatina en la activación de los genes α -globina de pollo en una línea celular de eritroblastos transformados.

1.3 Papel de la estructura y organización de la cromatina en la expresión diferencial

El establecimiento de dominios específicos de expresión o represión en el genoma se logra a través de: 1) la modificación química del ADN y las histonas que conforman los nucleosomas en una región genómica determinada; 2) la incorporación de variantes de histonas; 3) el reclutamiento de factores de transcripción; 4) la influencia de elementos de regulación distales como *enhancers*, silenciadores e *insulators*; y 5) la organización tridimensional del genoma al interior del núcleo (Figura 2).

La apertura o cierre de la estructura de la cromatina es uno de los principales factores que influye sobre el control de la expresión de los genes. Conceptualmente se conoce como eucromatina a aquellas regiones de cromatina abierta o laxa donde se presenta actividad transcripcional; mientras que a las regiones con una estructura compacta de cromatina, a las que no tienen acceso los factores de transcripción, se les llama heterocromatina. A continuación describo el papel de algunas modificaciones químicas sobre la fibra de cromatina en la formación de regiones de cromatina abierta o compacta. También menciono algunos aspectos relevantes de los elementos distales de regulación y la organización del genoma en la activación o silenciamiento de genes.

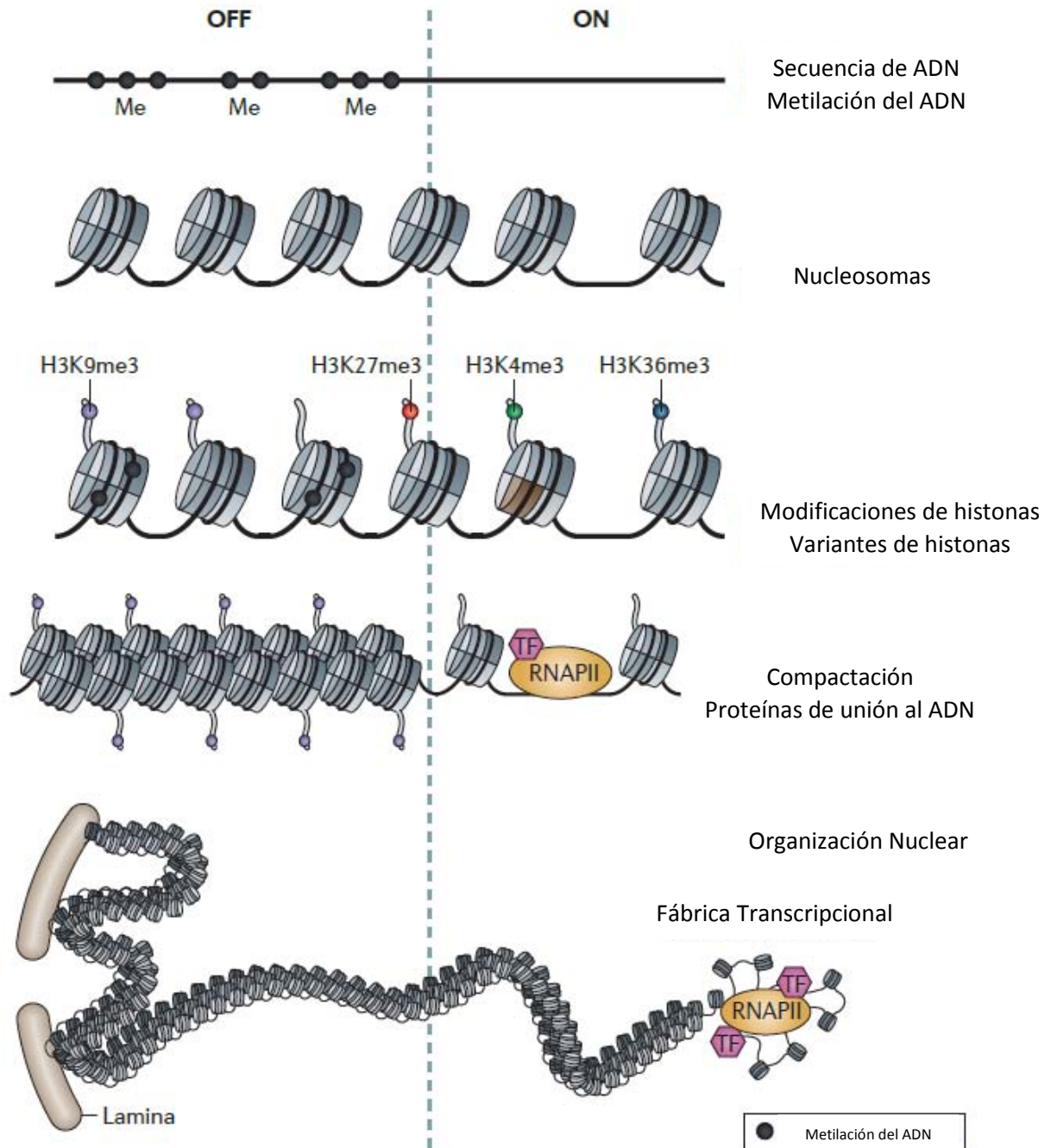


Figura 2. La organización y modificación de la cromatina influye sobre la expresión del genoma. Existen secuencias específicas en el genoma que pueden ser reconocidas por distintos factores que regulan la expresión de los genes. Sin embargo, la unión de éstos también depende de la combinación de modificaciones químicas que se presentan sobre la secuencia de ADN o las histonas que se presentan en una región determinada. Dichas modificaciones son reconocidas por cofactores y complejos remodeladores de la cromatina, regulando el grado de compactación del genoma; lo cual influye sobre la capacidad de unión al ADN de los factores de transcripción que finalmente van a participar en la expresión regulada de los genes. Existe además una localización nuclear determinada para las regiones de la cromatina que se transcriben activamente (eucromatina) y aquellas que están silenciadas (heterocromatina) (Tomada de Zhou *et al.*, 2011).

1.3.1 Metilación del ADN

La expresión de los genes α -globina de pollo que son objeto de estudio en este proyecto es influida por los estados de metilación del ADN, por ello describo ahora algunas de las características generales de la metilación del ADN en eucariontes. La incorporación de un grupo metilo en el carbono 5 de una base de citosina (5mC) es la modificación química más estudiada y tiene un papel importante en la regulación de la expresión génica de la mayoría de los modelos de plantas, animales y hongos; siendo asociada con el silenciamiento a través de fomentar una cromatina compacta. En mamíferos la metilación ocurre de manera simétrica en las dos citosinas de dinucleótidos CpG, los cuales se encuentra sobre-representados en el genoma (~28 millones de CpGs en el genoma humano) (Smith & Meissner, 2013). El hecho que la metilación ocurra en las citosinas palíndromes de los dinucleótidos CpG permite un mecanismo sencillo de transmisión de la marca a la siguiente generación de células; lo anterior a través de la propagación semiconservativa del patrón de metilación durante la duplicación del genoma (Allis *et al.*, 2015).

Existen principalmente dos formas en las que la metilación del ADN influye sobre las regiones promotoras. Una de ellas es a través del mecanismo directo, al metilarse los dinucleótidos CpG que coinciden con sitios de unión a factores de transcripción. Al estar metiladas dichas CpGs, los factores de transcripción no pueden reconocer su secuencia específica de unión en el promotor, comprometiéndose con ello la transcripción del gen al que regula (Della Ragione *et al.*, 2016). No sólo la unión de factores de transcripción puede ser afectada por la metilación; también la unión de proteínas que organizan la estructura tridimensional de la cromatina. Por ejemplo, se ha observado que la unión de la proteína multifactorial CTCF, que es objeto de estudio en esta tesis y cuya función describiré detalladamente más adelante, es metilo-sensible en algunos de sus sitios de unión en el genoma de vertebrados (Maurano *et al.*, 2015; Flavahan *et al.*, 2016; Liu *et al.*, 2016).

En el mecanismo indirecto, las CpG metiladas son blanco de unión de proteínas que reconocen de manera específica al ADN metilado, las cuales atraen co-represores y remodeladores de la cromatina que fomentan una cromatina compacta que no permite la

expresión de los genes (Della Ragione *et al.*, 2016). A estas proteínas se les conoce como *methyl binding proteins* (MBDs) por su dominio de unión al ADN metilado; un ejemplo es la proteína de unión a CpG metilado 2 (MeCP2). MeCP2 promueve la compactación de la cromatina en las regiones a las que se une al reclutar metiltransferasas de histonas y/o desacetilasas de histonas (Della Ragione *et al.*, 2016).

Como se ha mencionado, la metilación del ADN es importante para mantener genes y ciertas regiones del genoma silenciadas en un determinado tipo celular. Sin embargo, si dicho tipo celular se diferencia a otro tipo celular en el cual se requiere de la activación de algunos genes metilados, ocurre un proceso de desmetilación de las CpGs para descompactar esa región de la cromatina. La desmetilación activa del ADN, recientemente descrita, es catalizada por las enzimas *Ten Eleven Translocation* (TET), las cuáles oxidan de manera sucesiva a la 5mC hacia 5hmC (5-hidroximetilcitosina), 5fC (5-formilcitosina) y 5caC (5-carboxilcitosina) (Rasmussen & Helin 2016). Posteriormente estas bases modificadas son reconocidas y escindidas por una timina-ADN-glicosilasa (TDG); lo cual activa el sistema de reparación del ADN por escisión de base (BER) y restaura una citosina no metilada en esa posición (Rasmussen & Helin, 2016). Esto permite el reconocimiento y unión de los remodeladores de la cromatina y factores que relajan la estructura de la cromatina y promueven la actividad transcripcional.

En conclusión, la metilación de las citosinas en los dinucleótidos CpGs de regiones promotoras puede tener un efecto directo sobre la unión de factores de transcripción y/o favorecer el establecimiento de una cromatina compacta a través del reclutamiento de cofactores. De esta forma la metilación del ADN influye sobre la estructura de la cromatina y, junto con ello, en la regulación de la actividad transcripcional. A continuación se describe la importancia de la estructura de la cromatina en la activación o silenciamiento de genes.

1.3.2 Modificaciones post-traduccionales en las histonas

Se sabe que las proteínas que se unen al ADN, especialmente las histonas, tienen un efecto sobre el grado de compactación de la cromatina y en la regulación de la expresión génica. Sin embargo, el mecanismo por el cual las histonas tienen un efecto sobre la expresión diferencial de genes sigue siendo un campo de estudio en pleno desarrollo. Por muchos años se pensó que las histonas, al interactuar con el ADN, formaban una estructura compacta incompatible con la expresión de los genes. Con base en esta teoría, para activar un gen había que retirar completamente las histonas del promotor y del cuerpo del gen (Allis *et al.*, 2015). La validez de dicha teoría se puso en duda al no encontrar evidencia de que existieran regiones extensas de ADN desnudo en las células eucariotes (Clark & Felsenfeld, 1971), lo cual indicaría que hay regiones transcripcionalmente activas.

Fueron dos los descubrimientos que permitieron dilucidar el mecanismo por el cual las histonas influyen sobre la expresión génica (Allis *et al.*, 2015). El primero fue la correlación entre la acetilación de las histonas y la activación de los genes (Allfrey & Mirsky, 1964) y el segundo el descubrimiento del nucleosoma como unidad básica de la cromatina (Kornberg & Thomas, 1974). Los nucleosomas están compuestos por segmentos de ~147 pb de ADN que rodean un octámero de las histonas canónicas H2A, H2B, H3 y H4; aunque existen excepciones en los que alguna de las histonas canónicas es reemplazada por una variante de histona para conferir nuevas funciones al nucleosoma. Con la estructura cristalográfica del octámero de histonas conformando el nucleosoma (Luger *et al.*, 1997), se hizo evidente cómo las regiones amino-terminales de las histonas pueden sufrir modificaciones químicas con consecuencias estructurales y regulatorias. Lo anterior generó una gran cantidad de estudios para relacionar cada tipo de modificación química con estados específicos de estructuración de la cromatina y de actividad transcripcional. Todo esto llevó a que los investigadores de ésta área propusieran la existencia de un código de histonas con un papel preponderante en el fenotipo celular a través del reclutamiento de proteínas efectoras (Strahl & Allis, 2000; Allis *et al.*, 2015).

En este contexto es importante tener en mente que los nucleosomas están posicionados en todo el genoma, no sólo en regiones promotoras y cuerpos del gen. Con esto quiero decir que los nucleosomas también tienen un papel importante en regular el nivel de compactación y actividad de elementos regulatorios de la transcripción que se encuentran alejados de los genes. Con relación a esto, se ha observado por experimentos de inmunoprecipitación de la cromatina con anticuerpos específicos que ciertas modificaciones post-traduccionales, principalmente en los residuos de lisina de la histona 3, están relacionados con el estado activo o inactivo de los elementos de regulación. Por ejemplo, la trimetilación de la lisina 4 en la región amino-terminal de la histona 3 (H3K4me3), es una marca que se ha visto generalmente asociada a promotores de genes transcripcionalmente activos. Mientras que la trimetilación de la lisina 27 en la histona 3 (H3K27me3) es una modificación que se encuentra en regiones de heterocromatina; es decir, dónde no se espera actividad transcripcional.

A nivel nucleosomal, la activación de un gen, o grupo de genes, depende de las modificaciones químicas en las regiones amino-terminales de las histonas y del reposicionamiento de los nucleosomas mediado por la acción de complejos remodeladores de la cromatina cuya actividad requiere de ATP. Cabe señalar que el reposicionamiento de nucleosomas durante la transcripción ocurre de una forma coordinada y, una vez que se ha generado el transcrito, los nucleosomas vuelven a su posición original.

Las variantes y modificaciones de histonas juegan entonces un papel crucial en la apertura y cierre de la fibra de cromatina, modulando con ello la transcripción o silenciamiento de genes durante el ciclo celular, el desarrollo y la diferenciación (Figura 2). Gracias a los ensayos de inmunoprecipitación de la cromatina y secuenciación masiva (ChIP-seq); es decir, a escala de genoma completo, se ha observado que generalmente las marcas y variantes de histonas se posicionan en regiones de regulación génica específicas. Lo anterior favorece la hipótesis del código de histonas y facilita la identificación de promotores, sitios de inicio de la transcripción (*TSS*), y de elementos distales de regulación como *enhancers*, silenciadores e *insulators*; de los cuáles hablaré a continuación.

1.3.3 Elementos distales de regulación: *Enhancers*, Silenciadores e *Insulators*

Los elementos distales de regulación se asocian con secuencias en el genoma que influyen sobre la expresión de un gen o grupo de genes. Gracias a ciertas características en su composición genómica, la unión de factores de transcripción, la presencia/ausencia de nucleosomas y las modificaciones químicas en las regiones amino-terminal de las histonas es posible inferir si una región en el genoma puede tener función de *enhancer*, silenciador o *insulator* (Ernst & Kellis, 2012). Sin embargo, en el pasado la función de los elementos de regulación se definió a partir de su efecto sobre genes reporteros al incorporarlos en construcciones transgénicas; lo cual, a pesar de ser útil en la caracterización de la funcionalidad de la secuencia, no aportaba la suficiente información para entender su importancia en el contexto natural/ *in vivo* de la región de cromatina en la que se encuentran. A continuación resumo algunas de sus características.

Enhancers

Los *enhancers* se definen como regiones del genoma que tienen la capacidad de aumentar la transcripción de un gen independientemente de su posición y orientación respecto al promotor del gen que regula. Los *enhancers* pueden encontrarse a miles de pares de bases de sitio de inicio de la transcripción de su gen blanco. La distancia a la que se encuentran se sus promotores blanco es una de las principales limitantes para identificarlos, ya que dicha distancia es tan variable que de hecho pueden encontrarse hasta en las regiones intrónicas del gen al que regulan (Pennacchio *et al.*, 2013).

No existe una secuencia consenso que permita identificarlos por métodos computacionales pero, por otra parte, tienden a ser secuencias no-codificantes altamente conservadas entre distintas especies del mismo grupo filogenético (Pennacchio *et al.*, 2006). Gracias a los datos de ChIP-seq de distintas marcas epigenéticas y modificaciones de histonas se puede facilitar su identificación. Las marcas más utilizadas para identificar posibles *enhancers* son la presencia de la acetilasa de histonas p300 y las modificaciones H3K27ac y H3K4me1 (Heintzman *et al.*, 2007; Visel *et al.*, 2009; Creighton *et al.*, 2010). Otra de sus características es que se encuentran en regiones de cromatina abierta por lo

que los ensayos de hipersensibilidad a la DNasa I también resultan útiles para su identificación (Dorschner *et al.*, 2004). La validación final se realiza con pruebas *in vitro* (construcciones transgénicas) e *in vivo* (mutaciones en la secuencia del *enhancer*).

El modelo de acción de los *enhancers* más aceptado es el modelo de asas de cromatina. En este modelo se propone que los *enhancers* y promotores, que se encuentran lejanos a nivel de secuencia genómica, están cercanos físicamente al interior del núcleo gracias a la formación de asas de cromatina (Maston *et al.*, 2006). Las secuencias *enhancer* reclutan a factores generales de la transcripción y a la ARN Polimerasa II; lo cual indica que pueden estar fungiendo como centros para el ensamblaje de complejo pre-iniciador (PIC) de la transcripción. La Dra. Ann Dean considera que el acercamiento de los *enhancers* con sus promotores blanco podría: 1) ayudar a incrementar la concentración local de componentes de la maquinaria de transcripción en las regiones aledañas al gen blanco o, 2) un *enhancer* podría servir para transferir el PIC a un promotor (Pennacchio *et al.*, 2013). Lo anterior produciría entonces un incremento en la actividad transcripcional del gen blanco.

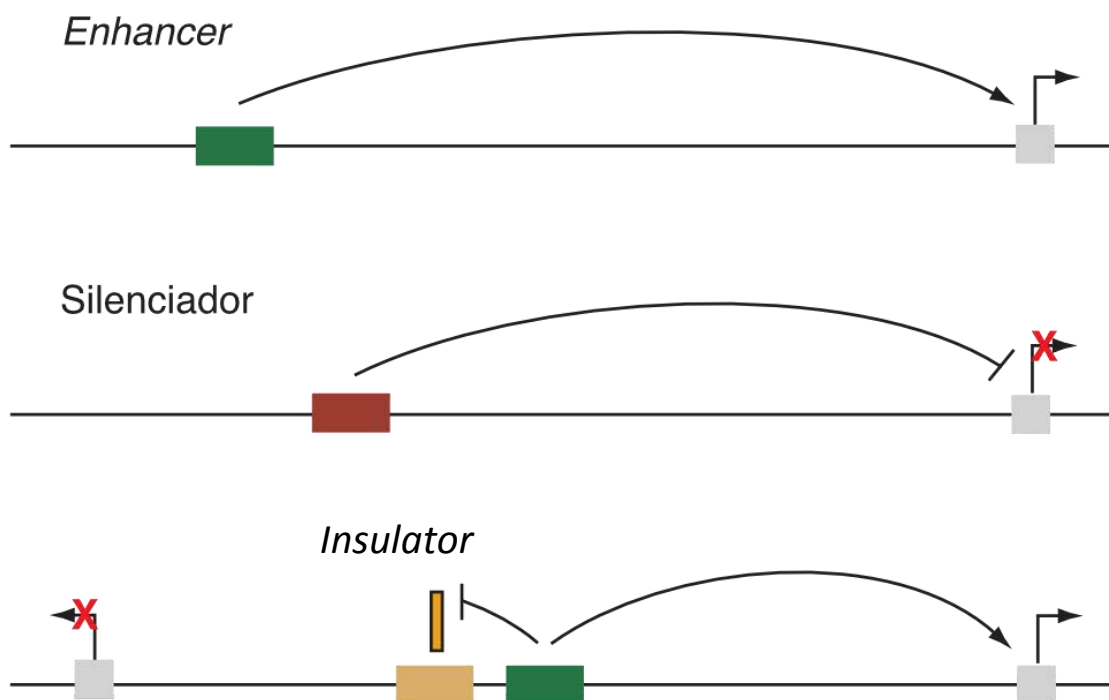


Figura 3. Elementos distales de regulación de la expresión génica. Los *enhancers* (cuadro verde) son secuencias que tienen el potencial de activar y/o aumentar la expresión de genes que se encuentran a miles de kilobases de distancia; mientras que los silenciadores (cuadro rojo) tienen el efecto opuesto, favorecen el silenciamiento de genes. Los *insulators* (cuadro amarillo) son regiones que limitan la actividad de los *enhancers* hacia un gen o genes específicos, además de que delimitan regiones de eucromatina y heterocromatina (Tomada de Maston *et al.*, 2006).

Silenciadores

Los silenciadores son secuencias reconocidas por factores represores y co-represores de la transcripción. De manera similar a los *enhancers*, su efecto sobre los promotores que regulan es independiente de su orientación y distancia. Con relación a la distancia, los silenciadores pueden: 1) encontrarse aledaños al promotor o al *enhancer* que regulan, o 2) presentarse como secuencias independientes en regiones intergénicas, intrónicas o 3'-UTR (Maston *et al.*, 2006). Uno de los modelos de acción de los silenciadores propuesto involucra la unión de los factores represores al silenciador. La unión de los represores pueden prevenir la unión de otros factores de transcripción activadores, ya sea porque los sitios de unión se encuentran cercanos entre sí, o al competir directamente por un mismo sitio de unión (Maston *et al.*, 2006). Los represores además pueden reclutar remodeladores de la cromatina que promueven la compactación y silenciamiento de la región blanco del silenciador (Maston *et al.*, 2006). Un ejemplo de secuencias que pueden tener un papel de silenciadores son los denominados *Polycomb Response Elements* (PREs) que se encuentran en el genoma de *D. melanogaster*. Se ha observado que la unión del factor Ying Yang-1 (YY1) a los PREs en embriones de *D. melanogaster* recluta a proteínas represoras del grupo Polycomb (PcG) a los sitios de unión. Lo anterior causa la desacetilación de histonas, la metilación de la lisina 9 y lisina 27 de la histona 3 y una cromatina compacta en las regiones blanco (Srinivasan & Atchison, 2004).

Insulators

En una célula diferenciada se pueden identificar puntualmente dominios alternos de heterocromatina y eucromatina, los cuales se definen durante el proceso de toma de decisión de la diferenciación celular (Hawkins *et al.*, 2010; Barkess & West, 2012; Zhu *et al.*, 2013). Dado lo anterior, es necesario definir límites para aislar las actividades opuestas de dominios de cromatina contiguos (Ali *et al.*, 2016). Las preguntas fundamentales que surgen sobre los dominios de cromatina están relacionadas con su formación y mantenimiento; así como con el establecimiento de sus límites. Sobre esto último se han identificado secuencias específicas en la mayoría de los límites de los dominios de cromatina que evitan

la propagación de heterocromatina hacia los genes activos (Barkess & West, 2012), razón por la cual se les conoce como *insulators*.

Barkess & West (2012) definen puntualmente a los *insulators* como elementos de ADN que tienen la habilidad de proteger a un gen de la influencia de elementos regulatorios no deseados presentes en su ambiente cromosomal. Por varias décadas se han utilizado ensayos de actividad reportera con construcciones transgénicas para caracterizar la función *insulator* de determinadas secuencias de nucleótidos (Figura 4). Con las construcciones transgénicas se pueden estudiar dos propiedades de los *insulators*. La primera es la actividad como barrera a la propagación de la heterocromatina (Figura 4A) (Recillas-Targa *et al.*, 2002). La segunda es la actividad de bloqueo de *enhancer* (Figura 4B) (Recillas-Targa *et al.*, 1999); es decir, la capacidad de una secuencia de evitar que un *enhancer* ejerza una función sobre un promotor determinado.

Sin embargo, aunque los ensayos reporteros permiten identificar fácilmente secuencias con actividad *insulator* (de acuerdo a su definición funcional), una limitante es que en las construcciones transgénicas no se respeta el contexto genómico y cromatínico del *locus* en particular sobre el cuál actúan (Barkess & West, 2012). Por ésta razón permanecen las preguntas sobre de qué manera actúan los *insulators* en el contexto *in vivo* regulando la actividad de *enhancers* y promotores específicos, y cuáles factores les permiten tener su función de barrera.

Gracias a los experimentos de captura conformacional de cromosomas, los cuáles permiten estudiar las interacciones físicas de regiones genómicas mediadas por proteínas, se ha observado que su actividad la llevan a cabo a través de la formación de asas de cromatina (Ong & Corces, 2014). Dado que es poco probable recuperar dichas interacciones con las construcciones transgénicas, se vuelve importante estudiar los mecanismos funcionales de los *insulator* dentro del contexto cromatínico en el que residen. Hasta hace algunos años era muy difícil estudiar *in vivo* el papel de una región *insulator* en la estructuración de la cromatina; sin embargo, gracias al desarrollo de técnicas de edición genómica se han observado tanto la función de barrera (Narendra *et al.*, 2015), como de

bloqueo de *enhancer* de elementos *insulator* específicos en *loci* particulares (Lupiáñez *et al.*, 2015).

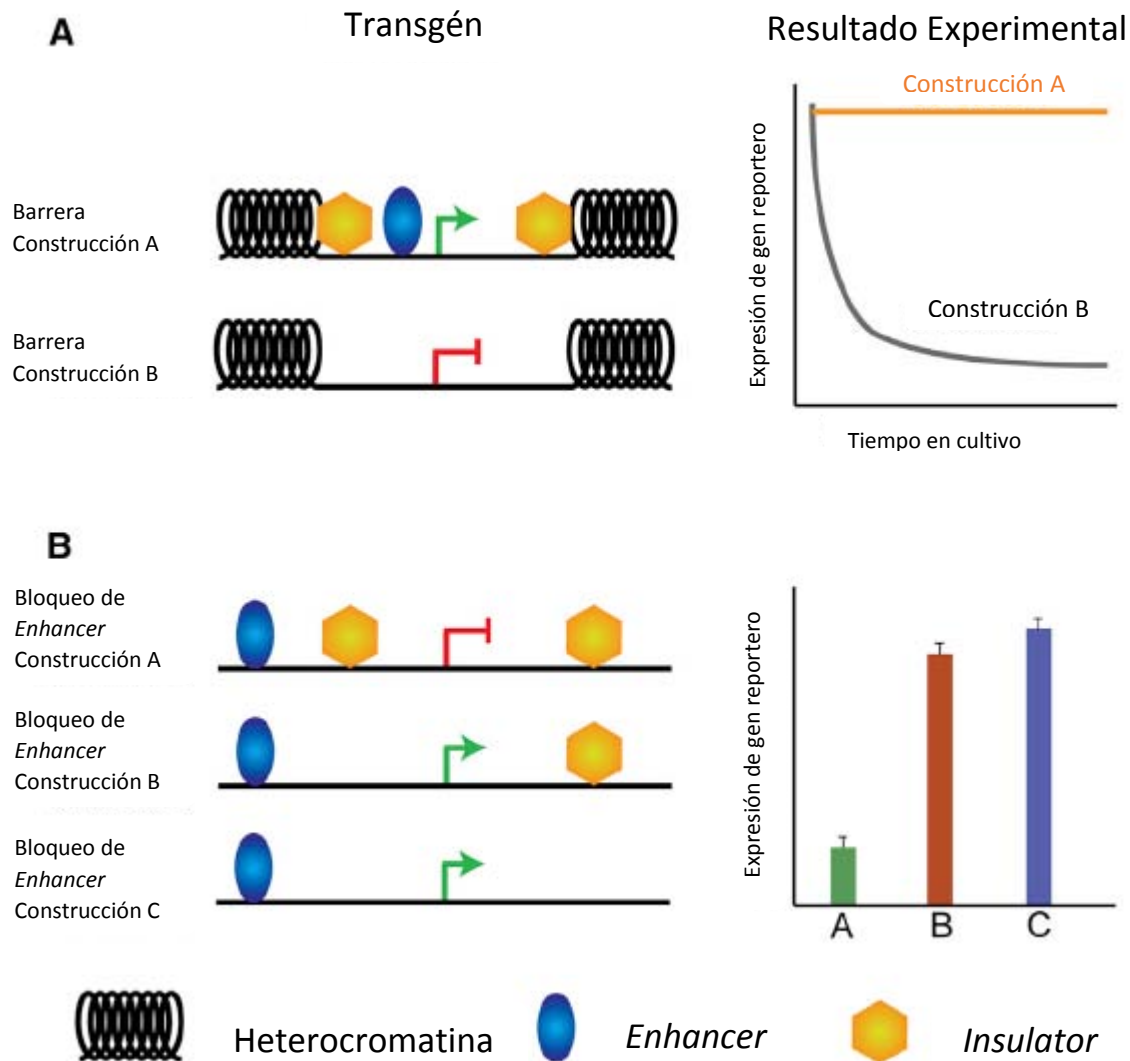


Figura 4. Caracterización de secuencias *insulator* utilizando construcciones transgénicas. A) Las secuencias *insulator* con actividad de barrera pueden proteger a un gen reportero de la heterocromatinización cuando ésta construcción se inserta en el genoma de una célula transfectada, de manera que la expresión del reportero se mantiene a través del tiempo (Construcción A). En cambio, la expresión del gen reportero en una construcción transgénica sin *insulator* va disminuyendo progresivamente después de mantener las células transfectadas en cultivo por cierto tiempo (Construcción B). **B)** La actividad de bloqueo de *enhancer* de un *insulator* se puede observar al colocar dicha secuencia entre la de un *enhancer* y un gen reportero. Si el *insulator* impide la actividad del *enhancer* sobre el gen reportero habrá una menor expresión del gen reportero (Construcción A) respecto a las construcciones control en las que no se encuentra el *insulator* entre el *enhancer* y dicho gen (Construcciones B y C) (Tomada de Phillips-Cremins & Corces, 2013).

Con respecto al mecanismo funcional de los *insulator*, éstos tienen sitios para la unión de factores de transcripción que contribuyen en mantener un estado abierto de la cromatina al reclutar modificadores de histonas e impedir la metilación del ADN (Raab & Kamakaka, 2010). Los ejemplos característicos están dados por los factores de transcripción USF1/USF2 y VEZF1. La unión de USF1/USF2 en el *insulator* CHS4 del dominio β -globina de pollo es necesaria para reclutar remodeladores de la cromatina que establezcan marcas de histonas de cromatina abierta que impiden la propagación de heterocromatina hacia el locus de los genes β -globina (West *et al.*, 2004). Mientras tanto, por ensayos de genes reporteros se comprobó que el sitio de unión a VEZF1 en el *insulator* CHS4 es necesario para evitar la metilación del ADN (Dickson *et al.*, 2010).

Los *insulator* tienen también sitios de unión de proteínas arquitectónicas, para el establecimiento de las interacciones de cromatina (Van Bortle *et al.*, 2014). En *Drosophila melanogaster* se han descrito más de 10 proteínas arquitecturales; mientras que en vertebrados, aunque se han encontrado varias proteínas colocalizando en las regiones de interacción de largo alcance, sólo una (CTCF) se ha caracterizado ampliamente en regiones *insulator* (Barkess & West, 2012; Gómez-Díaz & Corces, 2014). En el presente trabajo se muestran resultados que evidencian la importancia de la presencia de CTCF en la expresión de genes que se activan durante la diferenciación terminal de los eritrocitos de pollo. A continuación se describen algunas de las características más relevantes del papel de CTCF en la estructuración de la cromatina.

1.4 El factor nuclear CTCF y la organización tridimensional de la cromatina al interior del núcleo

El factor de unión a CCCTC (CTCF, por sus siglas en inglés) es una proteína nuclear de 11 dedos de zinc que en sus orígenes fue identificada como un factor de transcripción en los promotores MYC de aves y mamíferos (Lobanenkov *et al.* 1990; Klenova *et al.* 1993) y posteriormente descrito como una proteína con actividad *insulator* (Bell *et al.*, 1999). Sin embargo, recientemente se ha descubierto también que CTCF es la principal proteína que participa en la formación de asas de cromatina en mamíferos (Phillips-Cremins & Corces 2013; Rao *et al.* 2014).

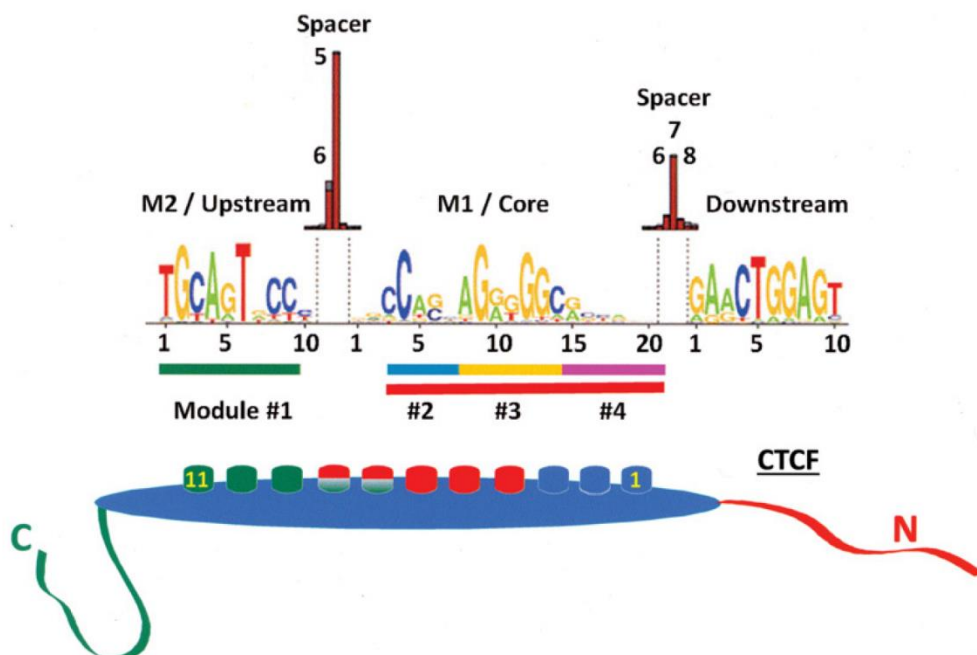


Figura 5. CTCF es un factor multivalente. A través de experimentos de ChIP-seq se ha podido determinar una secuencia de 20 nucleótidos (M1/Core) a la cual se unen preferencialmente los dedos de zinc 4-8 de CTCF. La secuencia M2 se encuentra presente también en algunos de los sitios de unión a CTCF con mayor afinidad para proveer estabilidad durante la unión. Mientras que la secuencia *downstream* es probable que tenga un papel relevante en la separación de CTCF a su sitio de unión. Existen además secuencias espaciadoras enmarcando el motivo *core* que le proveen estabilidad a la unión de CTCF a sus secuencias consenso. La orientación con la que se une CTCF al sitio M1 mostrado corresponde a la región N-terminal apuntando hacia la posición *downstream* de la secuencia *core* (Ghirlando & Felsenfeld 2016).

CTCF se une a secuencias de ADN de ~50 pb y se le ha llamado factor multivalente por su capacidad de reconocer múltiples sitios blanco en el genoma utilizando combinaciones distintas de sus dedos de zinc individuales (Filippova, 2008). No obstante, por análisis computacionales, se ha identificado que alrededor de 20 nucleótidos (*binding motif*) son críticos para el reconocimiento por CTCF (Figura 5) (Kim *et al.* 2007; Cuddapah *et*

al., 2009). Por experimentos *in vitro* e *in vivo* se ha demostrado que al mutar los nucleótidos que se encuentran en el *binding motif* se reduce o pierde totalmente la unión de CTCF en sondas marcadas radiactivamente y en el genoma, respectivamente; lo cual tiene repercusiones a nivel de organización de la cromatina, transcripción de genes y fisiológicamente (Bell *et al.*, 1999; Valadez-Graham *et al.*, 2004; Narendra *et al.*, 2015; Flavahan *et al.*, 2016).

CTCF lleva a cabo su función como *insulator* a través de la formación de asas de cromatina. El mecanismo de formación de asas de cromatina involucra la interacción de dos proteínas CTCF que reconocen sus sitios de unión en la base del asa de cromatina (Ghirlando & Felsenfeld, 2016). Este tipo de configuración mantiene regiones del genoma aisladas de su entorno cromatínico y de elementos regulatorios adyacentes. Las asas de cromatina pueden: 1) contener genes que están co-expresados o en un estado silenciado, 2) separar genes físicamente de *enhancers* y así evitar su función activadora, o 3) agrupar *enhancers* junto con genes al interior del asa para que su actividad sea específica y favorecer su localización en espacios determinados al interior del núcleo (Downen *et al.*, 2014).

La importancia de mantener regiones de la cromatina aisladas de elementos reguladores o cromatina adyacentes es tal que, se ha demostrado que mutaciones que impidan la unión de CTCF en los sitios frontera puede promover la activación anómala de genes de los *loci* HoxA y HoxC, por expansión de marcas de histonas de cromatina abierta (H3K4me3), en neuronas motoras diferenciadas a partir de células troncales embrionarias de ratón (Figura 6A) (Narendra *et al.*, 2015). Dichas mutaciones pueden también causar la activación de proto-oncogenes a través de la acción inespecífica de *enhancers* aledaños, causando el desarrollo de un proceso tumoral (Figura 6B) (Hnisz *et al.*, 2016). También ésta desestructuración de la cromatina se ha visto asociada con malformaciones de las extremidades en ratones y humanos (Figura 6C). La eliminación de la frontera entre los dominios de cromatina Elpha4 y Pax3 permite la interacción del *enhancer* interno mm1036 con Pax3, generando braquidactilia. Mientras que la eliminación de la frontera del dominio de cromatina Elpha4 con el dominio Wnt6/lhh hace lo propio para la interacción inespecífica del *enhancer* con el gen lhh, produciendo polidactilia (Lupiáñez *et al.*, 2015).

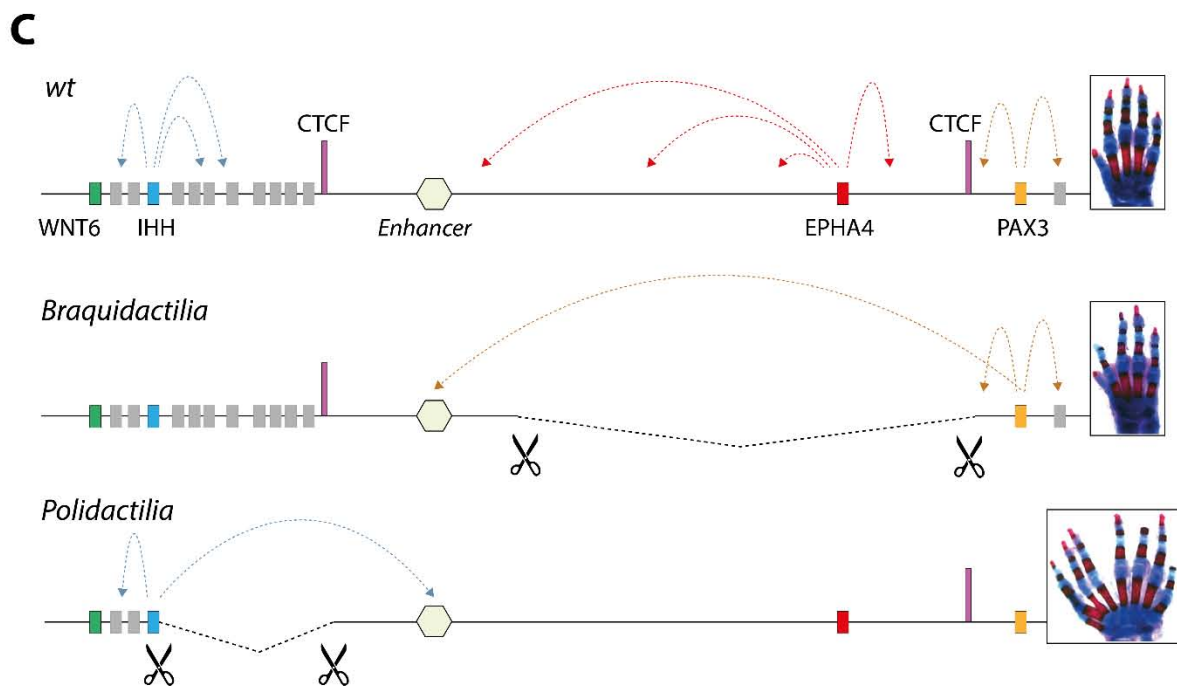
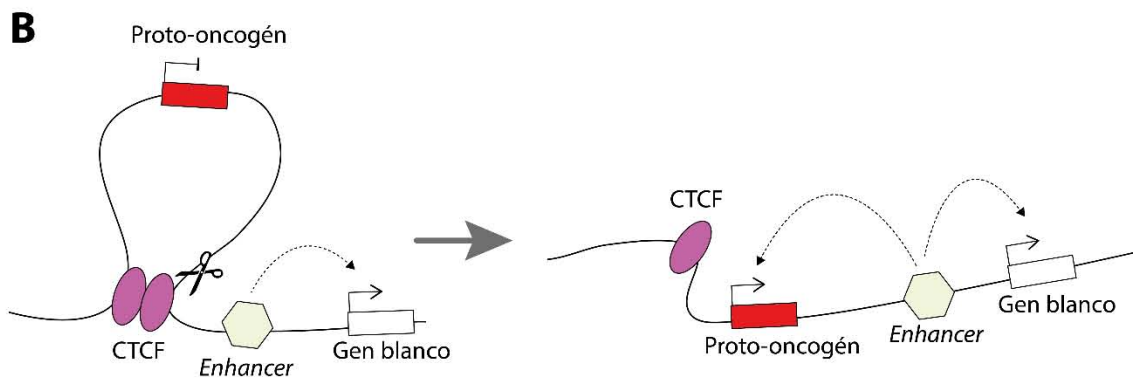
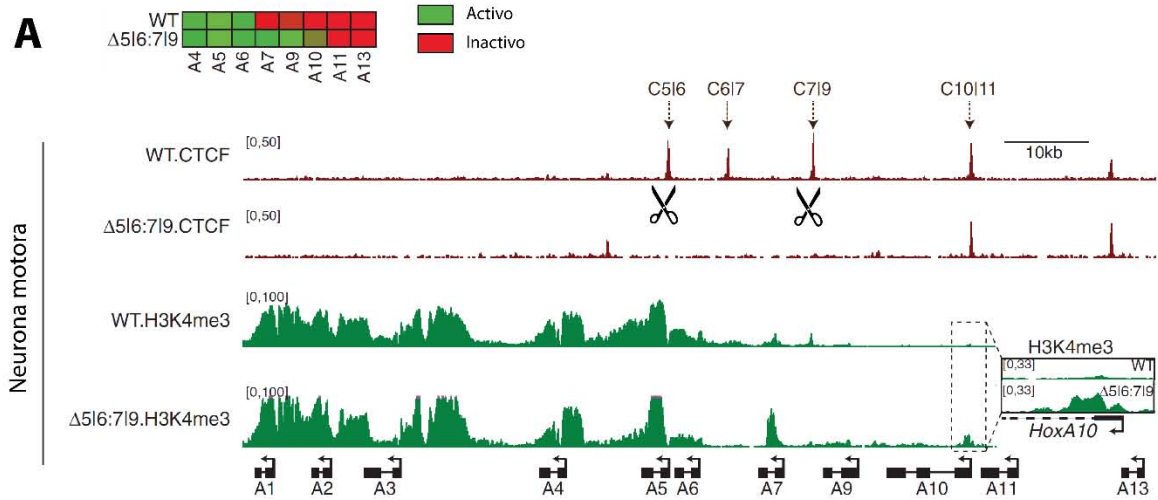


Figura 6. Mutaciones en los sitios de unión a CTCF producen re-arreglos en la estructura de la cromatina. Se muestran algunos ejemplos de experimentos en los que se comprueba *in vivo* el papel de CTCF en la organización de la cromatina, y la importancia de dicha organización en la actividad transcripcional específica de distintos *loci* de genes. **A)** Actividad transcripcional y enriquecimiento de CTCF y la marca H3K4me3 en el *locus* HoxA de neuronas motoras de ratón silvestres y mutantes. La gráfica de colores indica si el gen correspondiente está encendido (verde) o apagado (rojo). Se muestran los picos de enriquecimiento de unión de CTCF en el *locus* HoxA en neuronas motoras silvestres y mutantes, obtenidos por inmunoprecipitación de la cromatina con el anticuerpo para CTCF y secuenciación (ChIP-seq). Se muestra en verde la señal de enriquecimiento de la marca de cromatina H3K4me3 obtenida por ChIP-seq. La eliminación de los sitios de unión a CTCF C516 y C719 produce la propagación de la marca de cromatina abierta hacia la región downstream del locus, activando con ello los genes HoxA7 – HoxA10 (Tomado de Narendra *et al.*, 2015). **B)** Mecanismo propuesto por Hnisz *et al.* (2016) para la activación de proto-oncogenes por desorganización de la cromatina. La mutación de sitios de unión a CTCF que participan en el aislamiento de proto-oncogenes puede producir la interacción no-específica de un *enhancer* con el proto-oncogén, causando la activación anómala de este último. **C)** Modelo de interacciones de cromatina que se encuentran en los dominios de cromatina EPHA4, WNT6/IHH y PAX3 en las extremidades de embriones de ratones silvestres y con malformaciones. CTCF se une en los límites del dominio de cromatina EPHA4. Las flechas representan algunas interacciones de cromatina de los genes *Ihh*, *Epha4* y *Pax3* con zonas del dominio de cromatina en el que se encuentran (*wt*) y con una región *enhancer* del dominio EPHA4 (braquidactilia y polidactilia); las interacciones fueron recuperadas por experimentos de 4C (captura conformacional de cromosomas *on-chip*). Las líneas punteadas representan fragmentos del genoma eliminados con el sistema CRISPR-Cas9 y que representan algunas de las mutaciones que se pueden encontrar en humanos y ratones con braquidactilia y polidactilia de manera natural. Las eliminaciones que incluyen a los límites del dominio de cromatina EPHA4 producen un re-arreglo de la estructura tridimensional de la cromatina, permitiendo con ello la interacción de *enhancers* internos al dominio EPHA4 con genes de los dominios de cromatina contiguos (PAX3 y WNT6/IHH), lo cual ocasiona malformaciones en las extremidades de ratones y humanos (Modificado de Lupiáñez *et al.*, 2015). Las tijeras indican los sitios sobre los cuales se diseñaron los ARN guías para los cortes de doble cadena por la endonucleasa Cas9 del sistema de edición genómica CRISPR-Cas9.

Los ejemplos anteriores proveen evidencia del papel de CTCF en la organización tridimensional del genoma y en la separación física de dominios de cromatina. Los dominios de cromatina son unidades de cientos a miles de kilobases en las que los genes presentan cercanía espacial y, la mayoría de las veces, una misma regulación transcripcional (Dixon *et al.*, 2016). La unión de CTCF en los límites de los dominios de cromatina confiere estabilidad a la arquitectura genómica (Nora *et al.*, 2012; Phillips-Cremins & Corces, 2013; Lupiáñez *et al.*, 2015), regula el balance de modificaciones de cromatina activa y represiva (Splinter *et al.*, 2006; Narendra *et al.*, 2015), y permite la interacción específica de *enhancers* y promotores (Flavahan *et al.*, 2016; Hnisz *et al.*, 2016).

En resumen, CTCF es una proteína multifuncional porque está involucrada en distintos aspectos de la regulación de la expresión génica, de la estructuración de la cromatina y de la organización tridimensional del genoma. Por ello, estudiar sus mecanismos de acción permite obtener un mayor entendimiento de los procesos moleculares que establecen regiones de cromatina abierta o compacta, y que influyen sobre la activación o silenciamiento transcripcional de un gen durante la diferenciación celular. Esto es uno de los aspectos abordados en este trabajo.

Como observamos en los ejemplos descritos (Figura 6), la edición genómica de *loci* específicos que están asociados con estados de regulación de la transcripción y de la organización de la cromatina puede ayudar a discernir la relación entre la arquitectura del genoma y su función (Hsu *et al.*, 2014). En el presente trabajo eliminamos el sitio de unión a CTCF en el promotor del gen αD para estudiar su contribución en la activación de los genes α -globina de pollo.

1.5 El sistema de edición del genoma CRISPR-Cas9

En la búsqueda para asociar a cada gen una función específica en un determinado tipo celular, enfermedad o momento del desarrollo de un organismo se han generado distintas técnicas de ingeniería genómica como: a) recombinación homóloga CRE-loxP, para generar *knockouts* o *knockins* de un gen o región determinada; b) ARNs interferentes, para reducir la cantidad de un transcrito específico (*knockdowns*); c) ensayos de actividad reportera, para delimitar promotores, *enhancers*, silenciadores o *insulators*; d) *zinc finger* nucleasas y TALENs, para generar eliminaciones, inserciones y modificaciones.

Sin embargo, aunque la mayoría de dichas técnicas tienen ciertas ventajas para el estudio funcional de un gen, existen también ciertas desventajas. Por ejemplo, la frecuencia de recombinación homóloga es de 1 en $10^6 - 10^9$ células y la eficiencia de degradación de un transcrito por medio de un ARN interferente puede llegar a ser menor al 50%. Por otra parte, los ensayos de actividad reportera no ofrecen la capacidad de explorar la función de los elementos de regulación en un contexto genómico real y la edición genómica utilizando nucleasas que necesitan de la interacción proteína-ADN para un reconocimiento específico se puede comprometer dependiendo del contexto cromatínico de la región blanco (Hsu *et al.*, 2014).

En este sentido, el reconocimiento de una región genómica blanco por apareamiento de bases Watson-Crick, utilizando un oligonucleótido como guía para las endonucleasas, ofrece una solución a la mayoría de las limitantes mencionadas. Esto es ahora posible gracias al desarrollo del sistema de edición genómica CRISPR Cas9 por Charpentier y Doudna (2013). El sistema funciona así: la endonucleasa Cas9 es dirigida por un ARN guía (gRNA) al sitio blanco de interés a editar, generándose un corte de doble cadena que al ser reparado por el mecanismo de unión de extremos no homólogos, puede producir inserción o eliminación de una o varias bases en la secuencia de ADN (Cong *et al.*, 2013).

El acrónimo CRISPR hace referencia a los elementos repetidos llamados en inglés *Clustered Regularly Interspaced Short Palindromic Repeats* encontrados en el genoma de

una gran diversidad de bacterias y arqueas. Las regiones CRISPR en éstos organismos están compuestos por 1) conjuntos agrupados de genes CRISPR-asociados (Cas) codificantes para proteínas y 2) una serie de secuencias repetidas (repetidas directas) espaciadas por secuencias variables (*spacers*), estas últimas similares a las secuencias de distintos bacteriófagos (Hsu *et al.*, 2014).

El gRNA debe contener una secuencia de 20 nucleótidos que sea complementaria a la secuencia blanco. Una característica particular para el diseño de ésta secuencia es que en el extremo 3' de la región que será reconocida debe haber una secuencia de tres nucleótidos de la forma NGG (secuencia PAM), esto para el caso del sistema CRISPR-Cas9 derivado de *Streptococcus pyogenes* (Ran *et. al.*, 2013). Al unirse el gRNA a su secuencia complementaria, la endonucleasa Cas9 realiza el corte en las dos cadenas del ADN a 3 nucleótidos de distancia de la secuencia PAM. Lo cual es reconocido y reparado por el mecanismo de reparación del ADN.

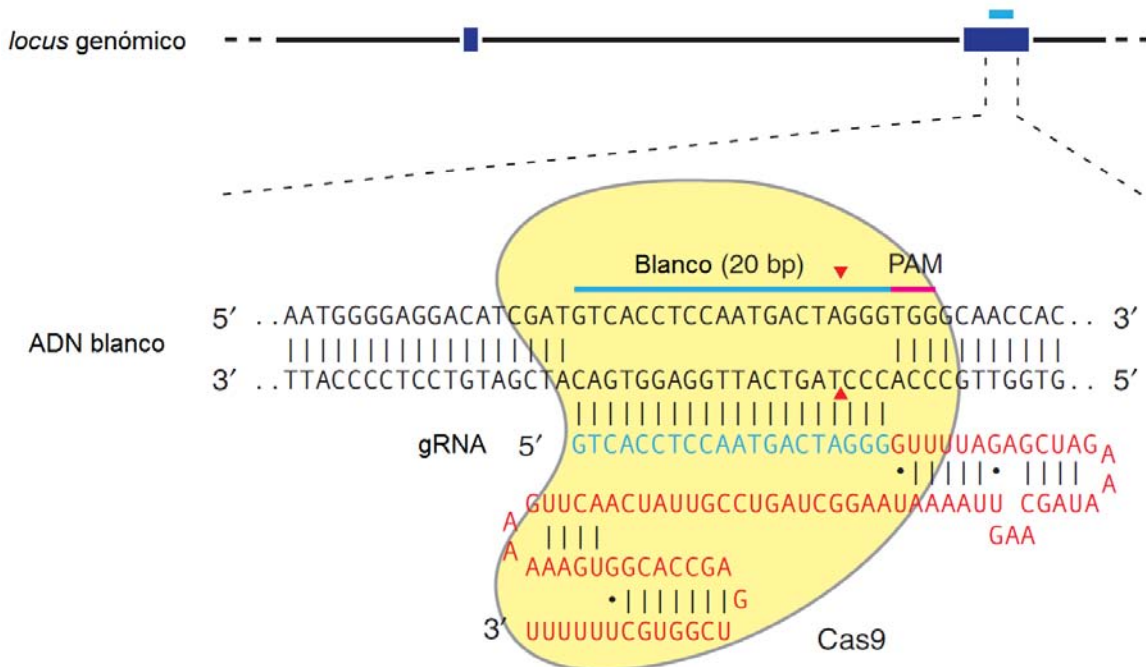


Figura 7. Esquema del sistema de edición genómica CRISPR-Cas9. La endonucleasa Cas9 es dirigida a una secuencia blanco específica en el genoma a través de su interacción con un gRNA que reconoce y complementa a dicho blanco. Para que la Cas9 realice el corte debe existir una secuencia PAM (Protospacer Adjacent Motif) inmediatamente después de los nucleótidos que forman apareamiento Watson-Crick con el gRNA. El corte de doble cadena por la Cas9 se realiza ~3 pb *upstream* a la secuencia PAM (triángulos rojos) (Tomada de Ran *et al.*, 2013).

Esta técnica ofrece un sinnúmero de posibilidades en el campo de la investigación científica, el sólo hecho de poder eliminar, insertar o modificar secuencias de ADN en células y organismos permite estudiar puntualmente la función de genes específicos y elementos regulatorios.

2. ANTECEDENTES

2.1 El dominio α -globina de pollo durante el desarrollo embrionario

En nuestro grupo de investigación se ha adoptado al dominio α -globina de pollo como modelo de estudio para entender los mecanismos asociados a la cromatina que regulan la expresión diferencial de los genes. El modelo es de gran utilidad dado que los genes que pertenecen al dominio se expresan coordinadamente en distintas etapas del desarrollo embrionario del pollo para regular los cambios en la necesidad de transporte de oxígeno. Además de que su configuración cromatínica cambia durante la diferenciación eritroide para permitir la acción de distintos elementos de regulación sobre los genes globina (Gavrilov & Razin, 2008; Valdes-Quezada *et al.*, 2013).

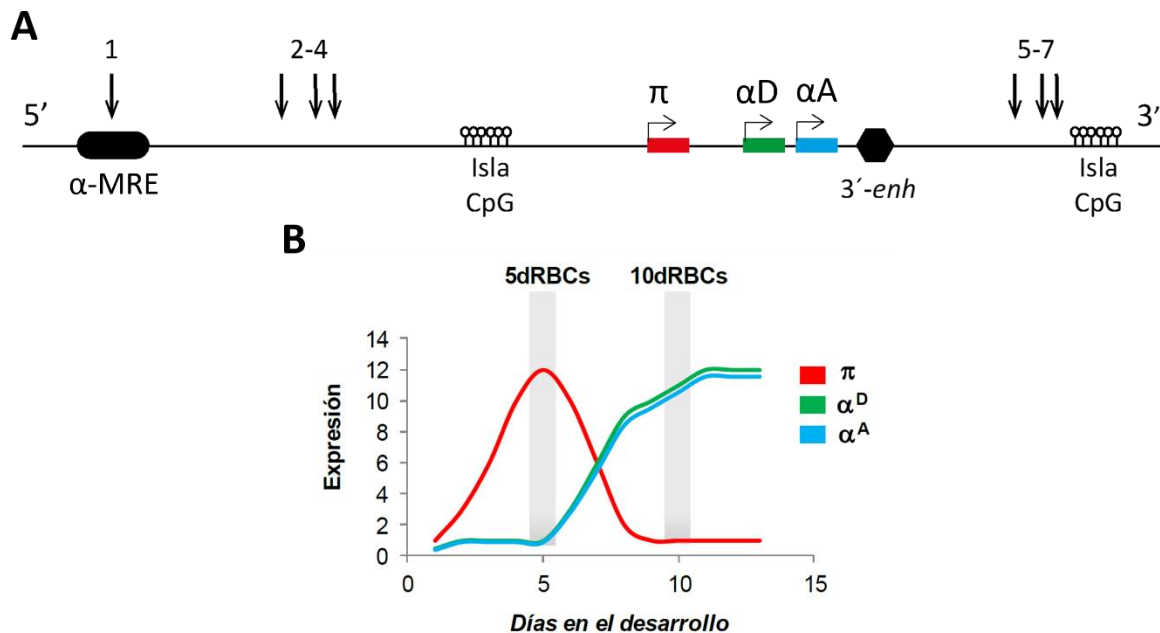


Figura 8. Los genes del dominio α -globina de pollo se expresan diferencialmente durante el desarrollo embrionario. A) Modelo del dominio α -globina de pollo en el que se muestran: los elementos potenciadores de la expresión α -MRE y $3'$ -enhancer, distintos sitios de hipersensibilidad a la DNasa, dos islas CpG y los genes globina π , α^D y α^A . **B)** Gráfica de expresión de los genes α -globina de pollo en eritrocitos a lo largo del desarrollo embrionario.

El dominio α -globina de pollo tiene tres genes codificantes para proteínas: π , α^D y α^A (Figura 9A), los cuáles residen en una configuración de cromatina abierta durante los primeros días del desarrollo de los eritrocitos de pollo (Rincón-Arano *et al.*, 2009). El gen π se expresa en los primeros días del desarrollo embrionario del pollo (1 – 5 días) (Figura 9B).

Sin embargo, en etapas posteriores del desarrollo (10 días), la expresión de π se abate como consecuencia de la hipermetilación del ADN, la unión de MeCP2 y la desacetilación de histonas (Valdes-Quezada *et al.*, 2013). Por su parte, los genes α D y α A presentan un patrón de expresión tardío en los eritrocitos embrionarios del pollo y una vez que se activa su expresión (día 6) (Figura 9B), ésta se mantiene el resto de la etapa embrionaria y de vida adulta del pollo, por ello se les conoce como los genes adultos del dominio α -globina (Rincón-Arano *et al.*, 2009).

La metilación del ADN que silencia al gen embrionario π se propaga por todo el gen π y hacia la región intergénica con el gen α D hasta alcanzar una zona específica en la que se encuentran dos dinucleótidos CpG constitutivamente metilados. Una observación original realizada por nuestro grupo demostró que CTCF se une al promotor de α D en una región inmediatamente río abajo a la zona hipermetilada en eritrocitos de 5 y 10 días de desarrollo (Martin *et al.*, 2011; Valdes-Quezada *et al.*, 2013). Se detectaron dos posibles motivos de unión al factor CTCF en el *locus* mencionado con la ayuda del *software MathInspector* para la identificación de secuencias consenso a factores de transcripción (Figura 10). Uno de estos motivos (CTCF-1) tiene además una gran similitud con el *logo* identificado para la unión de CTCF en el genoma del pollo, el cual representa la probabilidad de que ocurra un determinado nucleótido en una posición específica del motivo de unión a CTCF y que fue generado por el análisis de 15,000 sitios de unión a CTCF en el genoma del pollo con datos de ChIP-Seq (Figura 10) (Martin *et al.*, 2011).

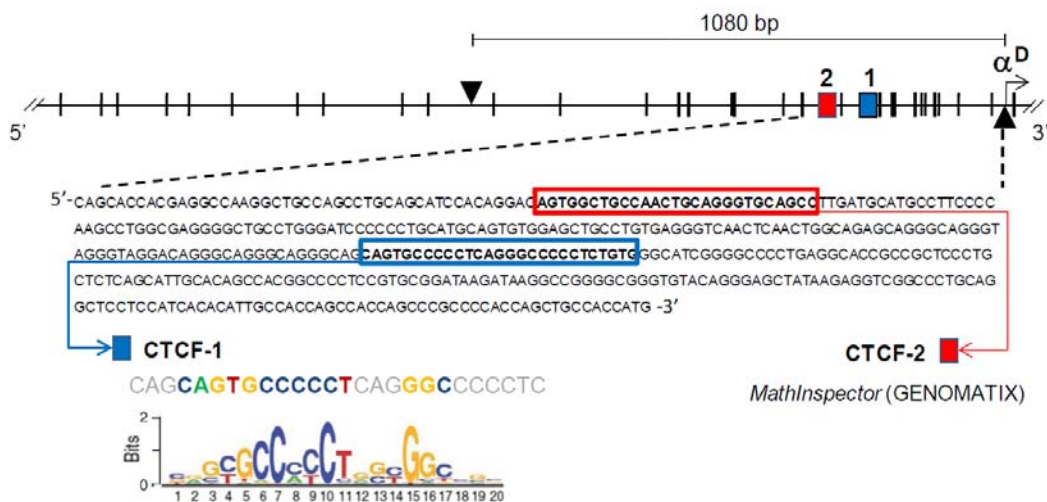


Figura 9. Posibles motivos de unión a CTCF río arriba al promotor del gen α D, obtenidos a partir de un análisis computacional (Tomado de Valdes-Quezada *et al.*, 2013).

Al realizar construcciones transgénicas con el gen *GFP* como reportero para estudiar la contribución de la unión de CTCF en la regulación de la expresión de los genes globina, se observó que la presencia del sitio de unión CTCF-1 es importante para evitar la metilación del promotor del transgén y con ello prevenir también el silenciamiento transcripcional del reportero GFP (Valdes-Quezada *et al.*, 2013).

A lo largo del dominio α -globina de pollo existen además una serie de elementos distales de regulación que parecen tener una función vinculada con la expresión coordinada de los genes globina durante el desarrollo embrionario y la diferenciación eritroide. Estos elementos se han caracterizado por ser sitios constitutivos o eritroide-específicos de hipersensibilidad a la DNasa I. En particular, existe un *enhancer* (*3'-enh*) que se encuentra a ~ 1 kb del gen αA en la posición 3' del dominio que ha sido ampliamente estudiado en nuestro grupo de investigación (Figura 8A; Targa *et al.*, 1993; Rincón-Arano *et al.*, 2005; Escamilla-Del-Arenal & Recillas-Targa, 2008). Se obtuvo evidencia, por ensayos de captura conformacional de cromosomas (3C), de que los promotores de π y αD interaccionan físicamente con el elemento *enhancer 3'-enh* del dominio α -globina en etapas tempranas del desarrollo del pollo (embriones de 5 días). Posteriormente (embriones de 10 días), la interacción con la región promotora del gen π se pierde y sólo se mantiene la del promotor αD -*3'-enh* (Figura 10; Valdes-Quezada *et al.*, 2013). De ésta manera el *3'-enh* está participando activamente en el recambio en la expresión de los genes α -globina de pollo durante el desarrollo embrionario. Dado que el *3'-enh* también tiene un sitio de unión a CTCF que está siendo ocupado *in vivo*, se propone que CTCF tiene un posible papel estructural para mediar la interacción entre el promotor αD y *3'-enh* (Figura 10).

2.2 Regulación del dominio α -globina de pollo en la diferenciación eritroide

La activación de la expresión de los genes globina que ocurre durante la diferenciación eritroide puede estudiarse a partir de líneas celulares arrestadas en distintos estadios de la eritropoyesis, lo que significa que dichas células están comprometidas a diferenciarse hacia un linaje eritroide adulto (Beug *et al.*, 1979). En particular, las células HD3 de pollo son un ejemplo de este tipo de líneas celulares. Las células HD3 son eritroblastos de pollo arrestados en su diferenciación que no expresan los genes globina; sin embargo, en condiciones experimentales es posible inducir su diferenciación y la activación de la expresión de los genes adultos α D y α A.

Algunas de las diferencias en las características cromatínicas del dominio α -globina de pollo entre células HD3 y eritrocitos embrionarios son: 1) la región promotora del gen α D se encuentra hipermetilada; 2) CTCF no se une al promotor del gen α D; y 3) la organización tridimensional del dominio (Figura 10). Una vez inducida la diferenciación, el promotor del gen α D se desmetila, se une CTCF en ésta región y se adquiere una configuración tridimensional del dominio similar a la de los eritrocitos embrionarios de 10 días de desarrollo (Figura 10; Gavrillov & Razin, 2008; Valdes-Quezada *et al.*, 2013); siendo éstos algunos de los mecanismos que permiten la activación de la expresión de los genes adultos.

Respecto a la organización tridimensional del dominio α -globina de pollo es interesante que el promotor de α D interactúa con otros elementos del dominio, y no así con el 3'-*enh*, en estadios más tempranos de la diferenciación eritroide (Gavrillov & Razin, 2008). Por lo que la interacción promotor α D – 3'-*enh* debe ser muy importante en el proceso de diferenciación y activación de los genes α -globina de pollo (Gavrillov & Razin, 2008; Valdes-Quezada *et al.*, 2013). Debido a las características mencionadas decidimos utilizar a las células HD3 como sistema biológico modelo para estudiar la contribución de CTCF en la regulación de la formación del dominio de cromatina que incluye a los genes α -globina de pollo y en la activación del gen adulto α D.

3. PLANTEAMIENTO DEL PROBLEMA

Los cambios en el estado de la cromatina y la formación de dominios a través de asas, son causa y/o consecuencia del encendido y apagado de genes durante la diferenciación celular. Por lo tanto, conocer los elementos que están involucrados en la formación regulada de dichas asas cromatínicas, permitirá comprender mejor los mecanismos que dan origen a determinado linaje celular.

En el dominio α -globina de pollo se propone que en eritrocitos de etapas tempranas del desarrollo embrionario el promotor del gen π está interactuando con el $3'$ -*enh* formando un asa de cromatina activa, la cual posteriormente se disocia permitiendo así: 1) la metilación del gen embrionario y consecuentemente su silenciamiento, y 2) la activación de la expresión de los genes adultos α^D y α^A , en particular el gen α^D que forma un asa cromatínica regulada por las proteínas CTCF que se unen su promotor y al $3'$ -*enh* (Figura 10). Abordar el estudio de la formación de dicha configuración activa del dominio desde el proceso de diferenciación eritroide, utilizando las células HD3 como modelo, permite preguntar puntualmente el papel de la unión de CTCF en el promotor de α^D para la activación transcripcional de los genes adultos α^D y α^A (Figura 10).

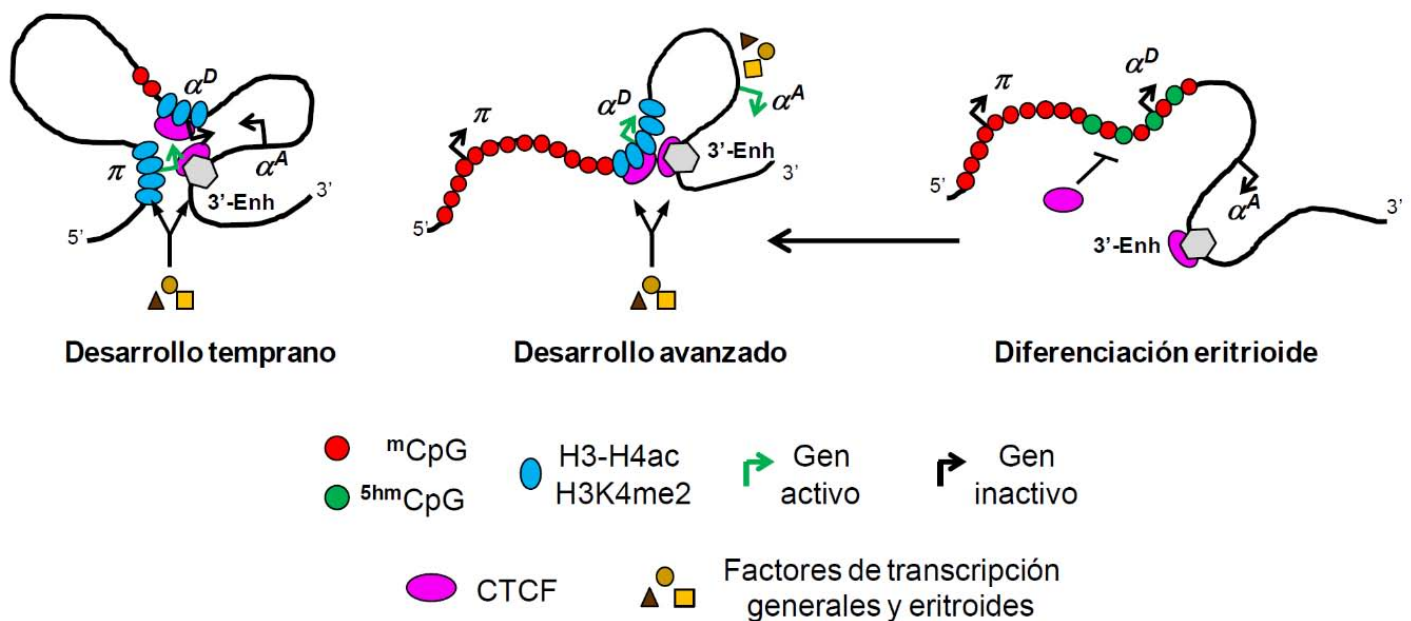


Figura 10. Modelo de regulación a nivel de cromatina del dominio α -globina de pollo (Tomada de Valdes-Quezada *et al.*, 2013).

4. HIPÓTESIS

El factor CTCF permite la interacción física entre el promotor α D y el 3'-*enh* en el dominio α -globina de pollo, a través de la formación de un asa de cromatina, activando la expresión de los genes adultos α D y α A.

5. OBJETIVO GENERAL

Determinar si la unión del factor CTCF a la región promotora del gen adulto α D favorece su interacción física con el 3'-*enh* y permite el establecimiento del estado de transcripción activa en los genes adultos del dominio en la etapa tardía de la diferenciación eritroide.

6. OBJETIVOS PARTICULARES

1. Validar, por medio de ensayos de retardo, la unión de CTCF a la secuencia predicha que se encuentra 190 pares de bases río arriba al sitio de inicio de la transcripción del gen adulto α D.
2. Remover, utilizando el sistema de edición CRISPR-Cas9, el sitio de unión a CTCF localizado río arriba del promotor del gen α D en el genoma de células HD3.
3. Comparar los cambios en la expresión de los genes del dominio α -globina de pollo en células HD3 Δ CTCF-1 y células HD3 diferenciadas.

7. MATERIAL Y MÉTODOS

1. Cultivo celular

La línea celular HD3, que corresponde a eritroblastos arrestados por la infección del virus termosensible AEV (*avian erythroblastosis virus*) (Graf 1975; Beug *et al.* 1979, 1982), se mantuvo en medio DMEM suplementado con 8% de suero fetal bovino (SFB), 2% de suero de pollo (ChS) y 1% de penicilina/estreptomicina (P/S). Se indujo la diferenciación de la línea celular HD3 utilizando 20 μ M del agente Iso-H-7 (Sigma, inhibidor de la proteína cinasa C) y 10 mM Hepes pH 8.0 en el medio de cultivo; en dichas condiciones se incubaron los cultivos por 48 h en una atmósfera de 1% CO₂ a 42°C, lo anterior inactiva el virus AEV y permite la diferenciación terminal de esta línea celular.

2. Plásmidos para edición con el sistema CRISPR Cas9: Alineamiento y clonación de gRNAs

Los pares de oligonucleótidos gRNAs complementarios fueron fosforilados y alineados agregando 1 μ L (100 μ M) de cada par de oligo, 1 μ L de 10X T4 Ligation Buffer (NEB), 0.5 μ L de la enzima T4 PNK (NEB) y 6.5 μ L de H₂O inyectable para un volumen final de 10 μ L de reacción en un tubo para PCR. La reacción se incubó en un termociclador a 37°C por 30 min, 95°C por 5 min y en un gradiente de reducción de temperatura de 5°C/min (0.1°C/s) hasta alcanzar los 25°C. Los oligos alineados se diluyeron 1:100 (100 veces) antes de clonarlos en los vectores CRISPR.

Los plásmidos PX330 (Addgene #42230) y lentiCRISPR-v2 (Addgene #52961) fueron digeridos con las enzimas BbsI (NEB) y BsmBI (NEB), respectivamente, como a continuación se describe. Para linearizar 1 μ g de PX330 se agregaron 2.5 μ L de 10X NEB Buffer 2.1, 1 μ L de enzima BbsI (NEB) y la cantidad faltante de H₂O inyectable para un volumen final de 25 μ L de reacción. La reacción se incubó a 37°C por 2 h y posteriormente se inactivó la enzima a 65°C por 20 min. Para la linearización de lentiCRISPR-v2 se agregó el volumen correspondiente a 2 μ g de plásmido, 2 μ L de BsmBI (NEB R0580), 5 μ L 10X NEB Buffer 3.1 y la cantidad restante de H₂O inyectable para un volumen final de 50 μ L. La reacción se incubó a 55°C por 1 hr. Los plásmidos linearizados fueron purificados utilizando el kit para extracción de gel QIAquick y eluidos en H₂O inyectable.

Las reacciones de ligación se llevaron a cabo en un volumen final de 20 μ L agregando 50 ng del plásmido linearizado y purificado, 1 μ L de oligo gRNA dúplex, 2 μ L de 10X T4 DNA Ligase Buffer, 1 μ L de T4 DNA Ligase y la cantidad restante de H₂O inyectable. Las reacciones se incubaron *overnight* (ON) a 16°C y después se inactivó la enzima a 65°C por 10 min. Se transformaron 5 μ L de reacción de ligación en células competentes TOP10.

3. Extractos nucleares de células HD3

Se transfieren 30 mL de cultivo de células HD3 a 80-90% de confluencia a un tubo falcon de 50 mL y se centrifuga a 1,500 rpm, 4°C por 5 min. El *pellet* de células obtenido se resuspende en 10 mL de buffer de fosfatos (PBS) para eliminar la mayor cantidad de residuos de medio de cultivo, se transfiere a un tubo falcon de 15 mL y se centrifuga nuevamente a 1,500 rpm, 4°C por 5 min. Se elimina el sobrenadante, se coloca el tubo con la muestra en hielo y se resuspende con 2.5 mL de buffer NI (15 mM Tris-HCl pH 7.5, 300 mM Sucrosa, 10 mM HEPES, 60 mM KCl, 4 mM MgCl, 5 mM NaCl, 1 mM PMSF). Una vez que las células se han resuspendido se agregan 2.5 mL de buffer NI + NP-40 1% y se pipetea lentamente 3 veces con una pipeta fría de 5 mL (1 min máx). Se centrifuga ahora por 5 min a 1,400 rpm, 4°C. Se retira el sobrenadante con pipeta Pasteur y vacío con cuidado de no tocar los núcleos, se resuspenden en 10 mL de buffer NI y se centrifuga nuevamente con las mismas condiciones. Se realiza otro lavado con 5 mL de buffer NI y después de la ronda de centrifugación, y de eliminar el sobrenadante, se resuspende el *pellet* en 50 μ L de buffer RIPA para extraer la proteína nuclear. Se transfiere la muestra a un tubo eppendorf de 1.5 mL, se incuba en hielo por 15 min y se centrifuga a 13,000 rpm, 4°C por 5 min. El sobrenadante se transfirió a otro tubo eppendorf y se cuantificó la proteína nuclear en el extracto por el método de Bradford.

4. Ensayo de movilidad electroforética (EMSA) y súper-retardo

Los oligonucleótidos se alinearon y marcaron radiactivamente usando PNK (Biolabs) y [γ ³²P] dATP. Para cada ensayo se utilizó 7.5 μ L de binding buffer (20 mM HEPES pH 7.9, 0.1 mM EDTA, 50 mM KCl, 5% Glicerol, 200 μ g/mL BSA), 1 μ L de poli dI-dC (como competidor inespecífico), 9-15 μ g de proteína nuclear y la cantidad de agua inyectable correspondiente para un volumen final de 25 μ L de reacción. Las muestras se incubaron 15 min en hielo y 15

min a temperatura ambiente antes de agregar la(s) sonda(s). En el caso de los súper-retardos el anticuerpo se incubó con la proteína 2-12 h antes de agregar la sonda. Después de la incubación se agregaron las sondas correspondientes (1 μL = 20,000 cpm), y se incubaron nuevamente las muestras por 15 min en hielo y a temperatura ambiente. Posteriormente, las muestras se cargaron en un gel de acrilamida al 5% no desnaturizante; la electroforesis se corrió primero a 100 V por 1 hr y después a 150 V por 2.5 h.

5. Transfección del sistema de edición genómica CRISPR-Cas9 (construcciones PX330)

Las células HD3 se lavaron dos veces con buffer de fosfatos salino (PBS) y se resuspendieron en medio DMEM. Posteriormente, se contaron en una cámara de Neubauer para utilizar 1×10^6 células en distintas condiciones de transfección. La cantidad de células indicada se transfirió a placas de 6 pozos y se transfectaron con 1 μg total de plásmidos (PX330-gRNA1 + PX330-gRNA2), con cantidades variables de Lipofectamine 2000 (2, 3, 4 y 5 μL) ó Fugene (4, 5 y 6 μL), de acuerdo a las indicaciones del protocolo correspondiente a cada reactivo. Después de 6 h se añadieron 4 mL de medio de cultivo HD3 a las células transfectadas. A las 72 h se extrajo ADN genómico para comprobar la eliminación de la región de interés por PCR.

6. Transducción del sistema de edición genómica CRISPR-Cas9 (construcciones LentiCRISPR-v2) y aislamiento de clones

Se generaron partículas virales, con los plásmidos pMD2.G y psPAX2, que contuvieran a las construcciones lentiCRISPR-v2 (gRNA1 + gRNA2 y gRNA3 + gRNA4). Las células HD3 fueron transducidas con las partículas virales producidas y 24 h posteriores a la infección se cambiaron a medio de cultivo HD3 para dejarlas recuperarse. Se agregó puromicina (0.5 mg/mL) a los cultivos 24 h después, y se extrajo ADN genómico en distintos días posteriores al inicio de la selección con puromicina para comprobar por PCR que la mutación estaba ocurriendo. Después de 14 días en selección con puromicina se realizaron diluciones clonales en cajas de 96 pozos, se dejó expandir las células por 7 días y posteriormente se verificó por PCR qué tipo de mutación tenían las poblaciones aisladas.

8. RESULTADOS

8.1 CTCF se une *in vitro* al sitio CTCF-1 del promotor del gen adulto α D

Valdés-Quezada (2013) comprobó por ensayos de ChIP que el sitio CTCF-1 es necesario para la unión de CTCF en la región *upstream* al promotor del gen α D. Para validar la interacción de CTCF con ese sitio específico se realizaron ensayos de retardo en gel (EMSA), para ello se diseñó un oligonucleótido de 63 pb que contuviera al centro la secuencia del motivo CTCF-1. La sonda CTCF-1 fue marcada radioactivamente y posteriormente incubada con distintas cantidades de proteína nuclear de células HD3 para determinar la cantidad de proteína óptima que nos permitiría detectar una banda de interacción sonda-proteína nuclear. Se decidió utilizar 15 μ g de extracto nuclear en los experimentos posteriores porque se logra obtener un retardo bien definido al incubar la proteína con la sonda marcada (Figura 11).

Ext. Nuclear HD3 (μ g)	-	3	6	9	12	15	30	40
CTCF-1	+	+	+	+	+	+	+	+

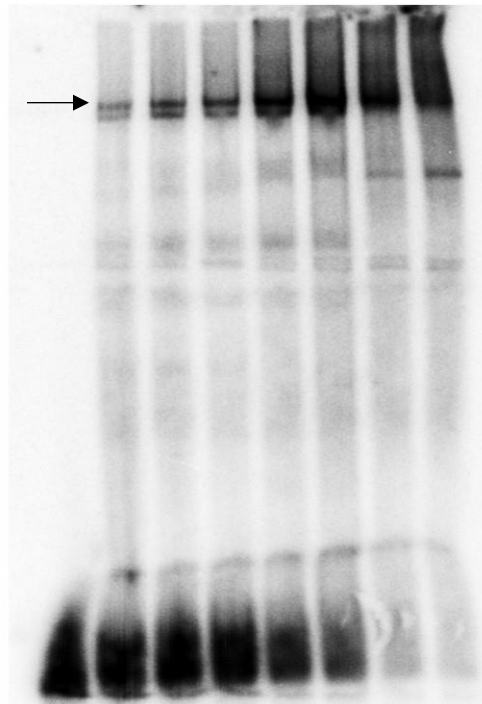


Figura 11. Ensayo de retardo con distintas concentraciones de extractos nucleares de HD3. La flecha indica el retardo detectado, y los números corresponden a las cantidades en μ g de extractos de proteína nuclear de células HD3.

Con el objetivo de confirmar que el retardo que se generó es producto de la interacción de CTCF con la sonda marcada se repitieron los ensayos poniendo a competir a CTCF-1 con una sonda fría (sin marcaje radiactivo) conteniendo la secuencia FII del *insulator* β -globina de pollo, por la cual CTCF tiene una alta afinidad de unión (Bell *et al.* 1999). También se utilizó un exceso de sonda fría CTCF-1 como competidor control (Figura 12). Además algunas muestras se incubaron con cantidades crecientes del anticuerpo anti-CTCF con el objetivo de generar un súper retardo dependiente de la interacción del anticuerpo con el complejo CTCF-sonda marcada. Las sondas frías FII y CTCF-1 compitieron con la sonda marcada por la unión a CTCF; es decir, al utilizar éstas en exceso molar de 100 veces no se observó el retardo que suponemos depende de la interacción de CTCF con la sonda marcada CTCF-1 (Figura 12). Las muestras incubadas con anti-CTCF no generaron súper retardos evidentes; sin embargo, al utilizar 3 y 5 μ g de anticuerpo se observa un enriquecimiento en la intensidad de la banda (Figura 12). Lo anterior pudiera estar relacionado con la formación de un complejo anticuerpo – proteína – sonda marcada. Estas observaciones proveen evidencia de que el retardo que se genera es producto de la interacción de CTCF con el oligonucleótido marcado.

Ext. Nuclear HD3	-	+	+	+	+	+	+	+	+	+
CTCF-1	+	+	+	+	+	+	+	+	+	+
CTCF-1 (sonda fría)	-	-	+	-	-	-	-	-	-	-
FII (sonda fría)	-	-	-	+	-	-	-	-	-	-
Anti-CTCF	-	-	-	-	1	3	5	1	3	5

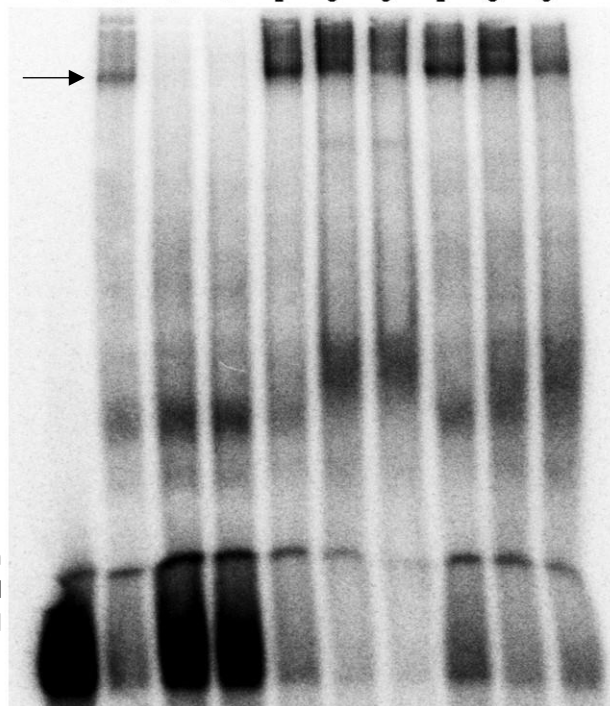


Figura 12. Ensayo de retardo y súper retardo con extractos nucleares de HD3. La flecha indica el retardo detectado. Los números indican la cantidad de anti-CTCF en μ g utilizada en los ensayos

En los ensayos de EMSA anteriores (Figuras 11 y 12) se observan barridos de la señal radiactiva de la sonda, esto puede deberse a su interacción con proteínas del extracto nuclear distintas a CTCF. Con el objetivo de reducir esas posibles interacciones inespecíficas de proteína – sonda marcada, se cambió el *buffer* de unión en el que se incuban las muestras por uno que contiene albúmina de suero bovino (*BSA*). Bajo estas condiciones, los ensayos de retardo con la sonda CTCF-1 pierden el barrido mencionado y aumentan la intensidad y el tamaño de la banda que correspondería al retardo (Figura 13). La presencia de *BSA* en el *buffer* debe estar ayudando a incrementar la unión específica de CTCF a la sonda CTCF-1 y por ello con 15 μ g de extractos nucleares se observa un retardo de mayor intensidad al obtenido anteriormente.

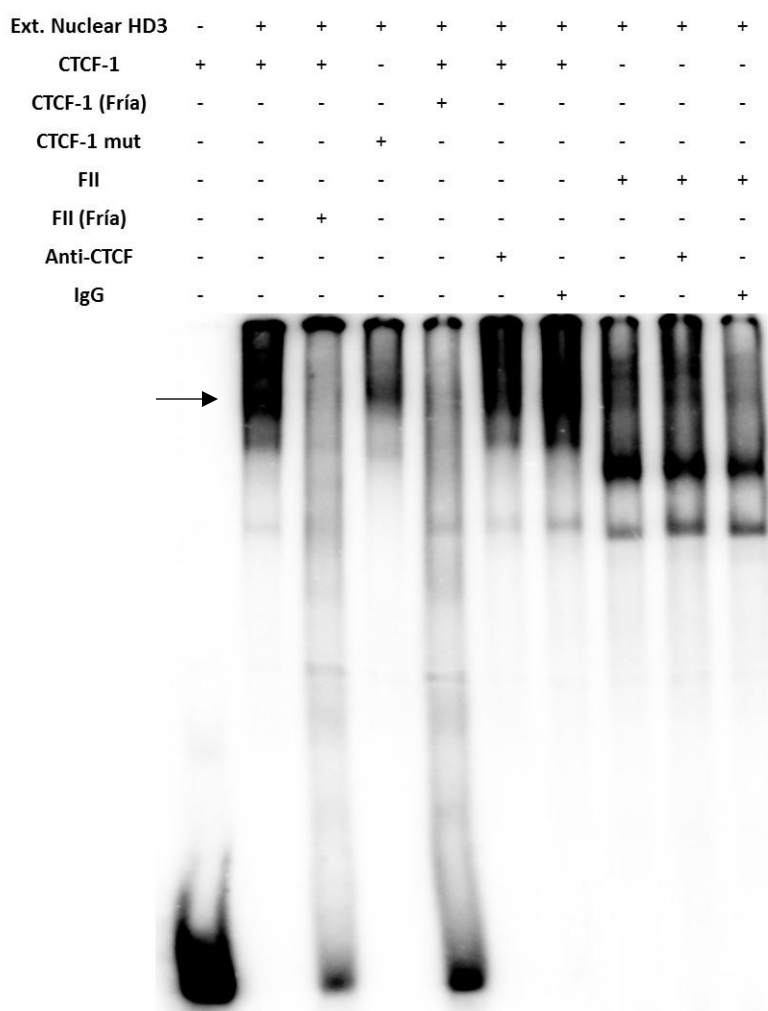


Figura 13. Ensayo de retardo y súper retardo de CTCF-1 y FII.

Respecto a los súper retardos, la incubación en presencia de IgG como control negativo generó un barrido de señal de mayor intensidad que el del anticuerpo anti-CTCF, lo cual es contrario a lo esperado (Figura 13, carriles 6 y 7). En el caso de FII marcado radiactivamente tampoco se observa un cambio evidente en el retardo que se genera en presencia – ausencia de anti-CTCF y la incubación control con IgG (Figura 13, carriles 8-10). Por lo tanto, es posible que el anticuerpo no esté uniéndose al factor CTCF eficientemente.

Para comprobar que el retardo observado se debe a la unión de CTCF al motivo de unión CTCF-1 en la sonda marcada se diseñó una sonda con las citosinas del *core* del motivo de unión reemplazadas por adeninas (CTCF-1 mut). Al incubar los extractos nucleares de HD3 con la sonda CTCF-1 mut se observó una disminución en el complejo de retardo pero no una pérdida total de éste (Figura 13). Este patrón se repitió al incubar ambas sondas en un gradiente de cantidad de extractos de proteína nuclear (Figura 14); es decir, es necesaria una menor cantidad de proteína nuclear para observar una banda de retardo intensa con la sonda CTCF-1 que con la sonda CTCF-1 mut.

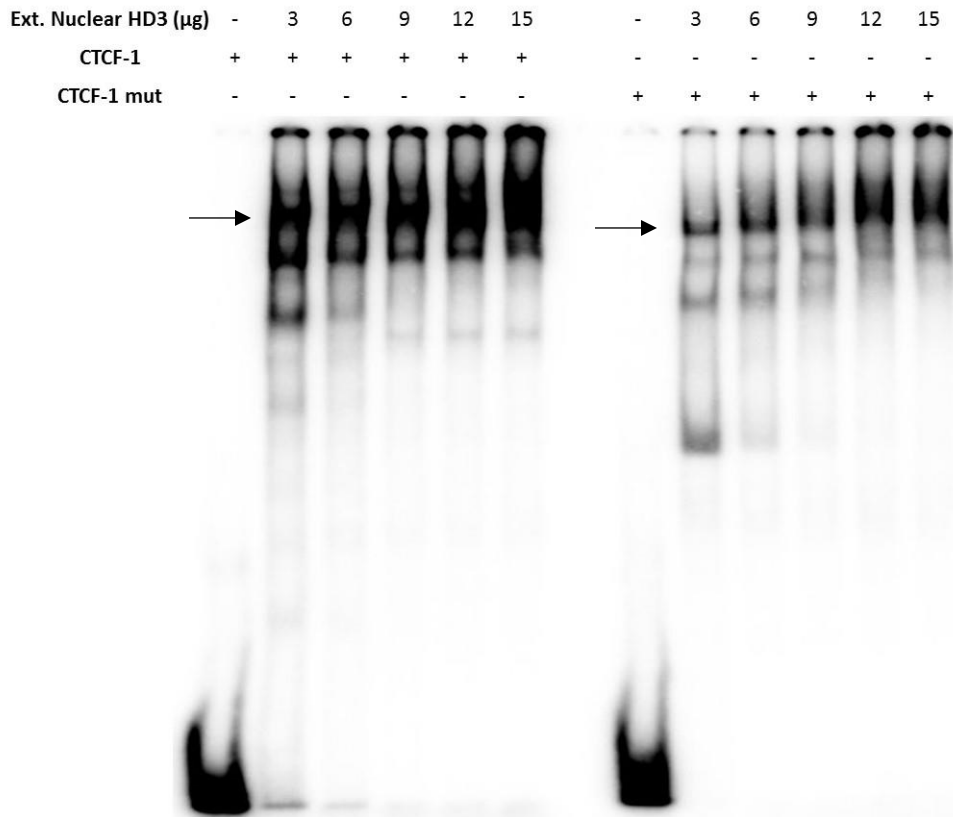


Figura 14. Ensayo de retardo de la sonda CTCF-1 y CTCF-1 mut con distintas cantidades de proteína nuclear. Los números corresponden a las cantidades en µg de extractos de proteína nuclear de células HD3.

Incluir *BSA* en el buffer de unión durante la incubación de la sonda marcada y los extractos nucleares de HD3 permitió detectar un retardo evidente al utilizar 9 μg de proteína nuclear (Figura 14). Utilizando esta condición se incubó nuevamente el anticuerpo anti-CTCF con los extractos nucleares por un período de 12 h a 4°C, esto con el objetivo de aumentar la probabilidad de interacción entre ellos y poder observar el súper retardo. Bajo estas condiciones se observó una disminución en la intensidad de la banda correspondiente al control de IgG (Figura 15, carril 7); sin embargo, no fue posible detectar un cambio evidente entre el retardo y súper retardo de CTCF-1 (Figura 15, carriles 2 y 6). Se observó nuevamente una disminución en el retardo de CTCF-1 mut, pero también la formación de nuevos complejos con mayor migración (Figura 15, carril 5). El cambio de citosinas por adeninas en la sonda podría estar generando sitios de reconocimiento para otras proteínas del extracto nuclear y por ello se observan los nuevos retardos. Nuevamente las sondas frías FII y CTCF-1 compitieron por la unión a CTCF con la sonda marcada CTCF-1; por lo que, este resultado provee la mayor evidencia de que el retardo que se genera es producto de la interacción de CTCF con el oligonucleótido marcado.

Ext. Nuclear HD3	-	+	+	+	+	+	+
CTCF-1	+	+	+	+	-	+	+
CTCF-1 (Fría)	-	-	+	-	-	-	-
FII (Fría)	-	-	-	+	-	-	-
CTCF-1 mut	-	-	-	-	+	-	-
Anti-CTCF	-	-	-	-	-	+	-
IgG	-	-	-	-	-	-	+

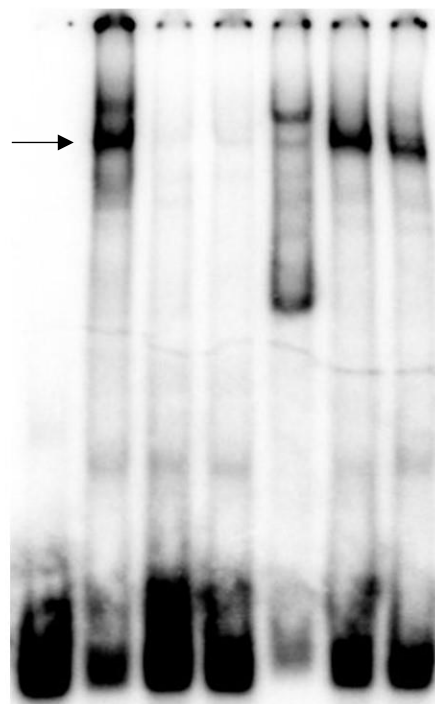


Figura 15. Ensayo de retardo y súper retardo de la sonda CTCF-1. Los extractos nucleares fueron incubados con el anticuerpo anti-CTCF de pollo o IgG por 12 h a 4°C para los súper retardos.

8.2 Diseño de gRNAs y oligonucleótidos para la eliminación del sitio de unión CTCF-1 por CRISPR-Cas9

Se hizo un análisis computacional para diseñar los posibles pares de gRNAs que permitan remover el motivo de unión a CTCF en la región promotora del gen adulto α^D con la endonucleasa hSpCas9, para ellos se utilizó el software para el diseño de gRNAs desarrollado en laboratorio del Dr. Feng Zhang (accesible en: <http://crispr.mit.edu/>). Se introdujo la secuencia (250 bp) de la región que contiene el motivo de unión CTCF-1, y se seleccionó como blanco el genoma de pollo para analizar también los posibles *off-targets* de las secuencias guías en todo el genoma, de los resultados de este análisis se eligieron dos pares de gRNAs, los cuáles se nombraron CRISPR1 (gRNA1 + gRNA2) y CRISPR2 (gRNA3 + gRNA4) para identificarlos fácilmente.

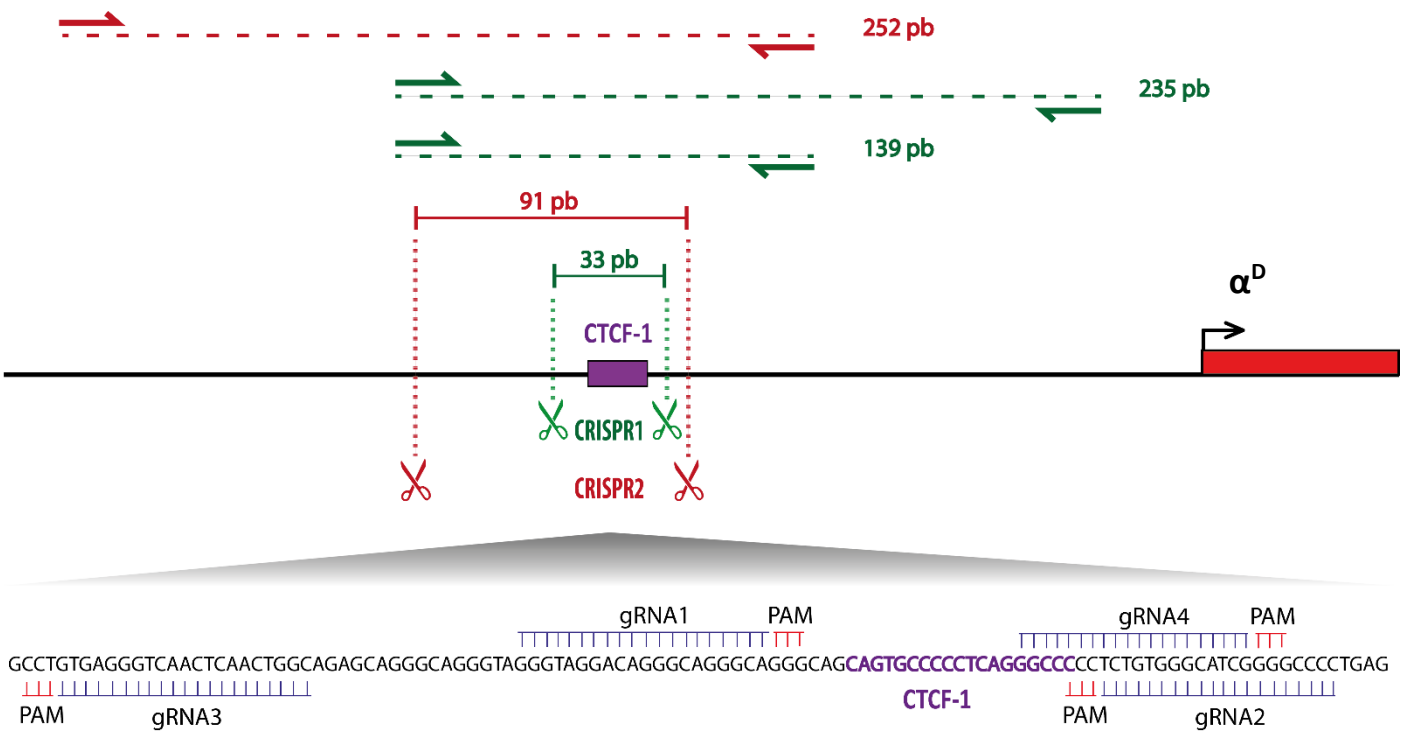


Figura 16. Sitios de unión de los gRNAs diseñados para eliminar el motivo de unión a CTCF en la región promotora del gen adulto α^D . Las tijeras y líneas punteadas indican los sitios de corte esperados con el sistema CRISPR-Cas9 al utilizar los pares de guías gRNA1 + gRNA2 (CRISPR1) y gRNA3 + gRNA4 (CRISPR 2); las eliminaciones predichas con estos pares de gRNAs corresponden a 33 pb y 91 pb, respectivamente. Las flechas representan la posición de los oligonucleótidos que permitirán la genotipificación por PCR de ambas eliminaciones y se esquematiza el tamaño de los amplicones esperado con ADN genómico de células HD3 wt. El color del amplicón indica si permite genotipificar la eliminación CRISPR1 (verde) o CIRPSR2 (rojo). Los sitios de unión de los gRNA diseñados se representan con esquemas azules y en rojo se indica la localización de la secuencia PAM (NGG) en el extremo 3' de la secuencia a la que se unen los gRNAs. En morado se indica la posición del sitio de unión a CTCF en la región promotora del gen α^D .

El par de gRNAs CRISPR1 se eligió con el objetivo de utilizar dos gRNAs que permitieran realizar la escisión del motivo CTCF-1 quitando la menor cantidad posible de bases adyacentes (Figura 16). Lo anterior con la finalidad de modificar lo menos posible el contexto genómico del promotor del gen α D. Al utilizar los gRNA1 y gRNA2 en el sistema de edición CRISPR-Cas9 se estima una eliminación de alrededor de 33 pb a una distancia de 181 pb del sitio de inicio de la transcripción del gen α D (Figura 16). Respecto al par de gRNAs CRISPR2, éstos se escogieron considerando aquellos gRNAs que tuvieran la menor probabilidad de producir *off-targets* dentro del genoma de pollo, de acuerdo a la predicción del *software* de diseño de gRNAs. Los gRNAs con estas características se nombraron gRNA3 y gRNA4, se estima que con ellos se elimine una región de 91 pb a una distancia de 174 pb del sitio de inicio de la transcripción del gen α D (Figura 16).

8.3 El sistema de edición genómica CRISPR-Cas9 permite eliminar la región correspondiente al motivo CTCF-1

a) Eliminación de 33 pb - CRISPR1

Como abordaje inicial se comenzó a trabajar con los gRNAs con la capacidad de producir la menor eliminación de bases posible, incluyendo el motivo CTCF-1; es decir, el par de guías CRISPR1. Para poder corroborar que el sistema de edición funciona después de transfectarlo en las células HD3, se diseñaron y estandarizaron oligonucleótidos para PCR que enmarcaran los sitios blancos del par de gRNAs (Figura 16). Si la región blanco es escindida, se espera obtener productos de PCR de menor tamaño que el de las células sin editar.

Se mandaron sintetizar oligonucleótidos con las secuencias de los gRNA1 y gRNA2, así como sus secuencias complementarias. Los pares de oligonucleótidos complementarios fueron alineados y clonados en el vector pX330 (Addgene). El par de plásmidos obtenidos fueron co-transfectados utilizando liposomas (Lipofectamina 2000 ó Fugene) en distintas condiciones de transfección (número de células, cantidad de ADN, relación liposomas : ADN) para obtener la condición óptima en que ocurre la eliminación esperada (Figura 17). A las 72 h de esto se extrajo ADN genómico. Se realizaron las PCRs correspondientes con los oligonucleótidos diseñados para verificar la eliminación por el par CRISPR1, y se corrieron

en un gel de acrilamida al 5%. Los productos de PCR de las células transfectadas con el sistema de edición CRISPR1-Cas9 mostraron el patrón de bandeo esperado, una banda de 235 pb (correspondiente a las células sin editar) y otra banda de ~205 pb, lo que indicaba que la eliminación del sitio de unión CTCF-1 parecía estar ocurriendo (Figura 17).

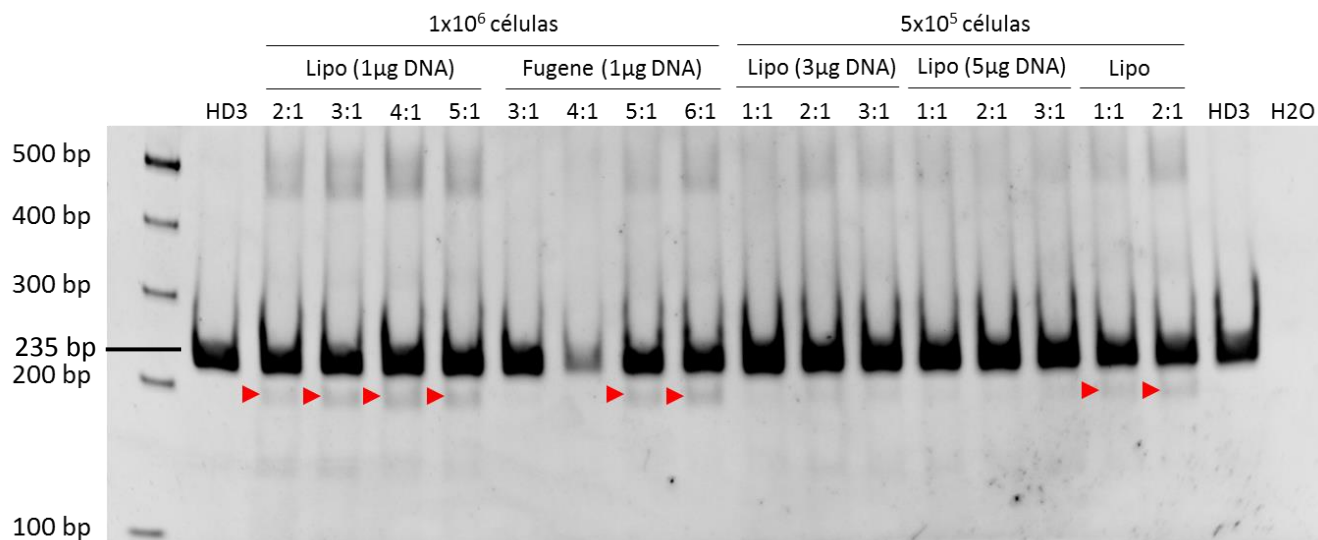


Figura 17. Eficiencia de sistema de edición CRISPR1 en distintas condiciones de transfección. Fragmentos amplificados a partir de ADN genómico de células HD3 control (HD3), células HD3 transfectadas con Lipofectamina 2000 (Lipo) o Fugene con el sistema de edición CRISPR1 (cantidad total de plásmidos PX330-gRNA1 y PX330-gRNA2 utilizados se muestra en paréntesis). Se indica el número inicial de células para la transfección y la relación Lipofectamina 2000/Fugene : ADN utilizada. Las flechas señalan el tamaño de fragmento esperado al eliminar el motivo CTCF-1.

A pesar de que las condiciones de transfección utilizadas permitieron detectar que la eliminación ocurre con mayor frecuencia en unas que en otras; la proporción de células editadas parecía ser muy baja como para obtener poblaciones enriquecidas de clones HD3 Δ CTCF-1. Por ésta razón se optó por cambiar a un sistema de transducción lentiviral del plásmido LentiCRISPR-v2 (Addgene) con los gRNAs clonados. LentiCRISPR-v2 contiene un gen de resistencia a puomicina por lo que es posible seleccionar a las células que han sido infectadas exitosamente. Las construcciones LentiCRISPR-gRNA1 y LentiCRISPR-gRNA2 fueron co-infectadas con partículas lentivirales en células HD3 y 24 h post-infección se seleccionaron con 0.5 μ g/mL de puomicina por 14 días. Se extrajo ADN genómico al día 0, 3, 5, 7, 10 y 14 de selección en puomicina con el cual se hicieron las PCRs de genotipificación para verificar la eliminación del sitio CTCF-1 y los productos se corrieron en un gel de acrilamida al 5% (Figura 18).

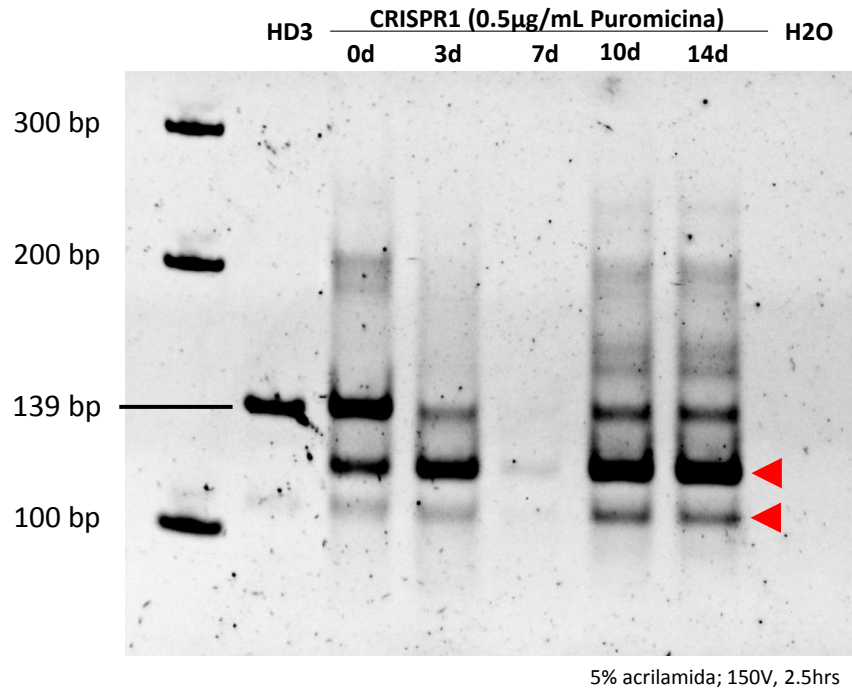


Figura 18. Selección de células HD3 CRISPR1 con puromicina y genotipificación por PCR. Los oligonucleótidos diseñados para genotipificar la eliminación CRISPR1 amplifican una región de 139 pb que contiene al motivo CTCF-1 a partir de ADN genómico de células HD3 wt. Las bandas de menor tamaño obtenidas al realizar la PCR de genotipificación con ADN de células HD3 infectadas con el sistema lentiviral CRISPR1 (flechas rojas), indican que ocurrieron eliminaciones de distintos tamaños en la región estudiada.

La utilización de este sistema lentiviral permitió detectar dos bandas enriquecidas de menor tamaño (Figura 18, flechas rojas) respecto al esperado para la banda de las HD3 sin tratamiento (139 pb). La eficiencia en la utilización del sistema de transducción lentiviral también se evidencia al observar una mayor intensidad en las bandas CRISPR inmediatamente 24 h posteriores a la infección de las células HD3 (Figura 18, 0d), respecto a la banda CRISPR observada en la genotipificación de células HD3 transfectadas con Lipofectamina o Fugene (Figura 17). Además, las poblaciones de células HD3 editadas se van enriqueciendo en función del tiempo de selección con puromicina (Figura 18).

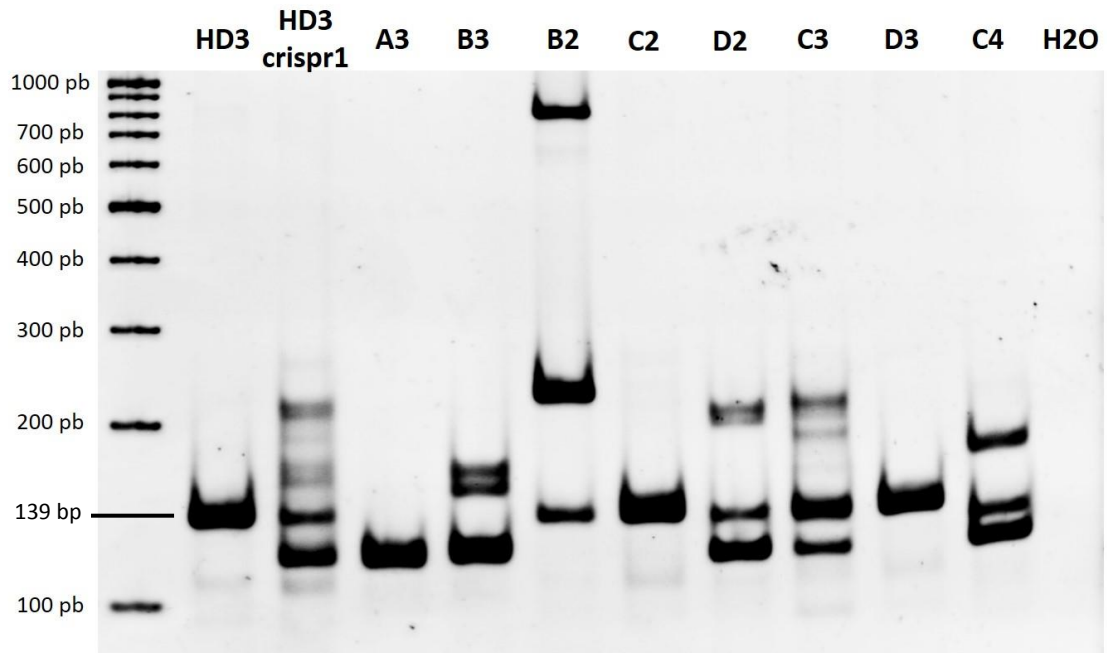


Figura 20. Eliminaciones e inserciones detectadas por PCR en las clonas de las células HD3 CRISPR1. Se muestra la genotipificación por PCR de la población de células HD3 CRISPR1 (*pool* de células infectadas con el sistema CRISPR1 y seleccionadas con 0.5 $\mu\text{g}/\text{mL}$ de puromicina por 14 días), y de algunas de las clonas obtenidas por dilución clonal de las células HD3 CRISPR1. De las clonas obtenidas, las células HD3 A3 y B3 mostraron la pérdida total del amplicón de 139 pb. En la clona HD3 B3 se detectaron además dos bandas de mayor tamaño, lo cual podría indicar la inserción de fragmentos de mayor tamaño en alguno de los alelos de las células HD3 al momento de reparar el corte de doble cadena.

b) Eliminación de 91 pb - CRISPR2

De igual manera que la edición de CTCF-1 con el sistema CRISPR1, se indujo la eliminación de CTCF-1 utilizando el sistema lentiviral para CRISPR2 (gRNA3 + gRNA4) en las células HD3. Al realizar la genotipificación por PCR con el ADN genómico de las células editadas con CRISPR2 y seleccionadas con puromicina se alcanzó a observar un barrido debajo de la banda wt (252 pb), además de una banda ~90 pb más pequeña (Figura 21, flecha roja). Sin embargo, a partir del día 7 en selección con puromicina se perdieron tanto la banda wt de 252 pb como la banda CRISPR2 de ~160 pb.

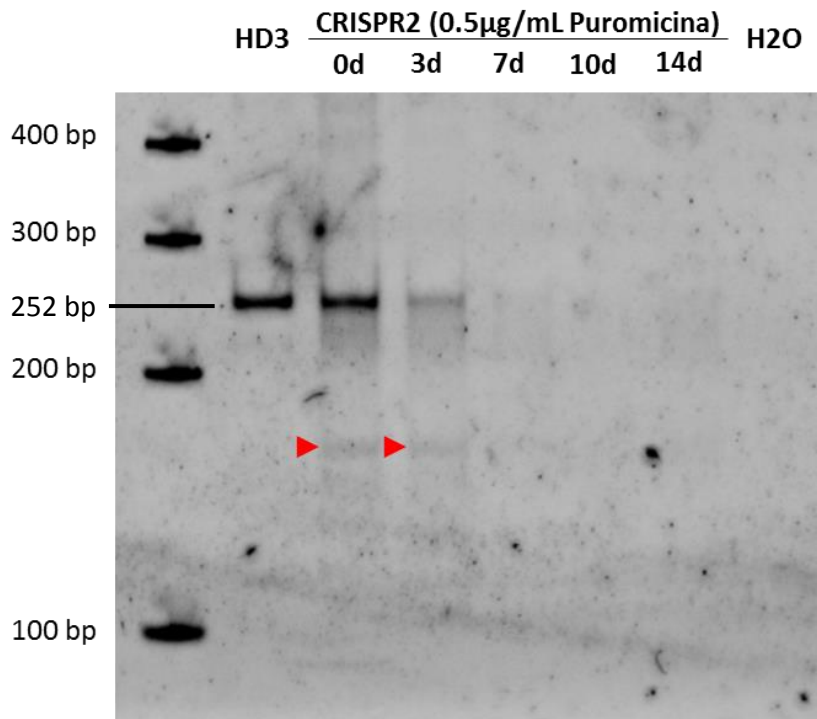


Figura 21. Selección de células HD3 CRISPR2 con puromicina y genotipificación por PCR. Los oligonucleótidos diseñados para genotipificar la eliminación CRISPR2 amplifican una región de 252 pb, que contiene al motivo CTCF-1, a partir de ADN genómico de células HD3 wt. Las bandas de menor tamaño obtenidas al realizar la PCR de genotipificación con ADN de células HD3 infectadas con el sistema lentiviral CRISPR2 (barrido), indican que ocurrieron eliminaciones de distintos tamaños en la región estudiada. Las flechas rojas señalan la banda correspondiente a la eliminación estimada de ~90 pb con el sistema CRISPR2.

8.4 La expresión del gen α D adulto disminuye en las células HD3 mutadas con el sistema CRISPR-Cas9

Los genes adultos α D y α A del dominio α -globina de pollo se encuentran transcripcionalmente activos a los 10 días del desarrollo en eritrocitos de pollo; mientras que el gen embrionario π se encuentra silenciado en ésta etapa del desarrollo. Las células HD3 son eritroblastos arrestados que una vez diferenciadas tienen un fenotipo similar al de eritrocitos maduros. Las células HD3 sin diferenciar mantienen los 3 genes en un estado inactivo. La inducción de su diferenciación a una temperatura de 42°C y la presencia del compuesto Iso-H7 por 48 h permite la detección de la expresión de los genes adultos por qRT-PCR (Figura 22A).

Sabemos que CTCF no se une al motivo CTCF-1 en el promotor del gen α D de las células HD3 hasta que se induce su diferenciación (Valdés-Quezada *et al.*, 2013), por lo que de acuerdo a la hipótesis presentada en este proyecto, su presencia es importante para la interacción del promotor con el 3'-*enh* y aumentar la expresión de al menos el gen adulto α D. Al cuantificar la expresión de dicho gen en células HD3 CRISPR1 diferenciadas se observó una reducción de ~ 4 veces de niveles del transcrito respecto a las HD3 diferenciadas (Figura 22C). Es probable que este resultado refleje el hecho de que CTCF no se está uniendo al promotor en una proporción de las células editadas HD3 CRISPR1; al no unirse, no se forma la interacción con el 3'-*enh* por lo que disminuye la expresión del gen α D.

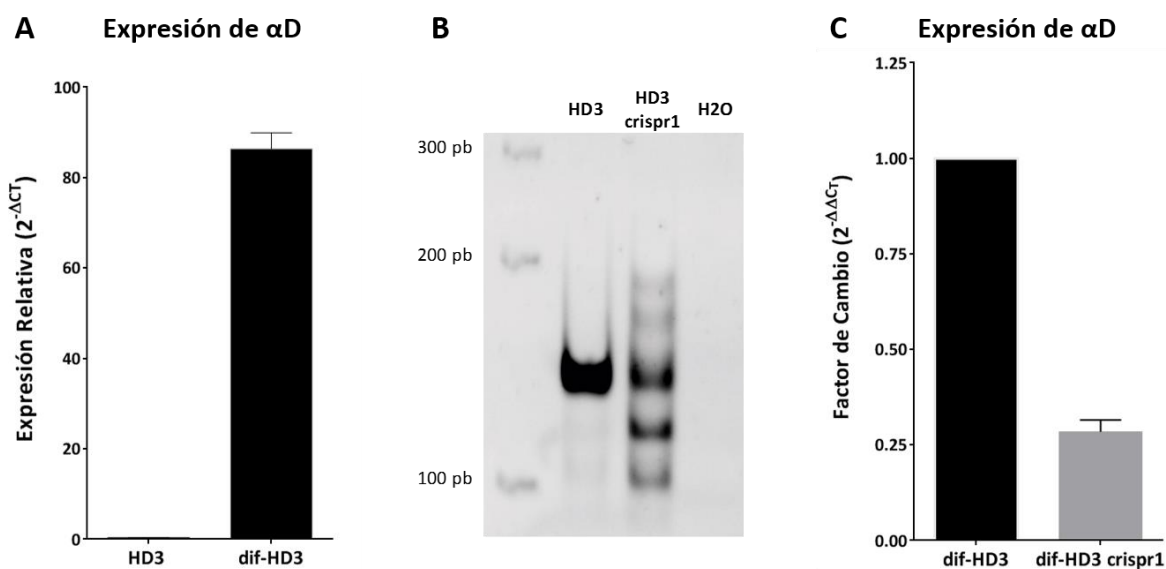


Figura 22. Expresión del gen α D en células HD3 y HD3 CRISPR1 diferenciadas. A) La inducción de la diferenciación en células HD3 activa la expresión del gen α D. B) Eliminaciones e inserciones detectadas por PCR en las células HD3 infectadas con el sistema de edición CRISPR-Cas9. C) La expresión del gen α D es menor en las células HD3 CRISPR1 diferenciadas que en las células HD3.

Sin embargo, al cuantificar los niveles del transcrito de α D en las monoclonas celulares HD3 A3 y B3, el efecto observado en su expresión fue distinto. En las células HD3 A3 no hubo cambio, mientras que en las células HD3 B3 se observó una tendencia a la disminución de la expresión del gen α D (Figura 23A). Estos resultados corresponden con las mutaciones que presentan las monoclonas celulares HD3 A3 y B3 en el sitio de unión a CTCF del promotor del gen α D. En las células HD3 A3 no se elimina ningún nucleótido del motivo de unión predicho, sino que se eliminan 21 pb que se encuentran *downstream* al motivo de unión (Figura 19). Para el caso de las células HD3 B3, la eliminación obtenida de 20 pb incluye los últimos 5 nt del extremo 3' del motivo de unión y 15 nt *downstream* al motivo. Una posible explicación a lo observado es que las bases eliminadas en las células HD3 B3 son importantes para el correcto posicionamiento de CTCF, por lo que su ausencia compromete la unión de CTCF y la expresión del gen α D.

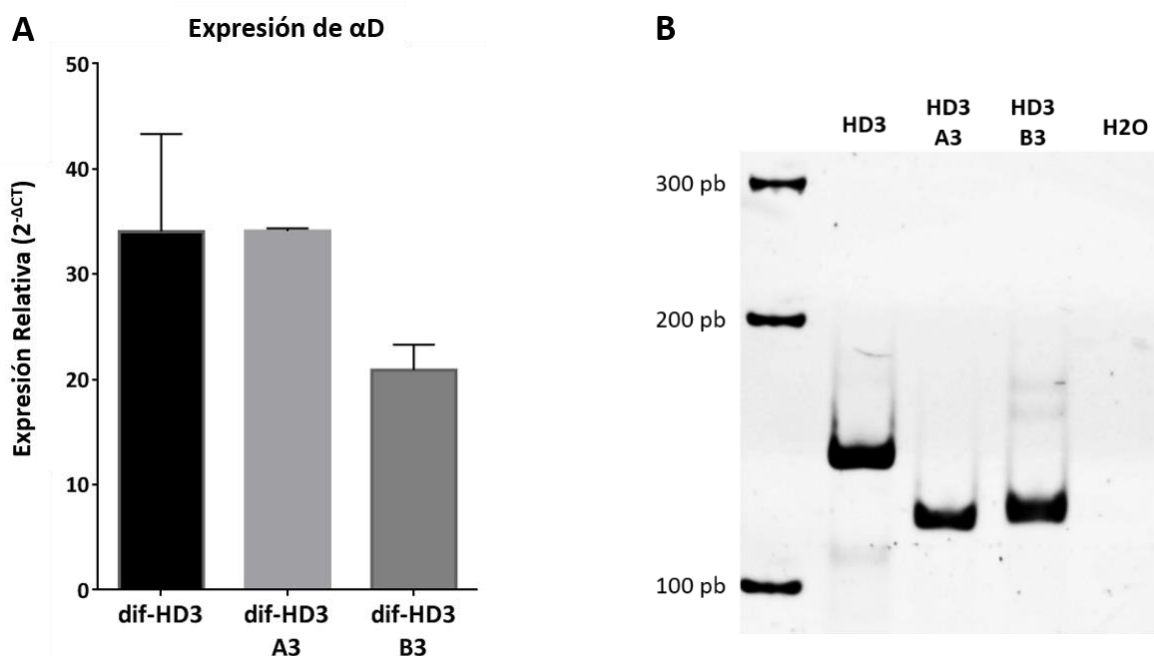


Figura 23. Expresión del gen α D en las monoclonas HD3 A3 y B3 diferenciadas. A) La expresión del gen α D es también menor en las monoclonas A3 y B3 diferenciadas respecto a las células HD3. B) Eliminaciones detectadas por PCR en las células HD3 A3 y HD3 B3.

9. DISCUSIÓN

En este trabajo de tesis se describe la importancia de la presencia de un sitio de unión a CTCF en el promotor del gen adulto αD del dominio α -globina de pollo para activar correctamente la expresión del gen correspondiente durante la diferenciación eritroide.

Los dominios génicos que contienen genes con expresión diferencial en tiempo y espacio son modelos biológicos relevantes para el estudio de los mecanismos que regulan la activación y silenciamiento de los genes. Los cambios en la estructura de la cromatina del dominio α -globina que ocurren en las células HD3 al inducir su diferenciación hacia eritrocitos maduros ofrecen la posibilidad de realizar preguntas puntuales acerca del papel de distintos elementos de regulación en la estructuración de un dominio de transcripción activa durante la diferenciación celular.

El dominio α -globina de pollo contiene 3 genes codificantes para la subunidad α de la hemoglobina que se expresan en distintos momentos del desarrollo en eritrocitos embrionarios. El gen embrionario π se expresa los primeros días del desarrollo y los genes αD y αA se activan a partir del día 6, momento en el que comienza a disminuir la expresión del gen π . El silenciamiento del gen π ocurre de manera autónoma al hipermetilarse el cuerpo del gen y su región intergénica aledaña al gen αD . La metilación se detiene antes de alcanzar el promotor del gen αD , lo cual permite que se mantenga en un estado de cromatina abierta que favorece su expresión. La presencia de un motivo de unión a CTCF (CTCF-1, identificado por experimentos de ChIP-seq y análisis de secuencia *in silico*) en el promotor del gen αD es aparentemente importante para evitar que se propague la metilación hacia el promotor y un gen reportero en experimentos con construcciones transgénicas.

En este trabajo se validó *in vitro* la unión de CTCF al motivo CTCF-1 con ensayos de retardo en la movilidad electroforética; esto porque al incubar la sonda CTCF-1 marcada radiactivamente con extractos nucleares de la línea celular eritroide de pollo HD3 se forman complejos de oligonucleótido-proteína. La presencia de CTCF en esos complejos se comprobó al utilizar la sonda fría FII (previamente caracterizada por unir a CTCF) como

competidor para la formación del complejo. En este ensayo se dejó de observar el retardo correspondiente al complejo proteína-oligonucleótido radiactivo porque FII titula a la proteína CTCF, impidiendo su interacción con la sonda radiactiva CTCF-1. Además al repetir el experimento con una sonda radiactiva mutada en el motivo CTCF-1 (CTCF-1_mut) también se perdió el retardo correspondiente; sin embargo, se observó la formación de complejos de proteína-oligonucleótido de distintos tamaños, los cuales podrían deberse a la interacción de otras proteínas con la secuencia mutada. Para realizar este experimento de forma más clara se propone mutar nuevamente el motivo CTCF-1 pero ahora utilizando la secuencia mutante previamente reportada en Valadez-Graham *et al.* (2004). Respecto al súper retardo con el anticuerpo para CTCF de pollo, es necesario estandarizar las condiciones para poder observarlo de forma evidente.

Después de comprobar que CTCF se une *in vitro* al motivo CTCF-1, procedimos a eliminar este sitio de unión en eritroblastos arrestados de pollo (HD3) utilizando el sistema de edición genómica CRISPR-Cas9. Se diseñaron los sistemas CRISPR1 para eliminar ~33 pb, y CRISPR2 para eliminar ~90 pb, ambas conteniendo el motivo de unión descrito. Las transfecciones utilizando el plásmido PX330 no permitieron enriquecer la población de células transfectadas eficientemente; por ello se optó por utilizar el plásmido LentiCRISPR-v2 que contiene un gen de resistencia a puromicina, el cual facilitó la selección de células transducidas y editadas con el sistema CRISPR-Cas9. Al amplificar las regiones blanco a mutar tanto con el sistema CRISPR1 como CRISPR2, se observaron bandas de distintos tamaños a la silvestre que indicaron la eliminación e inserción de fragmentos de ADN. Sin embargo, al prolongar la selección en puromicina de las células HD3 CRISPR2 se dejaron de recuperar bandas con la genotipificación por PCR. Una de las posibles explicaciones es que al repararse las lesiones ocasionadas por la endonucleasa Cas9 se modificó también uno o ambos sitios de apareamiento para los oligonucleótidos; es decir, generando una eliminación de mayor tamaño al esperado. Para comprobar lo anterior deben diseñarse oligonucleótidos externos a la región a editar que se encuentren al menos 50-100 pb más alejados, dado que a hasta ese rango abarcan los *indels* detectados por la edición CRISPR-Cas9 utilizando un solo gRNA.

En el caso de las células HD3 CRISPR1 fue sencillo realizar la genotipificación por PCR para comprobar la eliminación del sitio y al realizar diluciones clonales fue posible obtener poblaciones más homogéneas en cuanto a la mutación producida en la región promotora α D. Sin embargo, al analizar en el gel de electroforesis los tamaños de las bandas, ninguna de las poblaciones clonales obtenidas presentaba la eliminación esperada de 33 pb. No obstante, se secuenciaron las bandas para las poblaciones de células HD3 A3 y HD3 B3; y se observaron eliminaciones de 21 pb y 20 pb, respectivamente. La eliminación en las células HD3 A3 se encuentra inmediatamente río abajo al motivo CTCF-1, mientras que la de HD3 B3 elimina 5 pb de la región 3' del motivo CTCF-1. Ninguna de estas mutaciones fue similar a las obtenidas por secuenciación de los amplicones de menor tamaño de la población HD3 CRISPR1, las cuáles fueron de 35 pb y 38 pb y se llevan consigo al motivo CTCF-1.

La disminución en la cantidad del transcrito del gen α D observada en las poblaciones de células HD3 CRISPR1 y B3 diferenciadas sugiere que se afectó la unión de CTCF al promotor α D y que la unión de dicho factor es importante para la activación transcripcional del gen α D. Para comprobarlo es necesario realizar experimentos de inmunoprecipitación de la cromatina. Además, un aspecto interesante en los resultados obtenidos es que existe un decremento proporcional de la expresión de α D que parece estar relacionado con la cantidad de bases eliminadas del motivo CTCF-1. Es decir, en las células HD3 A3 que no eliminaron ninguna base del motivo de unión a CTCF predicho, sino bases adyacentes, los niveles de expresión de α D fueron similares a los encontrados en las células HD3 wt diferenciadas. En las células HD3 B3, donde se eliminaron 5 pb del motivo CTCF-1, el nivel de expresión de α D fue de ~60% respecto al calculado para las HD3 wt. Y en la población heterogénea HD3 CRISPR1, donde están presentes células con el motivo CTCF-1 completamente eliminado, la actividad transcripcional calculada para α D fue de 25% relativa a la de las células HD3 wt diferenciadas. Una posible interpretación de estos datos es que la secuencia consenso del motivo de unión a CTCF es necesaria y suficiente para el posicionamiento del factor nuclear CTCF. Por lo que sólo las mutaciones que ocurran sobre el motivo de unión, como en el caso de las células HD3 B3, pueden comprometer el correcto

posicionamiento de CTCF en alguna proporción de la población celular. Los experimentos de inmunoprecipitación de cromatina permitirán también dilucidar si esto es cierto.

Es importante recordar que el mecanismo propuesto para la función de CTCF en este *locus* consiste en la facilitación de la formación de una asa de cromatina con el 3'-*enh*; por lo tanto, la hipótesis que prevalece es que la eliminación del sitio de unión a CTCF en el promotor de αD de las células HD3 compromete el posicionamiento de CTCF en esta región al diferenciar las células HD3 mutantes. Consecuentemente, la formación del asa de cromatina con el 3'-*enh* se ve afectada, por lo que no se potencia la transcripción del gen adulto αD . Experimentos de captura conformacional de cromosomas (3C) permitirán comprobar si este es el mecanismo de acción.

La eliminación de este sitio de unión a CTCF pudiera tener repercusiones en la organización y regulación de la transcripción de todo el dominio α -globina de pollo. Esta idea surge porque: 1) CTCF se ha descrito como una proteína arquitectónica de la cromatina; y 2) existen datos de 3C en los que se observó que la región promotora del gen αD interactúa físicamente con otros elementos con posible función reguladora en el dominio desde etapas tempranas de la diferenciación eritroide (Gavrilov & Razin, 2008). Evaluar la actividad transcripcional del gen π y gen αA en las mutantes diferenciadas nos permitirá conocer si esto está ocurriendo, además de comparar si las interacciones con los otros elementos de regulación se mantienen o se pierden podría aportar mayor evidencia de la función de CTCF como organizador de la cromatina.

10. CONCLUSIONES

- CTCF se une *in vitro* al motivo de unión CTCF-1 en el promotor del gen adulto α D predicho por Valdes-Quezada *et al.* (2013) en el dominio α -globina de pollo.
- Hemos removido el sitio de unión a CTCF en el promotor del gen adulto α D en las células eritroblásticas arrestadas HD3 con el sistema de edición genómica CRISPR-Cas9.
- Eliminar el motivo de unión a CTCF (HD3 CRISPR1), o parte de éste (HD3 B3), en el promotor del gen α D de células HD3 ocasiona una disminución en la expresión del gen α D al diferenciarlas.
- Por lo tanto, la unión de CTCF en el promotor del gen α D es importante para la activación de dicho gen en las células HD3 diferenciadas y el mecanismo de acción podría ser a través de la formación de un asa de cromatina con el 3'-*enh*.

11. PERSPECTIVAS

- Comprobar que las eliminaciones obtenidas con el sistema CRISPR-Cas9 en la región promotora α D comprometen la unión de CTCF *in vitro*.
- Comprobar que disminuye el enriquecimiento de CTCF en el promotor de α D de las células editadas con CRISPR-Cas9.
- Demostrar si hay cambios en la conformación de la cromatina y en la configuración tridimensional del dominio α -globina en las células HD3 modificadas genéticamente.

12. BIBLIOGRAFÍA

- Ali T, Renkawitz R, Bartkuhn M. 2016. Insulators and domains of gene expression. *Curr Opin Genet Dev* **37**: 17–26.
- Allfrey VG, Mirsky AE. 1964. Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis. *Science* **144**: 559.
- Allis CD, Caparros ML, Jenuwein T, Reinberg D. 2015. Epigenetics. 2nd Ed. Cold Spring Harbor Laboratory Press, New York. 984 p
- Annunziato, A. (2008) DNA packaging: Nucleosomes and chromatin. *Nature Education* **1**, 26
- Barkess G, West AG. 2012. Chromatin insulator elements: establishing barriers to set heterochromatin boundaries. *Epigenomics* **4**: 67–80.
- Bell AC, West AG, Felsenfeld G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**: 387–96.
- Beug H, Doederlein G, Freudenstein C, Graf T. 1982. Erythroblast cell lines transformed by a temperature-sensitive mutant of avian erythroblastosis virus: a model system to study erythroid differentiation in vitro. *J Cell Physiol Suppl* **1**: 195–207.
- Beug H, von Kirchbach A, Döderlein G, Conscience JF, Graf T. 1979. Chicken hematopoietic cells transformed by seven strains of defective avian leukemia viruses display three distinct phenotypes of differentiation. *Cell* **18**: 375–90.
- Charpentier, E. & Doudna, J. A. (2013) Biotechnology: Rewriting a Genome. *Nature* **495**, 50 – 51
- Van Bortle K, Nichols MH, Li L, Ong C-TT, Takenaka N, Qin ZS, Corces VG. 2014. Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol* **15**: R82.
- Clark RJ, Felsenfeld G. 1971. Structure of chromatin. *Nature New Biol* **229**: 101–6.
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barreto, R., Habib, N., Hsu, P. D., We, X., Jiang, W., Marraffini, L. A. & Zhang, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819 – 823
- Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, *et al.* 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* **107**: 21931–6.
- Cuddapah S, Jothi R, Schones DE, Roh T-YY, Cui K, Zhao K. 2009. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* **19**: 24–32.

Della Ragione F, Vacca M, Fioriniello S, Pepe G, D'Esposito M. 2016. MECP2, a multi-talented modulator of chromatin architecture. *Brief Funct Genomics*.

Dickson J, Gowher H, Strogantsev R, Gaszner M, Hair A, Felsenfeld G, West AG. 2010. VEZF1 elements mediate protection from DNA methylation. *PLoS Genet* **6**: e1000804.

Dixon JR, Gorkin DU, Ren B. 2016. Chromatin Domains: The Unit of Chromosome Organization. *Mol Cell* **62**: 668–80.

Dorschner MO, Hawrylycz M, Humbert R, Wallace JC, Shafer A, Kawamoto J, Mack J, Hall R, Goldy J, Sabo PJ, *et al.* 2004. High-throughput localization of functional elements by quantitative chromatin profiling. *Nat Methods* **1**: 219–25.

Downen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, Weintraub AS, Schuijers J, Lee TI, Zhao K, *et al.* 2014. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**: 374–87.

Escamilla-Del-Arenal M, Recillas-Targa F. 2008. GATA-1 modulates the chromatin structure and activity of the chicken alpha-globin 3' enhancer. *Mol Cell Biol* **28**: 575–86.

Filippova GN. 2008. Genetics and epigenetics of the multifunctional protein CTCF. *Curr Top Dev Biol* **80**: 337–60.

Flavahan WA, Drier Y, Liao BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suvà ML, Bernstein BE. 2016. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**: 110–4.

Gavrilov AA, Razin SV. 2008. Spatial configuration of the chicken alpha-globin gene domain: immature and active chromatin hubs. *Nucleic Acids Res* **36**: 4629–40.

Ghirlando R, Felsenfeld G. 2016. CTCF: making the right connections. *Genes Dev* **30**: 881–91.

Graf T. 1975. In vitro transformation of chicken bone marrow cells with avian erythroblastosis virus. *Z Naturforsch, C, Biosci* **30**: 847–9.

Gómez-Díaz E, Corces VG. 2014. Architectural proteins: regulators of 3D genome organization in cell fate. *Trends Cell Biol* **24**: 703–11.

Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, *et al.* 2010. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**: 479–91.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, *et al.* 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**: 311–8.

Hnisz D, Weintraub AS, Day DS, Valton A-LL, Bak RO, Li CH, Goldmann J, Lajoie BR, Fan ZP, Sigova AA, *et al.* 2016. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**: 1454–8.

Hsu PD, Lander ES, Zhang F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**: 1262–78.

Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**: 1231–45.

Kornberg RD, Thomas JO. 1974. Chromatin structure; oligomers of the histones. *Science* **184**: 865–8.

Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, Shu J, Dadon D, Young RA, Jaenisch R. 2016. Editing DNA Methylation in the Mammalian Genome. *Cell* **167**: 233–247.e17.

Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**: 251–60.

Lupiáñez DGG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, *et al.* 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**: 1012–25.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E. & Church, G. M. 2013. RNA-guided human genome engineering via Cas9. *Science* **339**, 823 – 826

Martin D, Pantoja C, Fernández Miñán A, Valdes-Quezada C, Moltó E, Matesanz F, Bogdanović O, de la Calle-Mustienes E, Domínguez O, Taher L, *et al.* 2011. Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. *Nat Struct Mol Biol* **18**: 708–14.

Martin GR. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci USA* **78**: 7634–8.

Maston GA, Evans SK, Green MR. 2006. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* **7**: 29–59.

Maurano MT, Wang H, John S, Shafer A, Canfield T, Lee K, Stamatoyannopoulos JA. 2015. Role of DNA Methylation in Modulating Transcription Factor Occupancy. *Cell Rep* **12**: 1184–95.

Mojica, F. J., Díez-Villaseñor, C., Soria, E. & Juez, G. 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archea, Bacteria and mitochondria. *Mol. Microbiol.* **36**, 244 – 246

Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, Reinberg D. 2015. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**: 1017–21.

Nora EPP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, *et al.* 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**: 381–5.

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, *et al.* 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**: 499–502.

Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G. 2013. Enhancers: five essential questions. *Nat Rev Genet* **14**: 288–95.

Phillips-Cremins JE, Corces VG. 2013. Chromatin insulators: linking genome organization to cellular function. *Mol Cell* **50**: 461–74.

Raab JR, Kamakaka RT. 2010. Insulators and promoters: closer than we think. *Nat Rev Genet* **11**: 439–46.

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**: 2281–308.

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, *et al.* 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**: 1665–80.

Rasmussen KD, Helin K. 2016. Role of TET enzymes in DNA methylation, development, and cancer. *Genes Dev* **30**: 733–50.

Rincón-Arano H, Guerrero G, Valdes-Quezada C, Recillas-Targa F. 2009. Chicken alpha-globin switching depends on autonomous silencing of the embryonic pi globin gene by epigenetics mechanisms. *J Cell Biochem* **108**: 675–87.

Rincón-Arano H, Valadez-Graham V, Guerrero G, Escamilla-Del-Arenal M, Recillas-Targa F. 2005. YY1 and GATA-1 interaction modulate the chicken 3'-side alpha-globin enhancer activity. *J Mol Biol* **349**: 961–75.

Sanyal, A. Lahoie, B. R., Jain, G. & Dekker, J. 2012. The long-range interaction landscape of gene promoters. *Nature* **489**, 109 – 113

Splinter E, Heath H, Kooren J, Palstra R-JJ, Klous P, Grosveld F, Galjart N, de Laat W. 2006. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* **20**: 2349–54.

Srinivasan L, Atchison ML. 2004. YY1 DNA binding and PcG recruitment requires CtBP. *Genes Dev* **18**: 2596–601.

Strahl BD, Allis CD. 2000. The language of covalent histone modifications. *Nature* **403**: 41–5.

Targa FR, de Moura Gallo CV, Huesca M, Scherrer K, Marcaud L. 1993. Silencer and enhancer elements located at the 3'-side of the chicken and duck alpha-globin-encoding gene domains. *Gene* **129**: 229–37.

Valadez-Graham V, Razin SV, Recillas-Targa F. 2004. CTCF-dependent enhancer blockers at the upstream region of the chicken alpha-globin gene domain. *Nucleic Acids Res* **32**: 1354–62.

Valdes-Quezada C, Arriaga-Canon C, Fonseca-Guzmán Y, Guerrero G, Recillas-Targa F. 2013. CTCF demarcates chicken embryonic α -globin gene autonomous silencing and contributes to adult stage-specific gene expression. *Epigenetics* **8**: 827–38.

Van Bortle, K. & Corces, V. G. (2012) Nuclear organization and genome function. *Annu. Rev. Cell Dev. Biol.* **28**, 163-187

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, *et al.* 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**: 854–8.

West AG, Huang S, Gaszner M, Litt MD, Felsenfeld G. 2004. Recruitment of histone modifications by USF proteins at a vertebrate barrier element. *Mol Cell* **16**: 453–63.

Zhu J, Adli M, Zou JY, Verstappen G, Coyne M, Zhang X, Durham T, Miri M, Deshpande V, De Jager PL, *et al.* 2013. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152**: 642–54.