



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**DOCTORADO EN CIENCIAS BIOMÉDICAS**

**INSTITUTO DE FISIOLÓGIA CELULAR**

**REGULACION EPIGENETICA DEL GEN EMBRIONARIO PHI ALFA  
GLOBINA DURANTE EL DESARROLLO Y LA DIFERENCIACION  
ERITROIDE**

**TESIS**

**QUE PARA OPTAR POR EL GRADO DE:**

**DOCTOR EN CIENCIAS**

**PRESENTA:**

**GUSTAVO CHRISTIAN VALDES QUEZADA**

**TUTOR: DR. FÉLIX RECILLAS TARGA  
INSTITUTO DE FISIOLÓGIA CELULAR**

**MÉXICO, D.F. SEPTIEMBRE DE 2014**



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 INSTITUCIONALES

Este trabajo fue realizado en el Instituto de Fisiología Celular (IFC), Departamento de Genética Molecular, laboratorio 122-Norte de la Universidad Nacional Autónoma de México (UNAM) bajo la tutoría del Dr. Félix Recillas Targa quien recibió apoyos de DGAPA-UNAM (IN209403, IN203811 y IN201114) y del Consejo Nacional de Ciencia y Tecnología (CONACyT; 42653-Q , 128464 y 220503).

Los estudios de doctorado de Gustavo Christian Valdes Quezada fueron apoyados con una beca doctoral del CONACyT (207086), y por complementos económicos para la realización de la estancia de investigación y asistencia a congresos por parte de la Programa de Apoyos a los Estudios de Posgrado de la UNAM (PAEP-UNAM) y el Programa de Doctorado en Ciencias Biomédicas (PDCB-UNAM).

El Comité Tutorial que asesoró el desarrollo de esta tesis estuvo formado por:

Dr. Félix Recillas Targa	(IFC- UNAM)
Dra. María Imelda López Villaseñor	(IIB- UNAM)
Dr. Jesús Aguirre Linares	(IFC- UNAM)

La Bióloga Georgina Guerrero Avendaño brindó una excelente asesoría técnica durante el desarrollo de esta investigación.

El Dr. Fernando Suaste Olmos contribuyó con una excelente asesoría técnica durante el desarrollo del presente trabajo.

Los experimentos de citometría de flujo y la secuenciación de ADN se llevaron a cabo con la colaboración de la Unidad de Biología Molecular en el IFC-UNAM a cargo de la Dra. Laura Ongay.

Se agradece al Prof. Wouter de Laat por las facilidades para realizar la estancia de investigación de cuatro meses en el Hubrecht Institute en Utrecht, Holanda. Esta estancia fue posible gracias al Programa de Doctorado en Ciencias Biomédicas y el PAEP-UNAM.

Se agradece al Dr. Paul Delgado Olguín por la donación del anticuerpo anti-5hmC utilizado en los experimentos vinculados a la desmetilación del ADN.

El Jurado de Examen Doctoral estuvo constituido por:

Presidente:	Dra. Victoria Chagoya Hazas	(IFC-UNAM)
Secretario:	Dra. María Imelda López Villaseñor	(IIB-UNAM)
Vocal:	Dr. Horacio Merchant Larios	(IIB-UNAM)
Vocal:	Dr. Diego González Halphen	(IFC-UNAM)
Vocal:	Dr. David René Romero Camarena	(CCG-UNAM)

## AGRADECIMIENTOS

### **A mis padres,**

Gracias creer siempre en mí y por todo el apoyo incondicional y la confianza que en todos estos años me han brindado. Esta tesis está enteramente dedicada a ustedes y a toda la hermosa familia que siempre han sacado adelante. En verdad son lo máximo! Los quiero muchísimo.

### **A mi esposa,**

Marisol, mi panquecita tan hermosa, gracias por todo el impulso para salir siempre adelante. No encuentro las palabras para agradecer todo el apoyo y el amor que me has brindado en todo momento. En ti siempre encuentro las fuerzas para superar cualquier adversidad, pero sobre todo, en ti encuentro a mi Sol que me motiva y llena de felicidad mi camino. Tu empeño y dedicación han sido las mejores lecciones para salir adelante. Ahora sé que juntos podemos lograr todo lo que nos proponamos. Te amo todo y eres lo mejor que la vida me ha podido dar. Gracias..... Totales !!

### **A mis hermanos,**

Jazmín, Juan y Abraham, gracias por todo su apoyo y la paciencia que siempre me han tenido. En verdad son todo un ejemplo y orgullo para mí. Mil gracias, porque sé que siempre puedo contar con ustedes, los quiero mucho!

### **A mi creciente familia,**

Gracias Areli y Jorge por todo su apoyo y aguante, son unos grandes guerreros que admiro y quiero mucho. Gracias a todos los chamacos, Emilio, Esteban, Sara, Edahí e Iyari, que con toda su alegría fue posible superar los momentos más difíciles.

### **A Félix Recillas,**

Doc, Gracias por toda la confianza, pero sobre todo, muchas gracias por todo su apoyo y amistad. Me llena de mucho orgullo haber aprendido tanto de usted; aún recuerdo lo perdido que estaba cuando llegue al lab y ahora agradezco enormemente todo su apoyo para enderezar mi camino en la ciencia. Gracias por todas sus enseñanzas, las cuales me han mostrado que con esfuerzo y pasión, todo es posible.

### **A Georgina Guerrero,**

Geito, gracias por estar siempre apoyándome y ayudándome en todo momento para sacar todo adelante. Siempre he valorado mucho toda la experiencia que has compartido con todos en el laboratorio. De igual manera, quiero agradecerle no solo por todo el trabajo juntos, sino además, por esa bonita amistad que hemos forjado.

**Al Valdis,**

Jorge, gracias mi hermano, en verdad mil gracias. De no ser por todo tu apoyo, consejos y regaños, esto no hubiera sido posible. Siempre serás mi gran amigo a quien admiro, respeto y agradezco esta mistad súper chida.

A mis hermanos y grandes amigos: **Chucho, Jorge Paniagua, Javier y Cesar.** Les agradezco siempre estar allí, en las buenas y en las malas. Son todo un orgullo y tienen toda mi admiración, los adoro un chorro mis carnales!

A mis compañeros y amigos del laboratorio, a los viejos, contemporáneos y más nuevos: **Cristian (tocayo), Héctor, Inti, Martín Abrahan, Mayra, Eria, Ernesto, Pedro, Edgar, Paco, Charis, Erandi, Estela, Yael, Carla, Andrea, Mercedes, Gabriel, Ricardo, Mariana, Daniel, Rodrigo, David y Fernando.** Gracias a todos y cada uno de ustedes; sin sus consejos y ayuda esto no se hubiera concretado. En todas las etapas, han sido un grupo increíble del cual me he sentido muy orgullo de pertenecer. Muchas gracias por toda su amistad y por ser tan estupendas personas.

**A Dany Venegas,**

Chamaca, tengo mucho que agradecerte, todos tu consejos, todo tu apoyo, tus regaños y todas la porras que siempre con esa bonita sonrisa y cariño me brindaste para poder sacar todo adelante. Sabes que tú y toda tu familia siempre tendrán un lugar muy especial en mi corazón.

**A Gianelli Cortés,**

Giane, cómo podré agradecerte todo lo que me ayudaste y apoyaste en estos años. En verdad eres todo una trabajadora ejemplar!. Pero sobre todo, una gran amiga y persona. Muchas gracias por todo!

## INDICE

<b>I. Resumen</b>	
<b>II. Abstract</b>	
<b>III. Introducción.....</b>	<b>10</b>
1. El genoma eucarionte.....	11
2. Elementos genéticos de regulación.....	12
2.1. Los promotores de los genes	
2.2. Potenciadores o enhancers	
2.3. Las Regiones Control del Locus (LCRs)	
2.4. <i>Insulators</i> o delimitadores	
2.5. Elementos de regulación	
3. Regulación epigenética.....	23
3.1. La estructura de la cromatina	
3.2. La metilación del ADN	
3.3. Represión génica y metilación del ADN	
3.4. Metilación del ADN y su distribución en el genoma	
3.5. Metilación del ADN y cáncer	
3.6. Desmetilación del ADN	
4. Un código epigenético en las histonas.....	38
4.1. La acetilación de histonas	
4.2. La desacetilación de histonas	
4.3. La metilación de histonas	
4.4. Cromatina bivalente	
4.5. Resumen: epigenética y cromatina	
5. El factor multi-funcional CTCF y la regulación génica.....	50
5.1. CTCF y su papel en los dominios génicos	
5.2. CTCF y la metilación del ADN	
5.3. Resumen de las funciones del factor nuclear CTCF	
6. La regulación transcripcional en el espacio nuclear.....	60
6.1. Territorios cromosómicos	
6.2. Nuevas tecnologías para la visualización del arreglo 3D del genoma	

6.3.	Resumen de la dinámica nuclear	
7.	Los dominios de genes globina como modelos de estudio.....	71
7.1.	Regulación epigenética de los genes globina	
7.2.	Competencia por el enhancer y silenciamiento autónomo	
8.	Antecedentes particulares.....	78
8.1.	El dominio $\alpha$ -globina de pollo	
8.2.	La estructura de la cromatina y la regulación de los genes $\alpha$ -globina de pollo.	
8.3.	El enhancer 3' del dominio $\alpha$ -globina	
8.4.	Silenciamiento autónomo del gen embrionario $\pi$ $\alpha$ -globina de pollo.	
8.5.	Silenciamiento autónomo del gen embrionario $\pi$ y metilación del ADN.	
8.6.	Resumen de la regulación epigenética de los genes $\alpha$ -globina de pollo.	
<b>IV.</b>	<b>Planteamiento del problema</b> .....	<b>93</b>
<b>V.</b>	<b>Hipótesis</b> .....	<b>94</b>
<b>VI.</b>	<b>Objetivo General</b> .....	<b>94</b>
<b>VII.</b>	<b>Objetivos Particulares</b> .....	<b>95</b>
<b>VIII.</b>	<b>Estrategia experimental</b> .....	<b>96</b>
	1. Linajes eritroides primitivo y definitivo en el desarrollo del pollo.	
	2. Diferenciación eritroide definitiva en el pollo.	
<b>IX.</b>	<b>Material y Métodos</b> .....	<b>99</b>
<b>X.</b>	<b>Resultados</b> .....	<b>116</b>
	1. Reclutamiento de proteínas de unión a ADN metilado en el silenciamiento autónomo del gen embrionario $\pi$ .	
	2. Distribución de la metilación del ADN en la región genómica del gen embrionario $\pi$ .	

3. Estructura de la cromatina en la regulación de la expresión diferencial de los genes  $\alpha$ -globina.
4. Papel funcional de CTCF en el promotor del gen  $\alpha^D$ .
5. Organización espacial de los genes  $\alpha$ -globina y el elemento 3'-Enh a través del desarrollo del pollo.

<b>XI.</b>	<b>Discusiones.....</b>	<b>177</b>
	<ol style="list-style-type: none"> <li>1. El papel de la metilación del ADN en el silenciamiento del gen embrionario <math>\pi</math>.</li> <li>2. Desmetilación activa del ADN en el dominio <math>\alpha</math>-globina</li> <li>3. Delimitación de la metilación del ADN mediada por el factor CTCF.</li> <li>4. Participación del elemento 3'-Enh en la expresión diferencial de los genes <math>\alpha</math>-globina.</li> <li>5. Propuesta: Modelo de regulación de los genes <math>\alpha</math>-globina de pollo.</li> </ol>	
<b>XII.</b>	<b>Conclusiones.....</b>	<b>192</b>
<b>XIII.</b>	<b>Perspectivas.....</b>	<b>194</b>
<b>XIV.</b>	<b>Referencias.....</b>	<b>197</b>
<b>XV.</b>	<b>Publicaciones relacionadas al proyecto principal de tesis.....</b>	<b>210</b>
<b>XVI.</b>	<b>Artículo principal derivado del presente trabajo de investigación....</b>	<b>212</b>
<b>XVII.</b>	<b>Carátulas de los artículos publicados relacionados con el presente trabajo de investigación.....</b>	<b>224</b>

## I. Resumen

En el genoma de los eucariontes se encuentran codificados los genes necesarios para la ejecución de los diversos procesos biológicos. A través de la evolución, el genoma ha adquirido elementos de ADN que participan en la regulación de los programas de expresión de los genes. En el presente trabajo, se ha adoptado al dominio  $\alpha$ -globina del pollo como paradigma para comprender la participación de los elementos de regulación y el papel de la epigenética en la expresión de los genes  $\alpha$ -globina. El grupo de genes  $\alpha$ -globina presenta un patrón de expresión diferencial a través del desarrollo del pollo, donde el gen embrionario  $\pi$  es activo a la transcripción en las etapas tempranas y por su parte los genes adultos  $\alpha^D$  y  $\alpha^A$  presentan actividad en las etapas avanzadas. En el presente trabajo, se ha determinado que la expresión del gen embrionario  $\pi$ , en las etapas tempranas del desarrollo, involucra la interacción física con el elemento *enhancer* del dominio (3'-Enh) mediante la formación de un *loop* o asas de cromatina. En las etapas avanzadas del desarrollo, el gen embrionario  $\pi$  es silenciado en su transcripción mediante el establecimiento de un sub-dominio de metilación con la participación de la proteína represora de unión a ADN metilado, MeCP2. Como consecuencia de este sub-dominio de metilación y represión génica, el gen embrionario no presenta interacción física con el elemento 3'-Enh. Para contener contra la expansión del sub-dominio de metilación, el factor multifuncional CTCF es reclutado en la región del promotor  $\alpha^D$ , delimitando el sub-dominio de metilación y favoreciendo la correcta expresión de los genes adultos  $\alpha^D$  y  $\alpha^A$ . En células eritroblásticas inducidas a la expresión de los genes adultos, hemos encontrado una desmetilación activa del ADN asociada a la presencia de 5-hidroxi-metilcitocina y coincidente con la unión del factor CTCF en la región del promotor  $\alpha^D$ . Por tal motivo, proponemos que la presencia del factor CTCF es necesaria para el establecimiento de un estado de la cromatina abierto y permisible a la transcripción. En su conjunto estos resultados nos han permitido proponer un modelo de regulación de los genes  $\alpha$ -globina el cual involucra la participación de los elementos del dominio y la regulación epigenética.

## II. Abstract

Encoded in the eukaryotic genome, the genes are involved in the regulated execution of the diverse biological processes. Throughout the evolution, the genome has acquired DNA elements implicated in the regulation of the gene-expression programs. In the present work, we adopted the chicken  $\alpha$ -globin domain as a biological model system to understand the role of the gene regulation elements and the epigenetics in the  $\alpha$ -globin gene expression. The chicken  $\alpha$ -globin genes are expressed in a differential manner throughout the chicken development; the embryonic  $\pi$  gene is transcriptionally active in the early stages in turn the adult genes,  $\alpha^D$  and  $\alpha^A$ , are active in the late stages. In the present work, we have determined that the embryonic  $\pi$  gene expression involves a physical interaction with the domain *enhancer* (3'-Enh) by forming a chromatin *loop*. In late developmental stages, the embryonic  $\pi$  gene is transcriptionally silenced by the establishment of a DNA methylation sub-domain with the action of the DNA-methylated-binding protein, MeCP2. As a consequence of this DNA methylation and gene silencing, the embryonic  $\pi$  gene loses its physical interaction with the 3'-Enh element. In order to protect against the DNA methylation sub-domain expansion, the multifunctional factor CTCF is recruited at the  $\alpha^D$  promoter region, demarcating the DNA methylation sub-domain and favoring the appropriate expression of the adult genes. In an erythroblast cell context induced to the adult gene expression, we found the presence of the 5-hydroxy-methylcytosine at the  $\alpha^D$  promoter region associated with an active DNA demethylation in coincidence with the binding of CTCF. Therefore, we propose that the presence of the CTCF factor is needed for the establishment of an optimal chromatin conformation that favors regulated gene transcription. All together, these results enable us to propose a model for the  $\alpha$ -globin gene regulation on which the role of the domain DNA elements and the epigenetic regulation are discussed.

### **III. Introducción**

La comprensión de los procesos y mecanismos moleculares que constituyen a los seres vivos ha sido uno de los motores principales de las ciencias biológicas modernas. Gracias a los diversos avances tecnológicos ha sido posible el 'parcial' entendimiento de la vida inclusive a un nivel molecular, permitiendo la resolución de varias preguntas pero sobre todo y de manera altamente fascinante: la generación de nuevos cuestionamientos y paradigmas que nos hablan de la complejidad de los sistemas vivos.

Dentro de la biología se han desprendido diversas ramas, entre ellas la genética, y más recientemente la epigenética la cual se ha focalizado en la comprensión de los diversos procesos involucrados en la variedad y complejidad de los organismos, su relación con el medio ambiente así como en la perpetuación de su especie. En el presente trabajo de tesis se ha decidido contribuir en el conocimiento de los procesos que permiten la generación de tal complejidad en los seres vivos, en particular, en el establecimiento de diferentes programas de expresión génica. Para ello, se abordaron estudios de tipo genético y epigenético aplicados a un modelo de estudio que comprende la expresión diferencial de los genes codificantes para la síntesis de distintas hemoglobinas durante el desarrollo del organismo. Por tal motivo, es preciso detallar algunos aspectos introductorios al estudio de la expresión génica desde una perspectiva vinculada a la regulación genética y epigenética.

## 1. El genoma eucarionte

La información genética de los organismos se encuentra codificada en la molécula de ADN que constituye el genoma de cada especie. Contenidos en el genoma se encuentran los genes que codifican para las proteínas encargadas de los diferentes procesos y estructuras biológicas. Para 'cumplir' sus funciones, los genes son expresados mediante su transcripción a moléculas de ARN la cual puede, dependiendo de cada gen, ser traducida a proteína o bien cumplir alguna otra función catalítica, regulatoria, o estructural permaneciendo como molécula de ARN<sup>1</sup>.

Además de los genes, en el genoma se incluyen secuencias no-codificantes que corresponden con regiones genómicas las cuales participan en la regulación de la expresión de los genes. Estas fueron denominadas elementos genéticos de regulación. Por otra parte, en el genoma eucarionte la mayor parte del genoma está constituido por secuencias repetidas cuya función aun se desconoce<sup>2</sup>. A diferencia de estas secuencias repetidas, tanto los genes como los elementos de regulación, constan de secuencias únicas no repetidas.

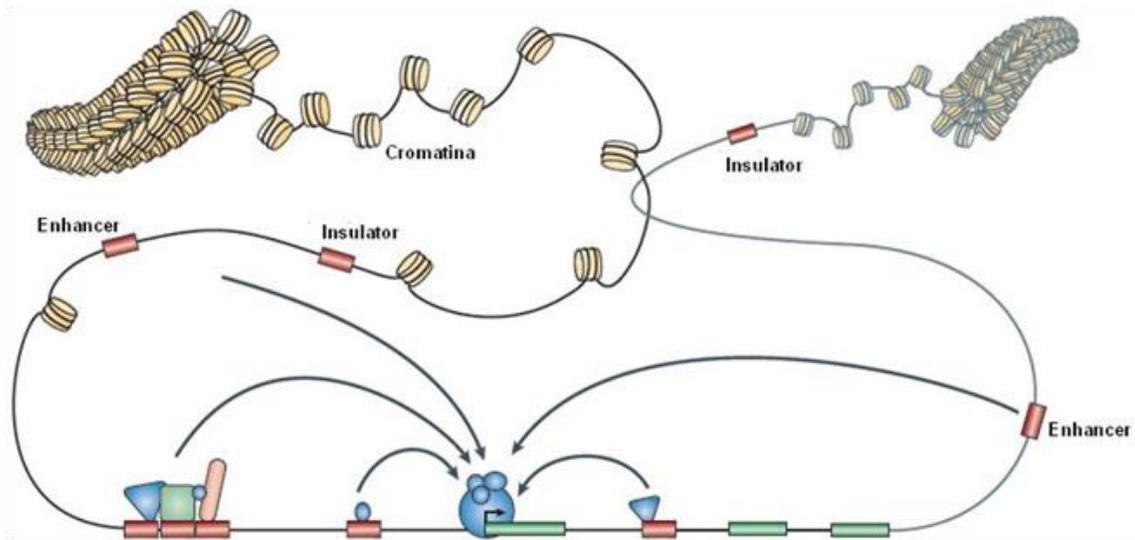
Como una consecuencia a tal diversidad de elementos contenidos, el genoma eucarionte ha adquirido un mayor tamaño (nucleótidos) a diferencia de los genomas procariontes. Por tal motivo, los eucariontes desarrollaron la capacidad de empaquetar un genoma de gran tamaño en el núcleo de cada célula<sup>1</sup>. Dicho empaquetamiento se logra mediante la asociación del ADN con proteínas llamadas histonas constituyendo así, junto con otras proteínas, a la

cromatina. Otra propiedad de este empaquetamiento en cromatina es la de regular la expresión génica<sup>3</sup>. En particular, la regulación de la estructura de la cromatina y su relación con la expresión de los genes es una de las áreas de estudio de la epigenética.

## **2. Elementos genéticos de regulación**

La manera en que la porción no-codificante del genoma puede ejercer un efecto sobre la expresión de los genes se ha descrito a partir de la caracterización de diversos elementos genéticos de regulación. Estos consisten de secuencias de ADN a las cuales se unen factores de transcripción con la finalidad de activar o reprimir la expresión de genes en cada tipo celular. En eucariontes, la transcripción de genes que codifican para proteínas se lleva a cabo por la enzima ARN polimerasa II (ARN pol II). Los elementos genéticos pueden actuar coordinando la acción de la ARN pol II en las diferentes fases de transcripción de los genes. Estos elementos pueden ser englobados en dos familias: *a) los promotores de los genes y b) los elementos distales de regulación (enhancers, silenciadores, LCRs e insulators).*

Así mismo, en la transcripción de los genes tanto los elementos genéticos como los epigenéticos actúan de manera coordinada para regular este proceso. La estructura de la cromatina presente en estos elementos puede influir directamente en los patrones de expresión, y más aún, conferir memoria epigenética relevante en la diferenciación celular y en el desarrollo de los organismos<sup>4</sup> (Figura 1).



**Figura 1.** Empaquetamiento del genoma eucarionte en cromatina. La acción coordinada entre los elementos de regulación, los factores de transcripción (óvalos azules) y la manera en que el ADN de la célula se encuentra organizado en cromatina tiene como finalidad la expresión regulada de los genes (cajas verdes). Tomada y modificada de <sup>4</sup>.

Para precisar la manera en que estos elementos funcionan como reguladores de la expresión génica en los eucariontes, a continuación se describen de manera general sus características y funciones.

## 2.1. Los promotores de los genes

Un promotor se define como una secuencia de ADN que es suficiente para dirigir la correcta iniciación de la transcripción de un gen mediada por la ARN polimerasa. Así mismo, se ha demostrado como la secuencia río arriba del sitio de inicio a la transcripción contribuye al posicionamiento del Complejo de Inicio de la Transcripción (CIT), el cual favorece y facilita el reclutamiento de la ARN

polimerasas. En general un promotor consiste de un conjunto de motivos de unión (secuencias de ADN) a factores nucleares conocidos como factores de transcripción<sup>5</sup>.

En cuanto a su estructura, los promotores se encuentran constituidos por dos elementos: *a) el elemento central* o 'core', el cual sirve como sitio de reconocimiento para la maquinaria basal de la transcripción y el sitio de ensamblaje del CIT. Además, define el sitio de inicio de la transcripción y la direccionalidad de la misma; *b) el elemento proximal*, el cual está río arriba del 'core' del promotor (a distancias variables en pares de bases), cuya función es la de atraer combinatorias de factores de transcripción tanto generales como tejido-específicos y además determina el tiempo y espacio en el cual un gen debe ser transcrito<sup>6</sup>.

En términos de su composición a nivel de secuencia de ADN, se puede mencionar que además de contener las secuencias motivo para la unión de los factores de transcripción, el 70% de los promotores de los vertebrados se encuentran asociados a regiones ricas en el di-nucleótido C-G (CpG) conocidas como islas CpGs<sup>7</sup>. Estas islas pueden ser categorizadas en islas pobres, intermedias o altas, dependiendo de la abundancia de dinucleótidos CpGs. Los promotores con islas CpG-altas están vinculadas con múltiples sitios de inicio a la transcripción (TSSs por sus siglas en inglés), controlando genes de expresión amplia o constitutiva y genes que se expresan en distintas etapas del desarrollo de un organismo. Por su parte, los promotores asociados a islas CpG-intermedias y CpG-bajas, se asocian con la regulación de genes tejido específicos<sup>8</sup>. En la

Tabla 1, se presenta una clasificación en al menos 4 tipos de promotores entre mamíferos e insectos propuesta por Saxonov y colaboradores <sup>7</sup>.

Tabla 1		Tipos de promotores (tomada y modificada de <sup>7</sup> )	
Tipo de promotor	Función del gen	Propiedades comunes	Composición de CpGs
Tipo I (adulto)	Expresión tejido-específica en tejidos periféricos adultos	TSS bien definido, enriquecimiento de caja TATA. Nucleosomas aleatoriamente posicionados	Principalmente no son islas CpG
Tipo II (ubíquos)	Expresión amplia	TSS dispersos, configuración ordenada o regular de nucleosomas	Islas CpG carentes de cajas TATA
Tipo III (regulación en el desarrollo)	Genes de expresión diferencialmente regulada, a menudo regulados en el desarrollo multicelular y la diferenciación.	Represión mediada por Polycomb, enriquecimiento en la marca de histona H3K27me3	Islas CpG que se extienden hacia el 'cuerpo' del gen

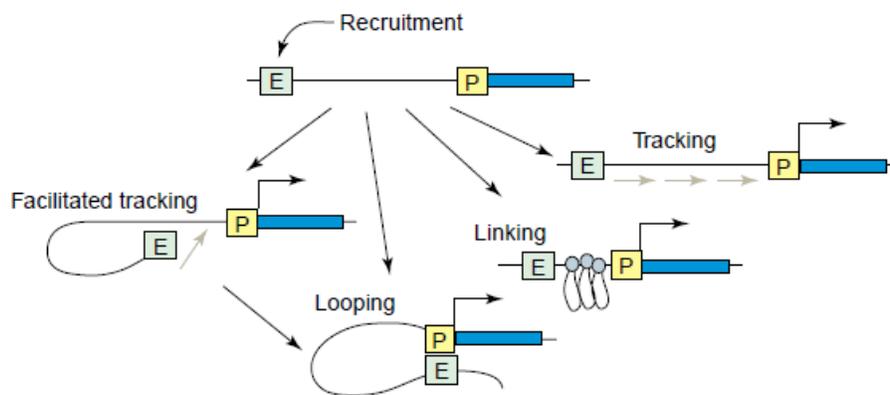
En términos de la cromatina asociada a estos elementos de regulación, se han observado diferentes propiedades que regulan la acción de los promotores a distintos niveles. Así mismo, la presencia de metilación en los dinucleótidos CpGs mayoritariamente presente en los vertebrados, confiere una característica de regulación y conservación de los promotores. Estos aspectos serán abordados en la sección de epigenética y regulación génica del presente trabajo.

## 2.2. Potenciadores o *enhancers*

Como se ha mencionado, cada gen para su transcripción depende de un promotor, sin embargo, existen un gran número de genes cuya regulación necesita ser controlada por uno o más elementos de regulación. Los elementos potenciadores o *enhancers*, constan de una serie de sitios de unión para factores de transcripción que contribuyen a la expresión regulada de un gen o genes blancos en tiempo y espacio. De manera general, se ha establecido que la presencia de estos *enhancers* proporciona a los genes una regulación específica de acuerdo al tejido y al momento del desarrollo del organismo en donde se deben de expresar<sup>9</sup>. En ensayos de transgénesis al incluir este tipo de secuencias *enhancer*, se ha observado que potencian o aumentan la actividad transcripcional del transgen<sup>10</sup>. Por ello, los *enhancers* además de darle especificidad a la expresión de los genes, también se consideran elementos necesarios para alcanzar los niveles adecuados de expresión de los genes.

El tipo de secuencias contenidas en los *enhancers* constan de motivos de unión a factores de transcripción tanto ubicuos como tejido específico. Básicamente, la función que desempeñan los factores de transcripción que reconocen estas secuencias es la de reclutar elementos activadores y/o co-activadores de la transcripción los cuales son 'transferidos' a la región promotora de su gen o genes blanco(s). Consecuentemente, los *enhancers* pueden encontrarse a 'grandes' distancias con respecto al promotor de su gen blanco, ya sea río arriba o río abajo en relación con el cuerpo del gen<sup>6</sup>. Se han propuesto diversos modelos para explicar la manera de actuar de un *enhancer* convergiendo

en la necesidad de una “comunicación” entre el *enhancer* y el promotor. Esta “comunicación” se logra a partir de diversos mecanismos entre los cuales, el modelo más favorecido es el de *looping* o de *formación de asas*, el cual muestra interacciones proteína-proteína formando asas de cromatina que facilitan el contacto físico entre estos dos elementos distales (Figura 2)<sup>11-13</sup>. En el *modelo de escaneo* (tracking), las proteínas que se reclutan en el *enhancer* se deslizan a lo largo de la cromatina hasta ‘encontrar’ al promotor. El *modelo de relevo* (*linking*), plantea que los factores transcripcionales se unen consecutivamente favoreciendo las interacciones a distancia entre el *enhancer* y el promotor<sup>5,11,14</sup>. Estos modelos se ejemplifican esquemáticamente en la Figura 2 y evidencias experimentales muestran que no son excluyentes.



**Figura 2.** Modelos de acción de un *enhancer*. Los modelos actuales se basan en la transferencia de los activadores y co-activadores reclutados en el *enhancer* (E) hacia el promotor (P). El cuadro azul representa el gen, tomado de<sup>14</sup>.

A nivel de la cromatina, los *enhancers* co-localizan con sitios de hipersensibilidad al corte por la endonucleasa DNasa I (DHS, por sus siglas en inglés), lo cual quiere decir que poseen una estructura de la cromatina relajada y

por ello se facilita, de manera regulada, el reconocimiento de los factores de transcripción a sus secuencias blanco en el ADN. Recientemente, se ha descrito que los *enhancers* poseen particularidades asociadas a su propia estructura de la cromatina. Inclusive un “*chromatin signature*” o firma en su estructura de la cromatina que puede ser característico a los *enhancers*<sup>15,16</sup>. Se ha demostrado que la actividad de los *enhancers* puede ser regulada a través de ARNs no codificantes denominados eARNs<sup>17,18</sup>. Este aspecto se aborda en la sección de epigenética y regulación génica.

### **2.3. Las Regiones Control del Locus (LCRs)**

Se ha observado que familias de genes (locus) que presentan un patrón ordenado de expresión durante el desarrollo o diferenciación a menudo se encuentran regulados por elementos de regulación que constan de varios *enhancers* que en su conjunto se denominan como regiones control del locus (LCRs, por sus siglas en inglés). A diferencia de los *enhancers* los cuales constan de un solo DHS, los LCRs pueden presentar varios DHS. En general su función es muy similar a la de un *enhancer*, es decir, potencian la transcripción de los genes de manera tejido-específica<sup>14</sup>. Sin embargo, los LCRs tiene la capacidad de impartir a los transgenes un patrón de expresión dependiente del número de copias e independiente del sitio de integración del transgen en el genoma<sup>10</sup>. Por lo tanto, los LCRs confieren un estado dominante de cromatina abierta en sus sitios de integración en el genoma<sup>19</sup>.

El entendimiento de los mecanismos de regulación derivados de la acción de los LCRs se ha logrado principalmente mediante su estudio en los dominios genes  $\alpha$  y  $\beta$ -globina en humano, ratón y pollo. El LCR de los genes  $\beta$ -globina está localizado de entre 6 a 22 kilo bases (kb) río arriba del primer gen globina del locus  $\beta$ -globina de humano. Está formado por un grupo de DHSs eritroide-específicos por el tipo de factores transcripcionales que se le asocian<sup>19</sup>. Los genes  $\beta$ -globina dependen en parte de la acción a distancia del LCR para poder alcanzar los niveles más altos de expresión<sup>20</sup>. Cuando este no está presente, se ha reportado en ratones transgénicos que el nivel de expresión del gen  $\beta$ -globina humano es cercana al 1%, comparado con los niveles endógenos<sup>19,21</sup>. La relevancia de este tipo de elementos de regulación se demuestra a partir de la remoción de este elemento, dado que no solo afecta dramáticamente los niveles de expresión de los genes globina, sino también la estructura de la cromatina y el tiempo de duplicación a lo largo de todo el locus en eritrocitos del humano<sup>19</sup>.

#### **2.4. *Insulators* o delimitadores**

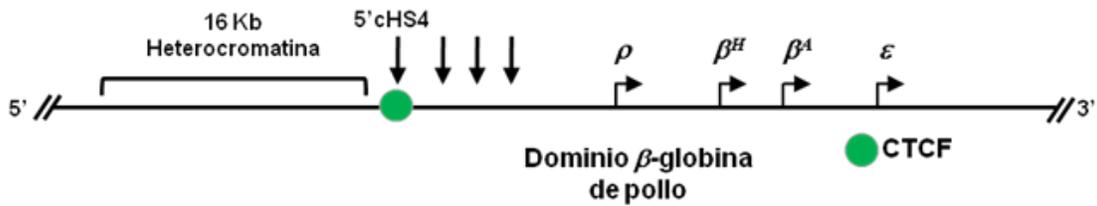
En el genoma eucarionte existen genes cuya expresión depende de momentos específicos del desarrollo o en un tipo de tejido en particular. En particular, estos genes están organizados en dominios génicos en los cuales se encuentran los elementos de regulación como lo son los promotores, potenciadores y los LCR, constituyendo unidades transcripcionalmente independientes. La separación espacial de distintos grupos de genes es uno de los posibles mecanismos que

promueven la formación de dominios funcionales. Con base a esta primicia, se han identificado secuencias de ADN que delimitan estos dominios. Estas secuencias se conocen como delimitadores o *insulators*. Los delimitadores o *insulators*, son secuencias que co-localizan con sitios de hipersensibilidad en el límite de los dominios génicos, donde comúnmente se define una clara transición entre estados de la cromatina “cerrada” y “abierta”. Los *insulators* están constituidos por secuencias neutras en cuanto a su actividad transcripcional, y de manera importante protegen a un dominio del efecto de la cromatina adyacente, previniendo la “promiscuidad”, es decir el efecto inespecífico de elementos de regulación entre distintos dominios génicos<sup>22</sup> (Figura 3). Los *insulators* que enmarcan a los dominios mantienen un estado específico de la cromatina del dominio, un estado “abierto” que favorece la expresión de los genes contenidos en él<sup>23</sup>. Sin embargo, el contexto de cromatina de los dominios génicos coincide también con cromatina “cerrada” o represora<sup>24</sup>.

En vertebrados el primer *insulator* descrito se localiza en el extremo 5' del locus  $\beta$ -globina de pollo<sup>25</sup>. Este elemento, se caracterizó como un sitio constitutivo de hipersensibilidad a la DNasa I (conocido como cHS4; Figura 3). El elemento cHS4 posee actividad de bloqueo de *enhancer* cuando este se sitúa entre un *enhancer* y el promotor. También se sabe que protege contra el efecto de posición causado por el entorno cromatínico en diferentes sitios de integración en el genoma presente en distintos organismos y tipos celulares<sup>26</sup>. Una de las principales funciones del cHS4 es evitar que elementos externos ejerzan una influencia en la expresión de los genes del dominio. La actividad funcional del

*insulator* cHS4 pudo aislarse a un fragmento de ADN de 250 pb, el cual se determinó como el “core” del *insulator*. Felsenfeld y colaboradores (1999) demostraron que 42 pb eran suficientes para mantener la función de bloqueo de *enhancer* con respecto a un promotor. En este elemento se demostró que se une la proteína CTCF<sup>27,28</sup>. CTCF es un factor nuclear altamente conservado en vertebrados; es una proteína cuyo peso molecular teórico es de 82 kDa compuesta por 11 dedos de zinc<sup>29,30</sup>. Se ha comprobado que CTCF está involucrado en la activación y silenciamiento de genes, dependiendo del contexto en que se encuentre, además de muchas otras funciones. El factor CTCF se ha identificado en todos los *insulators* de vertebrados conocidos a la fecha<sup>22,26,27</sup>.

Otra particularidad de esta proteína es la de promover la estructuración de la cromatina en asas; se ha demostrado que la unión de CTCF en el dominio  $\beta$ -globina de ratón promueve la formación de asas entre el LCR del dominio y cada uno de los promotores  $\beta$ -globina dependiendo de la etapa en la cual la expresión de cada gen es necesaria<sup>23,31</sup>. Cabe mencionar que CTCF no solo es considerada como una proteína *insulator*, por tal motivo aspectos más específicos de su función en la regulación génica serán abordados en la sección de cromatina, así como en los resultados derivados de la presente investigación.



**Figura 3.** Esquema del dominio  $\beta$ -globina de pollo. El *insulator* del dominio (5'cHS4) se localiza río arriba de los genes globina y río debajo de una región genómica de 16 kb de secuencias repetidas con características de heterocromatina constitutiva. En este elemento *insulator* existe un sitio de unión para el factor CTCF el cual puede contender contra la expansión de las 16 kb de heterocromatina río arriba de este. . Las flechas verticales representan sitios de DHS.

## 2.5. Resumen elementos genéticos de regulación

A lo largo de la evolución, el genoma de los eucariontes adquirió la capacidad de regular una gran variedad de genes. Mediante la existencia de elementos genéticos de regulación tal como los *enhancers*, LCRs e *insulators* la expresión génica puede ser coordinada para su correcta ejecución en tiempo y espacio. De manera general, estos elementos están constituidos por secuencias que incluyen motivos de unión para factores transcripcionales tanto ubicuos como tejido específicos. De tal manera que a los elementos de regulación pueden ser atraídos activadores, represores y otros reguladores de la transcripción génica. Sin embargo, la necesidad de empaquetar la gran cantidad de genoma en cromatina al interior del núcleo celular, genera un entorno en el cual el genoma es compactado a múltiples niveles, dificultando el reconocimiento de las secuencias blanco por parte de los factores de transcripción. Ante tal escenario, la necesidad de modular la estructura de la cromatina constituye otro nivel de regulación de la

expresión génica. Por tal motivo, a continuación se describen, de manera general, los mecanismos asociados a la estructura de la cromatina, su regulación y su efecto en la expresión de los genes.

### **3. Regulación epigenética**

La manera en que los genes son regulados en tiempo y espacio definen diferentes programas de expresión con los cuales se obtienen los diversos tipos y estructuras celulares de los organismos eucariontes a partir de un mismo genoma. En el ser humano se han descrito alrededor de 200 diferentes tipos celulares que van desde el momento de la fecundación hasta la constitución de una persona adulta. La manera por cual se puede obtener todas estas diferencias obedece a los diversos programas de regulación específicos y la expresión de genes característicos de cada tipo celular<sup>32</sup>.

Con la finalidad de que los genes puedan ser transcritos o reprimidos en un tipo y momento celular dado, la cromatina puede relajarse (activación) o compactarse (represión). Como una consecuencia de esto, el genoma de cada tipo celular estará estructurado en un entorno cromatínico distinto y específico a cada tipo celular, y esto debido a que los programas de expresión génica difieren entre cada una de ellas. De tal manera que la cromatina participa en el proceso de conferir identidad a cada tipo celular, una característica que es además heredable a cada célula durante la mitosis y meiosis. A lo largo de la diferenciación celular y del desarrollo es cuando estos distintos patrones de

regulación que llevan a cambios en la estructura de la cromatina y expresión génica se establecen. Sin embargo, tanto las células madres como los tejidos celulares derivados de estas, poseen en todo momento el mismo genoma en términos de la secuencia del ADN. Derivado de estas observaciones, históricamente se acuñó el término de epigenética a partir de la siguiente definición:

*La epigenética se refiere al estudio de todos aquellos factores no genéticos que intervienen en el desarrollo de un organismo, desde el óvulo fertilizado hasta su senescencia, pasando por la forma adulta (Waddington 1953<sup>33</sup>).*

En la actualidad, gracias a una mejor comprensión del papel que ejerce la cromatina con respecto a la expresión génica en cada tipo celular se ha redefinido el concepto de epigenética:

*Se consideran procesos de regulación epigenéticos todos aquellos que influyen en la expresión heredable de los genes sin que ocurran cambios en la secuencia del ADN<sup>33</sup>.*

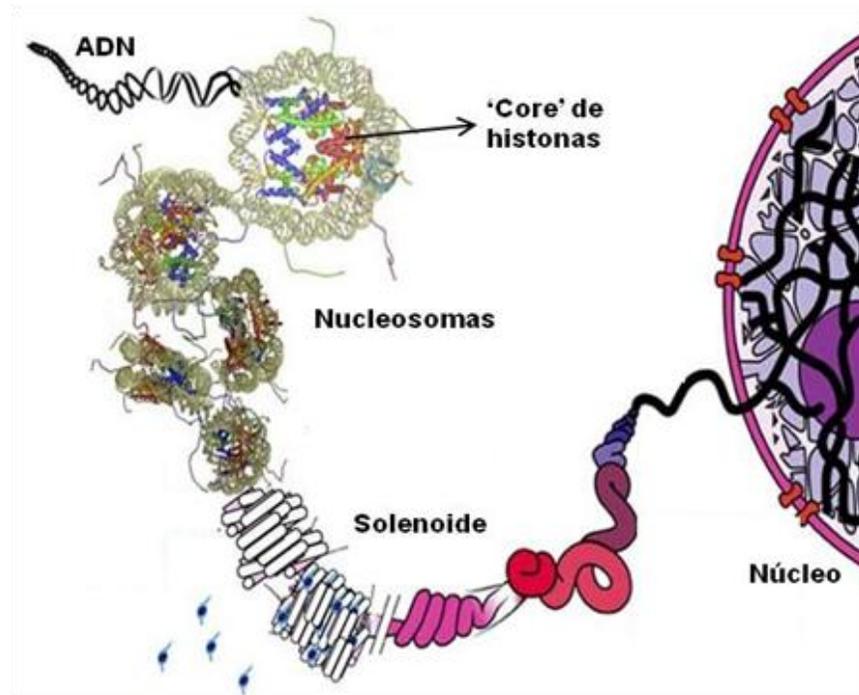
Se pueden distinguir varios procesos involucrados en la regulación epigenética : 1) la metilación del ADN, 2) las modificaciones post-traduccionales de las histonas, 3) los complejos remodeladores de la cromatina ATP-dependientes, 4) los complejos de proteínas represoras Polycomb y activadoras Trithorax, 5) mecanismos de regulación mediados por ARNs no-codificantes<sup>5</sup> y más recientemente el ordenamiento espacial (3D) del genoma al interior del

núcleo<sup>34</sup>. En su conjunto o en combinación de ellos, estos procesos modulan la estructura de la cromatina permitiendo el establecimiento de los diferentes programas de expresión génica que dan identidad a cada tipo celular. Por tal motivo, algunos aspectos generales de estos se abordan a continuación.

### **3.1. La estructura de la cromatina**

Como se ha mencionado, para su almacenamiento al interior del núcleo el ADN se asocia con un octámero de proteínas llamadas histonas y así formar la unidad mínima de la cromatina conocida como nucleosoma, y que en su nivel basal de organización se conoce como fibra de 10 nanómetros de diámetro. El octámero de histonas (“core” del nucleosoma) está constituido por dos copias de cada histona las cuales son: H2, H3, H2A y H2B e involucra la asociación de 146 pares de bases de ADN al “core” del nucleosoma<sup>35</sup> (Figura 4). Con subsecuentes niveles de organización de la cromatina confiere un grado de compactación del genoma de 5 a 10 veces, lo cual es necesario para contener los aproximadamente 2 metros de longitud que el genoma humano posee. Sin embargo, esta compactación resulta insuficiente para empaquetar al genoma en los 30 a 50 micrómetros de diámetro que en promedio posee el núcleo celular. Para lograr un mayor grado de compactación, se ha visto involucrada la asociación de una quinta proteína, la histona H1 (histona “linker”) a la región internucleosomal para constituir la estructura cromatínica conocida como solenoide o fibra de 30 nm (Figura 4). Esto otorga un nivel de compactación de 50 veces o

más del genoma, un mayor grado de estructuración se logra a partir arreglos de esta fibra de solenoide en estructuras de mayor orden hasta constituir, por ejemplo, los cromosomas metafásicos, un proceso muy poco comprendido hasta ahora<sup>36</sup>.



**Figura 4.** Empaquetamiento del genoma eucarionte al interior del núcleo celular. Se muestran los distintos niveles de compactación del genoma en asociación con las histonas.

Como puede suponerse, todos estos niveles de empaquetamiento del genoma eucarionte confieren *per se* un impedimento físico para la expresión de los genes. Dicho impedimento interfiere directamente con el proceso de transcripción de los genes, desde la dificultad para el reconocimiento de los factores de transcripción de sus secuencias blanco en el ADN, así como para la

elongación de la ARN polimerasa II en la generación del ARNm. Por tal motivo, se ha descrito a la cromatina como un entorno represivo para la expresión de los genes. Tal es así que para la expresión de los genes es necesario estructurar la cromatina a un nivel permisible para la transcripción, así como para otros procesos asociados al genoma como los son la replicación, la reparación del ADN y la recombinación, entre otros. Por ejemplo, las regiones genómicas en estado de eucromatina son duplicadas primero con respecto a las secuencias en estado de heterocromatina ocurriendo en la parte final de la fase S del ciclo celular.

Ante este escenario, en la evolución de las células eucariontes se han adquirido distintos mecanismos y procesos ligados a la regulación de la estructura de la cromatina, de manera tal que puedan generarse los escenarios permisibles para las diferentes funciones asociadas al genoma. En la actualidad, mucha de la investigación en torno a la cromatina se concentra en el estudio de las modificaciones covalentes y no covalentes del ADN y las modificaciones post-traduccionales de las histonas y la influencia que tienen en conjunto con la organización del genoma en cromatina.

La manera en que la cromatina está presente a lo largo del genoma no es homogénea en términos de su organización y estructura. Los diferentes niveles de organización de la cromatina pueden ser inclusive observados al nivel de la microscopia. En un núcleo celular interfásico típicamente pueden ser identificadas regiones de mayor densidad óptica y electrón-densa además de la del nucléolo. Estas se sabe que corresponden a cromatina de alto grado de compactación donde no existe expresión génica: a esta conformación de la cromatina se le

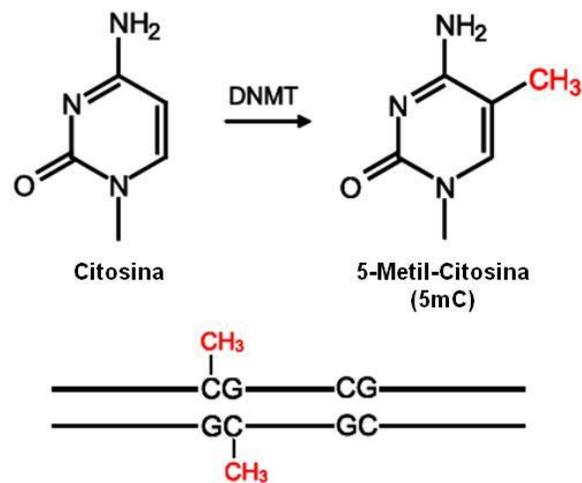
denomina heterocromatina. De manera contraria, existen regiones de menor densidad óptica y electrón-densas donde se ha distinguido la mayor parte de la actividad transcripcional: a esta cromatina se le denomina eucromatina<sup>37</sup>. Este dinamismo entre exponer y ocultar secuencias regulatorias, es lo que ha permitido a los organismos eucariontes la capacidad de desarrollar diferentes grados de complejidad funcional y regulatorio como por ejemplo la diferenciación celular entre muchos otros. Es decir, la ejecución de los programas de expresión génica propia a cada tipo celular en los metazoarios depende directamente en diferencias en la estructura de la cromatina en cada tipo celular.

### **3.2. La metilación del ADN**

La metilación del ADN es uno de los mecanismos epigenéticos mas estudiados<sup>33</sup>. La metilación ocurre en el contexto del di-nucleótido C-G (CpG) mediante la incorporación de un grupo metilo (CH<sub>3</sub>) en la posición 5 de la citosina generando la modificación conocida como 5-metilcitosina<sup>38</sup> (5mC; Figura 5). La incorporación de este grupo metilo a en los CpGs no modifica la secuencia primaria del ADN, por lo que no se considera una modificación genética. Sin embargo, esta modificación covalente puede influenciar directamente en la actividad transcripcional de los genes <sup>39</sup>. La 5mC es una modificación común para la mayoría de los organismos eucariontes como las plantas, los animales y ciertos hongos, por lo que se considera como un proceso critico en términos de evolución dada su conservación y presencia entre los organismos. A pesar de ello, no todos

los organismos tienen metilación del ADN como uno de sus procesos de regulación y estabilidad de su genoma<sup>40</sup>.

El proceso de metilación del ADN está dirigido por las enzimas denominadas ADN metil-transferasas o Dnmts (por sus siglas en inglés). La Dnmt1 es la encargada de metilar las CpGs en la horquilla de replicación a partir de un templado de 5mC pre-existente, es decir, se considera una enzima de mantenimiento y una de las responsables de segregar información epigenética en cada división celular (Figura 5). Las enzimas Dnmt3a y Dnmt3b tienen la capacidad de metilar CpGs que no se encuentren previamente metiladas, por lo que se les denomina enzimas *de novo*<sup>41,42</sup>.



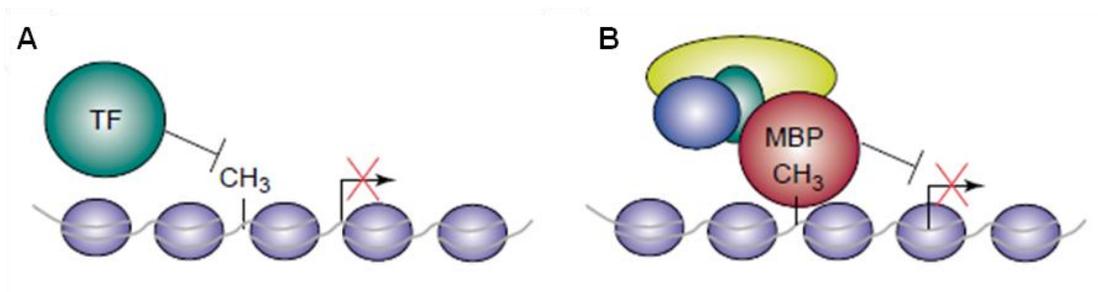
**Figura 5.** Metilación del ADN en CpGs de los eucariontes. Mediada por la actividad enzimática de las Dnmts, las citosinas en un contexto del di-nucleótido CpG (esquema de abajo) pueden ser selectivamente metiladas en el carbono 5 de la citosinas.

El principal papel atribuido a la metilación del ADN es el de la represión transcripcional y por ende con un estado de la cromatina cerrado. Un aspecto

importante de la metilación del ADN es su participación de manera normal en la inactivación de elementos móviles del genoma como lo son los transposones y secuencias de origen retroviral, por consecuencia tiene una función en el mantenimiento de la estabilidad del genoma evitando eventos de recombinación no deseados<sup>41,42</sup>. También se ha observado que está involucrada en varios procesos como en la expresión mono-alélica de genes improntados, manteniendo un estado cromatínico cerrado de secuencias repetidas y en la inactivación del cromosoma X de hembras con compensación de dosis<sup>39,43</sup>.

### 3.3. Represión génica y metilación del ADN

Los efectos represivos de la metilación del ADN en la expresión génica y de las secuencias repetidas involucran dos mecanismos principales: 1) la metilación del ADN presente en las secuencias que coinciden con motivos de unión para factores de transcripción interfiriendo con la unión de estos impidiendo ejerzan su efecto de activación génica (Figura 5); alternativamente, 2) la metilación del ADN puede ser una “marca” que es reconocida por co-factores que específicamente se unen al DNA metilado. Estas proteínas se denominan proteínas de unión a CpG-metilada (MBP, por sus siglas en inglés, Figura 6)<sup>39,44</sup>.



**Figura 6.** Mecanismos de represión mediados por la metilación del ADN. **A)** La presencia de 5mC puede interferir con la unión de los activadores de la transcripción o bien **B)** esta marca puede ser reconocida por las MBPs reclutando complejos remodeladores que pueden fomentar un estado de cromatina cerrada o represiva a la transcripción.

Estudios del grupo de Adrian Bird, basados en la capacidad de estas MBPs para unirse *in vitro* al ADN metilado, han determinado que este grupo de proteínas son las encargadas de “leer” esta marca, es decir, unirse a las C<sup>m</sup>pGs<sup>45,46</sup>. Además, como parte de su estructura peptídica, las proteínas MBD1, MBD2, MBD4 y MeCP2 presentan un dominio de interacción con remodeladores de cromatina, desacetilasas de histonas y metilasas de histonas aunado al dominio de unión al ADN metilado<sup>47</sup>. De tal manera que, el efecto de represión transcripcional de estas proteínas está mediado por la actividad que fomenta la formación de una cromatina compacta en las regiones del genoma donde estas se encuentran unidas (Figura 6B). Además, se ha descrito que estas proteínas desempeñan un papel crucial para el desarrollo del ratón, ya que ratones mutantes para los genes que codifican para las proteínas *Mbd1*, *Mbd2* y *Mbd4* presentan fenotipos que comprometen el desarrollo embrionario<sup>48</sup>. Inclusive, mutaciones en la MBP mas estudiada, MeCP2, causan un desorden neurológico conocido como el Síndrome de Rett en hembras siendo letal en machos<sup>42</sup>.

### **3.4. Metilación del ADN y su distribución en el genoma**

La distribución del di-nucleótido CpG no es homogénea a lo largo del genoma de los vertebrados. Existen regiones en el genoma con alta densidad de CpGs y se

conocen como islas-CpGs y la mayor parte de estas corresponden con promotores de genes. De hecho, los promotores de los genes asociados a islas-CpGs pueden ser clasificados en tres grupos de acuerdo con la densidad presente de este di-nucleótido; siendo de alta, intermedia y baja densidad de CpGs donde la presencia de la metilación y su efecto difiere entre ellos<sup>49</sup>. Los promotores asociados a islas-CpGs de alta densidad generalmente se encuentran hipo-metilados y no necesariamente transcripcionalmente activos, por lo que la metilación no parece gobernar por completo la regulación de estos promotores. Los de densidad baja de CpGs parecen ser altamente impactados por la presencia de la metilación ya que cuando se encuentran hiper-metilados no hay transcripción, mientras que en su estado de hipo-metilación los promotores son activos<sup>49</sup>.

Un aspecto muy interesante de la distribución del ADN metilado es su presencia en las porciones codificantes de los genes conocidas como cuerpo de los genes<sup>50,51</sup>. Lo que se ha podido observar es que genes con alta actividad transcripcional correlacionan con un estado de hipermetilación en su cuerpo del gen sugiriendo un papel en la elongación y la edición (*splicing*) de la transcripción<sup>50,52</sup>. Este comportamiento en la distribución de la metilación del ADN se ha observado en células madre así como de células diferenciadas a partir de estas, y células cancerígenas<sup>50</sup>.

En conclusión, el contenido de CpGs (bajo, intermedia y alto) impone en los promotores una regulación transcripcional puesto su pre-disposición a ser o no metilados. Además, la metilación del ADN presente en los cuerpos de los genes

coincide con actividad transcripcional, por lo que es necesario entender su papel más allá de la represión génica.

### **3.5. Metilación del ADN y Cáncer**

Existen alteraciones al nivel epigenético asociadas a enfermedades como lo es el cáncer. Como se ha mencionado, en condiciones normales las islas-CpGs se encuentran libres de metilación<sup>49</sup>. Sin embargo, en células neoplásicas se han caracterizado patrones de metilación distintos o aberrantes, definidos por una hipo-metilación generalizada<sup>53-55</sup>. Como una consecuencia de esto, los programas de expresión de genes se ven alterados y por ejemplo, genes que deberían estar silenciados por la metilación en el ADN, son activos transcripcionalmente. Además la desmetilación produce una inestabilidad genómica por la movilización de elementos transposones y retro-transposones<sup>54,55</sup>. Además de la hipo-metilación en el cáncer, existe una ganancia de metilación, o hiper-metilación de las islas CpG de ciertos genes asociados al control del ciclo celular la diferenciación celular<sup>56</sup>. Por ejemplo, el gen supresor de tumores *Retinoblastoma* (Rb) su promotor presenta un alto contenido de CpGs. Dicho promotor encuentra hipo-metilado en células que presentan expresión de Rb e hiper-metilado en células donde el gen no es activo a la transcripción<sup>57</sup>.

### **3.6. Resumen: la metilación del ADN**

La incorporación de un grupo metilo en las CpGs es un proceso altamente regulado y dirigido a regiones específicas en el genoma. En términos generales,

la presencia de la C<sup>m</sup>pG se asocia con un estado de represión transcripcional de los genes (mediante factores y co-factores de represión) y alteraciones a esta distribución se ven asociadas con enfermedades como el cáncer. También se ha observado que la metilación del ADN es un proceso dinámico ya que puede ser removida de los promotores en un tiempo y espacio específico. Sin embargo, hasta el momento no se conoce ninguna proteína con la actividad de des-metilasa de ADN. Estudios recientes han encontrado que la desmetilación del ADN puede estar mediada no por la actividad de una des-metilasa de ADN, sino por todo un mecanismo que involucra la reparación del ADN. Dado la importancia en el estudio de la metilación de ADN y en función de los resultados que se mostrarán más adelante, a continuación se describen los hallazgos en cuanto a la desmetilación del ADN y su papel en la regulación de los genes.

### **3.7. Desmetilacion del ADN**

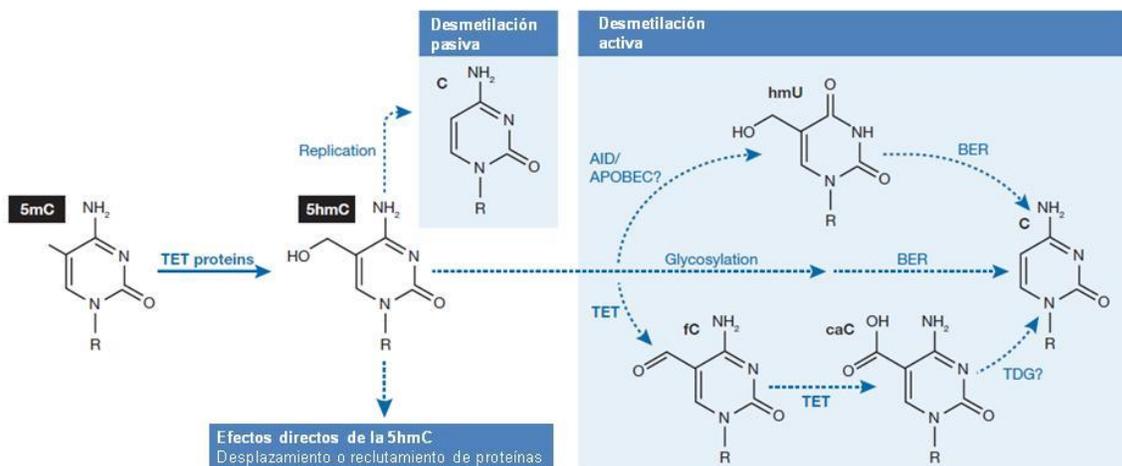
Como se ha mencionado, la metilación del ADN puede ser muy estable en términos de herencia a través de las divisiones celulares y la diferenciación celular, esto gracias a la presencia de las proteínas Dnmts. Sin embargo, se sabe que durante el desarrollo embrionario y en el establecimiento de la línea germinal ocurren ondas de desmetilación masivas del genoma<sup>58,59</sup>. Además, se ha reportado la desmetilación de *loci* específicos a través de la diferenciación celular somática<sup>50</sup>. Por consecuencia, se puede concluir que la desmetilación de regiones específicas del genoma es un mecanismo epigenético de regulación. No obstante,

a la fecha no se conoce actividad enzimática capaz de remover el grupo metilo de la 5mC<sup>33</sup>. Una explicación podría estar en que a través de las divisiones celulares, la metilación del ADN no es copiada en la horquilla de replicación. Sin embargo, la desmetilación puede ocurrir también en ausencia de mantenimiento de la 5mC a través de las divisiones celulares en cultivo<sup>59</sup>. En su conjunto, esto sugiere que la desmetilación del ADN puede ser un proceso pasivo (sin copiarse en las divisiones celulares) o activo (no dependiente de la división celular).

Como se acaba de mencionar, en las primeras etapas después de la fertilización las células embrionarias madre (ESCs, por sus siglas en inglés) experimentan una onda global de desmetilación. Esta desmetilación puede estar mediada por la capacidad que tienen estas células de dividirse rápidamente, por lo que la desmetilación puede ocurrir de manera pasiva<sup>60</sup>. Sin embargo, estudios recientes han detectado una gran abundancia de una forma hidroxilada de la metilación del ADN presente en estas células<sup>61</sup>. Esta modificación corresponde a la 5-hidroxi-metilcitosina (5hmC). De manera similar, en células del sistema nervioso la abundancia de la 5hmC es mayor en comparación con otros tejidos, pero a diferencia de las ESCs, las células del cerebro se dividen a una tasa muy baja o prácticamente nula por lo que la desmetilación pasiva no podría ser un mecanismo muy factible para la desmetilación del ADN<sup>62</sup> (Figura 6). Por lo tanto, la presencia de 5hmC se ha propuesto como un intermediario entre el estado metilado y no metilado de la citosina en su contexto CpG<sup>52,63,64</sup>.

Estudios recientes han descrito a un nivel enzimático la conversión de 5mC en 5-hidroxi-metilcitosina; la familia de enzimas 'Ten-Eleven Translocation' (TET)

tienen la capacidad de oxidar la 5mC teniendo como producto a la 5hmC<sup>65,66</sup> (Figura 7). La familia de proteínas TET consta de tres miembros, TET1, TET2 y TET3, todas ellas con la capacidad de oxidar la 5mC en 5hmC tanto *in vitro* como *in vivo*<sup>52,64</sup>. Este proceso de generación de 5hmC es altamente regulado y no ocurre de manera azarosa en el genoma, ya que el análisis del patrón de distribución de la TET1 en células ESCs de ratón a escala del genoma completo demostró que esta se encuentra preferentemente unida en las islas-CpGs con niveles bajos de 5mC<sup>67,68</sup>. Sin embargo, su papel en la activación y represión génica queda por ser clarificado puesto que la presencia de 5hmC coincide tanto con sitios de hipersensibilidad a la DNasa I pero de igual manera con regiones de cromatina cerrada o represiva mediada por los complejos Sin3A y Polycomb<sup>68</sup>.



**Figura 7.** Reacciones enzimáticas asociadas a la desmetilación del ADN. Cuando las proteínas TET generan la 5hmC a partir de la oxidación de la 5mC pre-existente, esta puede ser una marca que puede dirigir diversos procesos ya sea de ejecución directa al ser reconocida por otras proteínas o bien llevar a una desmetilación pasiva (replicación). Cuando la desmetilación es un proceso activo, involucra una serie de modificaciones químicas de la 5hmC por diversos complejos entre ellos AID/APOBEC, BER o TDG. Tomada y modificada de<sup>63</sup>.

Dentro de los mecanismo de desmetilación activa se ha propuesto que la generación de la 5hmC es una ruta intermediaria entre la 5mC y las citosinas no metiladas. De hecho, se ha observado que algunos de los mecanismos de reparación del ADN tales como los de deaminación mediados por el complejo AID/APOBEC tienen la capacidad de convertir la 5hmC en 5hmU (5-hidroximetiluracilo) seguido por la reparación por escisión de base (BER, por sus siglas en inglés) o bien la 5hmC es glicosilada seguido de la acción de BER llevando a la generación de citosinas no metiladas<sup>63,64</sup> (Figura 7).

Alternativamente a estos mecanismos de desmetilación, el grupo de Xo y colaboradores<sup>69</sup> describieron que la propia TET1 responsable de la 5hmC, tiene la capacidad de realizar dos oxidaciones subsecuentes llevando a la modificación hacia formil-citosina (5fC) y posteriormente a carboxil-citocina (5caC; Figura 7). De manera que una actividad de de-carboxilacion podría llevar a la generación de citosinas no metiladas. Este mismo grupo, describió que una enzima llamada ADN timidina glicosilasa (TDG por sus siglas en inglés) es la responsable de dicha de-carboxilación llevando a la desmetilación de las citosinas<sup>63,69</sup>.

En resumen, se ha descrito que la metilación del ADN es una modificación epigenética implicada en la regulación de la transcripción por diversos mecanismos sea a través de la interferencia con la unión de los factores de transcripción, o bien, mediante el reclutamiento de complejos remodeladores de la cromatina especializados en la represión génica mediante la compactación de la estructura de la cromatina. Por otra parte, la desmetilación del ADN parece ser un evento altamente conservado y de gran relevancia en términos de la regulación

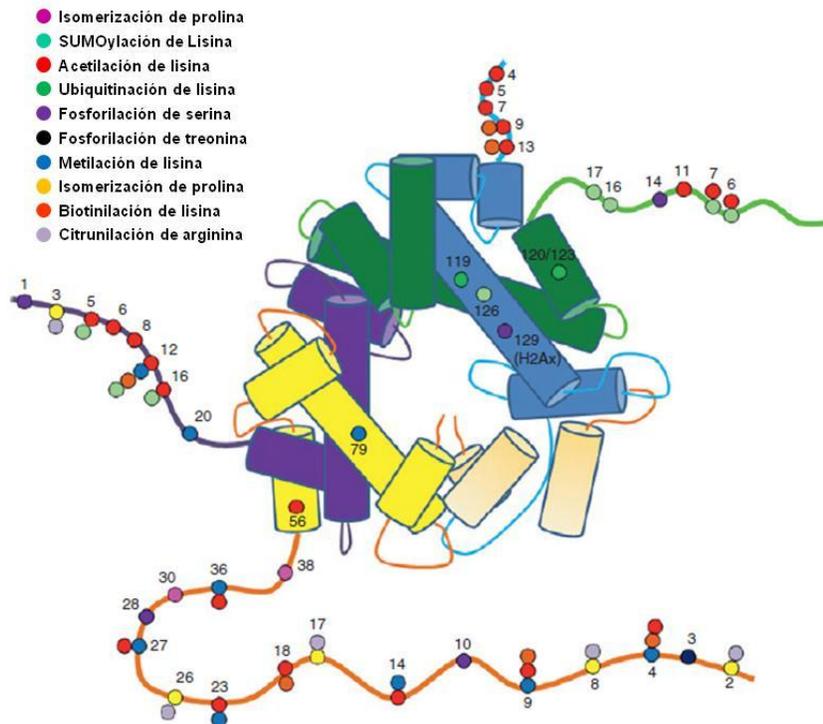
epigenética. Tanto la presencia de 5mC así como de la 5hmC y sus derivados, podrían ser vistas como marcas primarias para los procesos biológicos ligados a la regulación del epigenoma, por lo que es necesario considerar sus componentes y su participación en la estructuración de la cromatina. Por tal motivo, a continuación se describen el papel de la estructura de la cromatina en regulación de la expresión génica y otros procesos.

#### **4. Un código epigenético en las histonas**

El estado de cromatina en el cual el genoma de los eucariontes se encuentra empaquetado es blanco de múltiples niveles de regulación que por consecuencia permiten la ejecución de los diversos procesos ligados al genoma. El establecimiento de los diferentes niveles en que la cromatina puede estar estructurada es el resultado de una interdependencia entre las secuencias de ADN, los factores de transcripción y los diversos procesos epigenéticos de regulación. De entre ellos, las histonas que constituyen parte fundamental de la estructura de la cromatina, son blancos de una plétora de modificaciones post-traduccionales que permiten la formación de estados particulares y regulados de la estructura de la cromatina y además estas modificaciones pueden incorporar una marca heredable que es relevante para los procesos del desarrollo del organismos y la diferenciación celular.

Tanto el dominio globular como la región amino-terminal de las histonas son blancos de una gran serie de modificaciones post-traduccionales. Estas

modificaciones incluyen a la metilación, la acetilación, la ubiquitinación, la ADP-ribosilación y sumoilación de los residuos de lisina (K); también la metilación de residuos de arginina (R); y la fosforilación de serinas y treoninas; entre otras<sup>70</sup> (Figura 8). Sin embargo, de entre estas modificaciones la metilación/desmetilación y la acetilación/desacetilación, su establecimiento y papel en la regulación del genoma han sido las más estudiadas. Ante este escenario, se ha llegado a postular que la interdependencia existente entre las diferentes modificaciones puede ser un código de histonas que marcaría la ejecución de programas que podrían definir a cada tipo celular a través de patrones de modificaciones en las histonas específicos a los largo de todo el genoma<sup>3</sup>.



**Figura 8.** Modificaciones post-traduccionales presentes en las histonas. En esta representación de un nucleosoma se muestra la estructura de los cuatro poli-péptidos del core de histona (H2A azul; H2B verde; H3 amarillo; H4 magenta; tomado y modificado

de<sup>70</sup>. Resulta importante resaltar que existen modificaciones que incluso se localizan al interior del nucleosoma y cuyas funciones no han podido ser clarificadas aún.

La estrategia más utilizada para el estudio de estas marcas y su distribución en el genoma es una técnica experimental conocida como inmunoprecipitación de cromatina (ChIP, por sus siglas en inglés). Esta técnica permite conocer la presencia y abundancia relativa de una marca de histonas, factores de transcripción y otras proteínas en un contexto *in vivo* tanto de células en cultivo como de células provenientes de tejidos. El ChIP es un protocolo basado en el enriquecimiento de las secuencias donde modificaciones en las histonas o la proteína de interés asociada *in vivo* se enriquecen mediante su inmunoprecipitación con anticuerpos específicos. El enriquecimiento de las secuencias puede ser evaluado mediante ensayos de PCR o más recientemente con secuenciación masiva de todo el genoma<sup>71</sup>. De este modo, se ha podido describir la distribución de estas marcas a lo largo del genoma y su relación con la estructura de la cromatina y la expresión génica, además de otros procesos. A continuación, se describen algunos de los hallazgos a este respecto.

#### **4.1. La acetilación de histonas**

La asociación del ADN con el “core” de histonas se ve facilitada por diferencia de cargas entre estas moléculas<sup>35</sup>; la acetilación de las histonas neutraliza la carga positiva de las lisinas debilitando la interacción del nucleosoma con el ADN y por ende favoreciendo una mayor movilidad del nucleosoma exponiendo secuencias

en el ADN a ser 'leídas' para llevar a cabo la transcripción<sup>35</sup>. Las enzimas encargadas de acetilar a los residuos son las acetil-transferasas de histonas (HATs, por sus siglas en inglés) como lo son p300/CBP, Gcn5, PCAF entre otras<sup>72,73</sup>. Además de relajar la estructura de la cromatina, se ha observado que la acetilación puede ser una marca que es reconocida por proteínas que poseen un bromodominio las cuales podrían estar involucradas en la regulación positiva de la transcripción<sup>74</sup>.

Es importante mencionar que las HATs no poseen dominio alguno para el reconocimiento de secuencias de ADN, por lo que necesitan reconocer y asociarse a dominios peptídicos específicos en co-factores y factores de transcripción para ser reclutadas a las regiones genómicas donde las histonas deben ser acetiladas, como por ejemplo en promotores de los genes activos transcripcionalmente. Sin embargo, al menos para PCAF se ha demostrado la capacidad de reconocer tanto *in vivo* como *in vitro* la acetilación de la histona H4 a través del bromodominio presente en esta enzima, por lo que la acción HATs pudiera no depender de las secuencias y los factores de transcripción<sup>74,75</sup>.

Estudios donde se evaluó la abundancia y distribución de las histonas acetiladas en células T CD4<sup>+</sup> de humano y del linaje hematopoyético, determinaron que la acetilación de histonas se asocia principalmente en sitios de inicio de la transcripción de genes activos<sup>76-78</sup>. Sin embargo, una de las conclusiones más relevante de estos y otros estudios posteriores a escala de genomas completos, fue que las regiones de enriquecimiento son dinámicas y varían entre los diferentes tipos celulares, llevando a la propuesta de que la

acetilación de histonas, en su conjunto con otras modificaciones en las histonas contribuyen en la diferenciación de distintos tipos celulares<sup>79,80</sup>.

Otra particularidad de la distribución global de las histonas acetiladas es su presencia en los elementos *enhancers*. Estudios a nivel genómico o “genome-wide”, encontraron que la unión de PCAF y p300/CBP es preferente en los *enhancers* y que su efecto se ve reflejado por un enriquecimiento típico a estos elementos de la marca de acetilación de la lisina 27 de la histona H3 (H3K27ac)<sup>81</sup>. En la actualidad, la presencia de la H3K27ac es utilizada para la identificación *in vivo* de *enhancers* activos, en conjunto con la mono-metilación de la histona H3 en su lisina 4; H3K4me1. La presencia de la acetilación tanto en los promotores de los genes activos así como en los *enhancers* se cree que promueve un estado abierto y relajado de la cromatina lo cual facilita la unión de la maquinaria de transcripción, formación de asas de cromatina y la elongación de la transcripción<sup>81</sup>.

#### **4.2. La desacetilación de histonas**

El establecimiento de la acetilación de histonas no solo depende de la acción de las enzimas HATs; la acetilación es una marca reversible en las histonas y su remoción es regulada por la actividad de las enzimas desacetilasas de histonas<sup>82</sup> (HDACs, por sus siglas en inglés). Estas enzimas están altamente conservadas a nivel evolutivo desde la levadura hasta el humano. Existen alrededor de 18 distintas HDACs en humano las cuales pueden ser agrupadas en cuatro familias

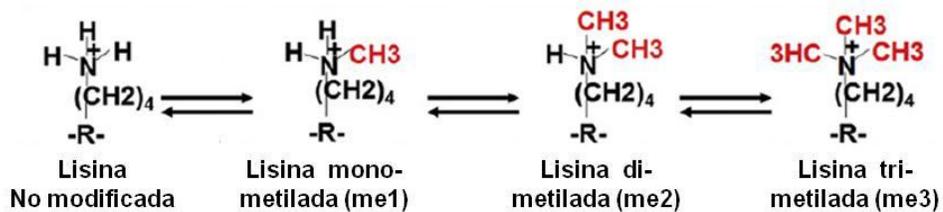
de acuerdo a su homología con las de la levadura: Clase I; HDAC1, 2, 3 y 8 con homología a Rpd3, la Clase II; HDAC4, 5, 6, 7, 9 y 10 con homología a Hda1, la Clase III; Sirt1-7 con homología a Sir2, y la clase IV; que incluye a HDAC11<sup>78</sup>. La desacetilación de histonas generalmente se asocia con represión génica al favorecer la formación de una cromatina compacta, las HDACs han sido identificadas como co-represores de la transcripción.

La manera de actuar de las HDACs es a través de complejos de represión como factores que reconocen secuencias de ADN, co-factores (co-represores) así como el reconocimiento de modificaciones post-traduccionales específicas en las histonas. En particular, se sabe que las MBDs tienen la capacidad de interactuar con HDACs para ejercer su efecto de represión de la transcripción fomentando un estado hipo-acetilado de la cromatina y por tanto compacto. De esta forma las proteínas de unión al ADN metilado MeCP2, MBD2 y MBD4 pueden interactuar directamente con HDACs (Figura 5B) o bien, en asociación con los complejos co-represores como Sin3A<sup>83,84</sup>.

### **4.3. La metilación de histonas**

La presencia de metilación en los residuos de lisina y arginina en las regiones amino terminales de las histonas puede estar involucrada tanto en procesos de activación o represión génica dependiendo del residuo que se encuentre metilado. Un aspecto relevante es que a diferencia de la acetilación de histonas, la metilación no altera las fuerzas electrostáticas del nucleosoma; la metilación de

histonas es una “marca” o “señal” que debe ser reconocida para llevar a la activación o represión génica. La metilación de dichos residuos puede presentarse en tres estados: la adición de solo un grupo metilo (mono-metilación, me1), dos metilos (di-metilación, me2) o bien tres metilos (tri-metilación, me3), como se ejemplifica en la Figura 9<sup>85</sup>. Además, es importante recalcar que este es un proceso dinámico, pues se ha descrito la desmetilación de histonas mediante una actividad enzimática y con consecuencias en la regulación de la estructura de la cromatina y la expresión génica.

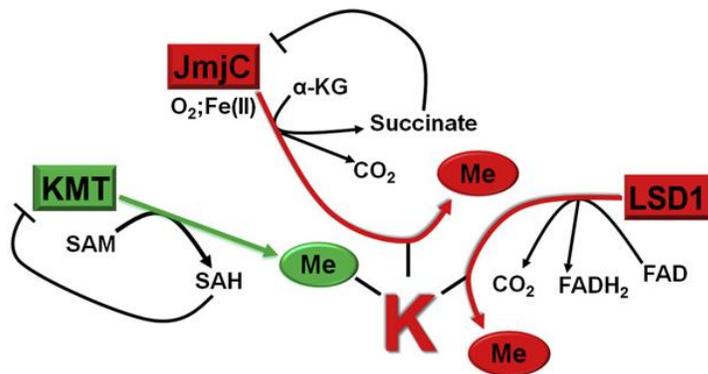


**Figura 9.** Metilación de residuos de lisina en las histonas. La metilación de histonas es un proceso reversible ya que se han descrito desmetilasas de histonas. Tomada y modificada de<sup>85</sup>.

La metilación de los residuos, en particular los de lisinas, se encuentra altamente regulada por la acción de las enzimas metil-transferasas de lisina (KMTs, por sus siglas en inglés) y lo logran a través de un dominio de metilación de histonas conservado denominado SET (Figura 9)<sup>86</sup>. Las posiciones más estudiadas de metilación de histonas son las siguientes: la lisina 4 de la histona H3 (H3K4), H3K9, H3K27, H3K36, H3K79 y la H4K20. Con respecto a las posiciones de arginina se incluyen la H3R2, H3R8, H3R17, H3R26 y la H4R3. Sin embargo, también puede presentarse metilación en residuos de las histonas H1,

H2A, H2B, sus efectos funcionales y la regulación de estas aun no se han determinado<sup>86</sup>. A la fecha, se han identificado varias enzimas con esta capacidad y sus blancos de metilación se han podido establecer en mamíferos, insectos y nematodos<sup>86</sup>.

En años recientes se ha descrito que la metilación de las histonas puede también ser revertida por un mecanismo de desmetilación mediado por la familia de enzimas desmetilasas de lisinas (KDMs, por sus siglas en inglés). Esta actividad de desmetilación fue descubierta en el 2004 por el grupo de Shi y colaboradores al observar que la enzima LSD1 podía desmetilar histonas tanto *in vitro* como *in vivo* (Figuras 9 y 10)<sup>87</sup>. Posteriormente, otras enzimas involucradas en este proceso fueron descubiertas, las cuales corresponden a KDMs con dominio Jumonji C (JmjC) de digoxigenasa dependiente de hierro<sup>88</sup> (Figura 10).



**Figura 10.** Esquema generalizado de los mecanismos de reacción de las KMTs y de las KDMs. Se muestran tanto las enzimas involucradas en la metilación de la lisina (verde), en la desmetilación (rojo), sus donadores y derivados. Tomada de<sup>85</sup>.

La metilación de la histona H3 en su lisina 4 (H3K4) es una de las modificaciones más estudiadas y se ha relacionado con la regulación positiva de la transcripción. Trabajos donde se ha evaluado la distribución global o genómica de este tipo de marcas han alcanzado las siguientes conclusiones: la mono-metilación (H3K4me1) es una modificación presente en elementos *enhancers* de células troncales embrionarias que no son activos pero están pre-determinados a ser activos en la diferenciación de las células<sup>89</sup>; la H3K4me2 se encuentra ampliamente distribuida en sitios de transcripción activa, *enhancers* y de cromatina relajada<sup>15,90</sup> y, la H3K4me3 es una marca específica de promotores de genes transcripcionalmente activos y de islas-CpG con hipometilación del ADN. El establecimiento de estos patrones específicos de marcas se atribuye a los factores de transcripción y complejos con los cuales las KMTs pueden interactuar. En particular, los complejos de Trithorax pueden reconocer secuencias específicas en el ADN llamadas TRE (del inglés: Trithorax Responce Elements) los cuales han sido ampliamente descrito en la mosca de la fruta<sup>86</sup>.

En cuanto al efecto en la represión génica por parte de la metilación de histonas, se sabe que la marca de la H3K9me2, H3K9me3, H3K27me3 y la H4K20me3 son abundantes en genes que no presentan actividad transcripcional. Inclusive, la H3K9me3 es una las marcas características de zonas de heterocromatina constitutiva como lo son los telómeros y centrómeros de los cromosomas. Se postula que estas marcas representan un código para la acción de otras proteínas que tiene que ver con la compactación de la cromatina, como lo son la proteína de heterocromatina HP1 y la histona H1<sup>85</sup>. En general, las

regiones genómicas ricas en secuencias repetidas desde levaduras hasta humanos tienen un estado de cromatina cerrada donde la metilación en las histonas es prevalente<sup>76</sup>. Para el caso de la regulación transcripcional, se sabe que genes silenciados por histonas con las marcas H3K9me2 y me3 en su secuencia promotora pueden ser blancos de acción de las KDMs para el “borrado” de estas marcas de metilación y posteriormente la actividad de las KMTs puede llevar a un estado de cromatina permisible a la transcripción<sup>91</sup>. Este tipo de proceso es de suma importancia dado que la reversibilidad de este tipo representa otro nivel de regulación epigenético.

Una marca muy particular en términos de memoria epigenética y diferenciación celular es la H3K27me3; esta modificación post-traducciona es catalizada por la acción de varias enzimas de entre ellas la más estudiada es Enhancer de Zeste (EZH2)<sup>92</sup>. EZH2 pertenece a un grupo de proteínas represoras Polycomb, en particular del grupo conocido como PRC2 (del inglés: Polycomb Repressive Complex 2). Conformando el complejo PRC2, también se encuentran las proteínas EED y SUZ12 las cuales son necesarias para el reconocimiento de la H3K27me3 en la fase S del ciclo celular, lo que permite que esta marca sea propagada a las células hijas como una señal epigenética<sup>93,94</sup>. Un aspecto muy interesante es que un gran grupo de promotores de genes involucrados en el establecimiento de linajes celulares específicos son silenciados por el enriquecimiento de H3K27me3 en ESCs<sup>95,96</sup>. A su vez, el borrado selectivo de la H3K27me3 a través de la diferenciación celular se puede llevar a la expresión regulada de los genes específicos de cada linaje celular.

#### 4.4. Cromatina bivalente

Mediante estudios de ChIP-seq los cuales permiten evaluar el enriquecimiento de marcas de cromatina o factores de transcripción a una escala del genoma completo, se pudo determinar en ESCs que existen regiones del genoma, en particular promotores de ciertos genes, donde co-existen tanto las marcas de apertura de la cromatina mediadas por Trithorax (activación) y las catalizadas por Polycomb (represión). En otros términos, son regiones con enriquecimiento tanto de H3K4me3 y H3K27me3, respectivamente<sup>97</sup>. A esta combinación de marcas tanto de cromatina represiva como de activación se le denomina “cromatina bivalente” y es una característica exclusiva de las células pluripotenciales<sup>98</sup>.

Los dominios de marcas bivalentes resultan de particular atención pues se cree que contribuyen a la ejecución de programas de expresión necesarios para mantener un estado pluripotente de estas células y en los eventos iniciales que llevan a la diferenciación celular. En particular, se ha propuesto que los dominios bivalentes pueden reprimir a los genes que controlan algún linaje vía la ganancia de la marca H3K27me3 durante la pluripotencialidad pero permaneciendo latentes a la activación en la diferenciación por la presencia de la H3K4me3. En este escenario, la H3K4me3 podría lograr que la inducción de los genes importantes para el desarrollo sea más eficiente. Adicionalmente, se ha sugerido que la H3K4me3 podría proteger a estos genes de un silenciamiento permanente o estable, por ejemplo, evitando la formación de una cromatina represora como lo hace la metilación del ADN<sup>98</sup>.

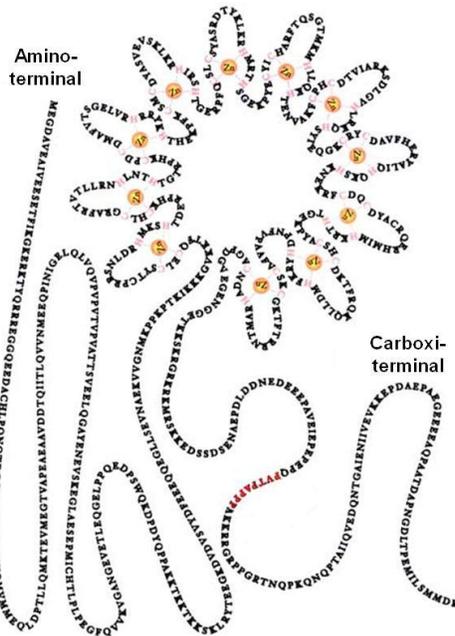
#### **4.5. Resumen: epigenética y cromatina**

La regulación de la expresión génica en los eucariontes se distingue por diversos procesos de regulación vinculados con el empaquetamiento del genoma en el núcleo. Esta organización del genoma en cromatina desempeña un papel clave modulando el tiempo y el lugar de la expresión de los genes. Se ha descrito la función represora de la metilación en el ADN, la activación génica ante la desmetilación y papel de ambos procesos en la memoria epigenética. De igual forma, la cromatina y sus modificaciones asociadas a esta desempeñan un nivel de regulación en los genes para la ejecución de los distintos programas de expresión génica que en parte, contribuyen a la pluripotencialidad de las células madre, la generación de precursores celulares, así como en los procesos de diferenciación celular. Sin embargo, un aspecto central que tiene relación con los resultados presentados en esta tesis, es lo poco que se sabe con respecto a los factores o mecanismos que pueden llevar a la coordinación de tan diversos procesos de regulación epigenética y génica. Esto sugiere la necesidad de elementos que orquesten a los mecanismos genéticos y epigenéticos algunos de los cuales incluye a los factores de transcripción tales como CTCF y otras moléculas tales como los ARNs no-codificantes. En el presente trabajo de tesis se ha estudiado la participación de factor CTCF en la regulación de la familia de genes  $\alpha$ -globina de pollo a través del desarrollo del pollo y la diferenciación eritroide. A continuación se describen algunos aspectos generales de CTCF y otros factores.

## 5. El factor multi-funcional CTCF y la regulación génica

En el año 1990, se describió la acción represora de la transcripción de un factor nuclear denominado CTCF o Nep1 sobre el gen *myc* de pollo<sup>99</sup>. Posteriormente, se pudo determinar su papel en la activación del gen de la proteína precursora  $\beta$ -*mieloide* por la unión *in vitro* de CTCF a la región promotora de dicho gen<sup>100</sup>. Por ello, se definió a la proteína CTCF como un factor transcripcional tanto represivo como activador. Estudios posteriores, lo han propuesto como un factor multi-funcional puesto que, como se abordará en esta sección, son múltiples las funciones atribuibles a CTCF<sup>30</sup>. CTCF está conservada en vertebrados, es una proteína de 82 kDa compuesta por 11 dedos de zinc<sup>29,30</sup> los cuales tienen cerca del 100% de conservación entre humano, ratón, y el pollo, dejando la mayor variabilidad en las regiones amino y carboxi-terminales de la proteína<sup>101</sup>(Figura 11).

El factor CTCF parecer ser una proteína crucial en términos del desarrollo y de la fisiología de los organismos. Existen evidencias de que el “knockout” de este gen en ratones produce letalidad en etapas previas a la implantación<sup>102</sup>. En el caso de la sobreexpresión inducida de este factor, se observan alteraciones en las tasas de proliferación y capacidad de diferenciación celular<sup>103</sup>. Su patrón de expresión es ubicuo y podría considerarse un gen de expresión constitutiva, sin embargo, los niveles varían entre los tejidos lo que pudiera estar implicado en la definición de diferentes programas de expresión.



**Figura 11.** El factor nuclear CTCF. Se muestra a la proteína CTCF (secuencia de amino ácidos) y la región altamente conservada de dedos de zinc (Zn) con los cuales puede unirse al ADN.

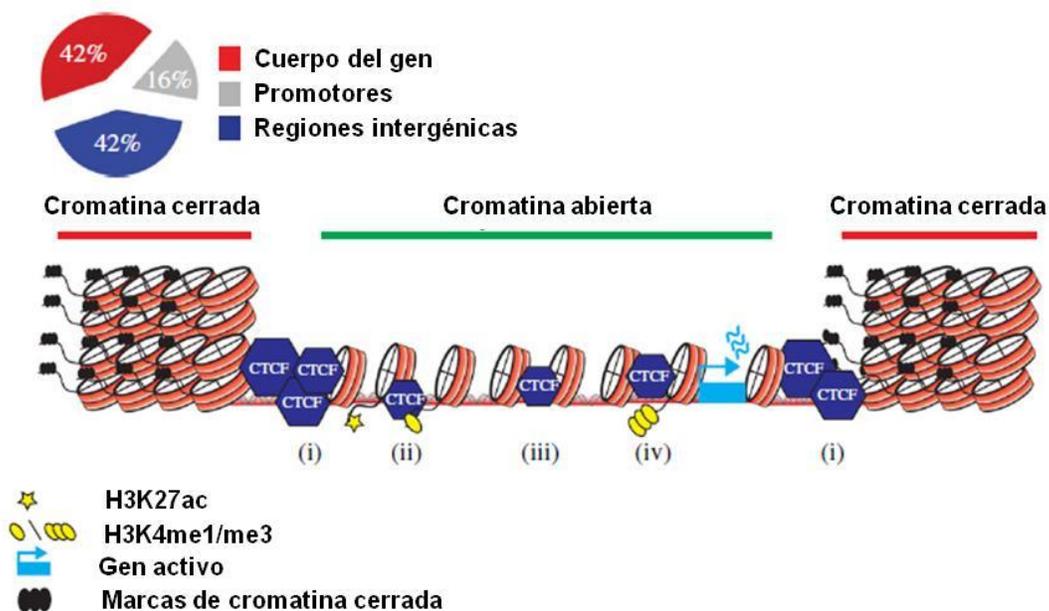
Un aspecto muy importante que vincula a CTCF con un regulador “maestro” de la transcripción es su capacidad para unirse a una amplia variedad de secuencias localizadas en distintas regiones del genoma. Estudios de los sitios a los cuales CTCF se une mediante ensayos de CHIP acoplados a secuenciación masiva del genoma (CHIP-seq), han revelado que el número de sitios de enriquecimiento pueden ir desde 5,000 en eritrocitos del pollo<sup>104</sup> hasta 20,000 o más en células humanas<sup>76</sup>. Cabe mencionar que en este trabajo de tesis presentaré resultados publicados que contribuyen al conocimiento de la distribución genómica de CTCF en distintos organismos<sup>104</sup>. Derivado de estos y otros estudios<sup>105,106</sup> se ha determinado además que la secuencia consenso de reconocimiento de CTCF puede ir de 20 a 50 nucleótidos atribuible a la presencia

de los 11 dedos de zinc. Sin embargo, cuando estos dedos de zinc son alterados solo 4 de los 11 son importantes para seguirse uniendo a la secuencia consenso<sup>107</sup>. No obstante, son necesarios más estudios asociados a la estructura de CTCF, sus modificaciones post-traduccionales y su regulación para tener un mejor entendimiento de su papel en la expresión de los genes.

La mayor parte de las funciones atribuibles a CTCF han sido elucidadas gracias a estudios del tipo de ChIP-seq, tanto de la misma proteína así como su relación con componentes asociados a la cromatina<sup>105,106</sup>. Con base a estos y otros trabajos, se sabe que CTCF participa como un delimitador definiendo transiciones entre cromatina abierta y cerrada, como un organizador topológico formando asas de cromatina (estructuración 3D del genoma), protector contra la metilación del ADN, regulando el “splicing” transcripcional y como reclutador de la ARN polimerasa II<sup>10,108</sup> (Figura 12).

Al observar tal variedad de funciones y procesos dependientes de CTCF surge la gran interrogante de cómo podría este factor ser tan ‘versátil. Como ya se ha mencionado, la gran variedad de sitios del genoma a donde CTCF se une *in vivo* mediada por sus 11 dedos de zinc sugiere un papel en la repartición de las tareas a realizar dependiente de su localización. Así mismo, se sabe que CTCF es capaz de interactuar con diferentes proteínas y complejos que tienen que ver con la regulación de la transcripción y la estructura de la cromatina. Por mencionar algunos de ellos: se ha descrito la asociación con la ARN polimerasa II lo que lo vincula como un factor regulador de los promotores de ciertos genes<sup>108</sup>.

Además, para llevar a cabo su papel como organizador topológico del genoma al interior del núcleo, CTCF se ha visto en asociación con los complejos de cohesinas (SMC3 y Rad21) los cuales participan en la condensación y organización 3D del genoma<sup>109</sup>. Mediante la asociación con las proteínas cohesinas, CTCF puede llevar a cabo la función de formar asas de cromatina entre secuencias o regiones distales del genoma con consecuencias en la regulación de la expresión génica.



**Figura 12.** Múltiples funciones de CTCF en la organización de la cromatina y la regulación transcripcional. La grafica de Pie muestra la distribución de CTCF en las distintas regiones del genoma de humano, la cual se conserva entre vertebrados. (i) Marcando las transiciones entre tipos de cromatina (abierta y cerrada) se puede encontrar al factor CTCF, a su vez, definiendo dominios de regulación donde este mismo participa: (ii) CTCF puede encontrarse en regiones *enhancers* (marcados por H3K4me1) y facilitar la interacción con su promotor blanco (iv) o bien entre el *enhancer* y promotores (iii) y llevar al bloqueo de la acción del *enhancer*. Tomada y modificada de<sup>108</sup>.

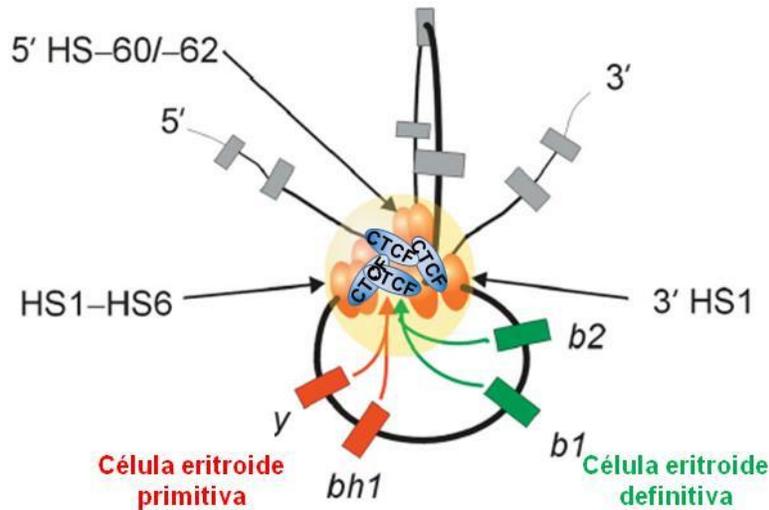
Aunada a su participación en la estructuración 3D del genoma al interior del núcleo, CTCF también se ha visto involucrado en el establecimiento de dominios de cromatina<sup>108</sup>. Derivado de estudios de ChIP-seq de CTCF y de la distribución de diversas modificaciones post-traduccionales en las histonas, se han podido encontrar coincidencias entre CTCF y ciertas marcas en las histonas sugiriendo que este factor puede estar involucrado en la determinación de regiones de cromatina abierta y cerrada. Esto es, un gran número de sitios de unión al factor CTCF se localizan en zonas del genoma que representan transiciones entre dominios de cromatina activa (H2AK5ac) y cromatina inactiva (H3K27me3)<sup>76,110</sup>. Más aun, existe una correlación entre la presencia de CTCF en zonas del genoma con alta densidad génica, coincidente con la marca específica de promotores activos, la H3K4me3, y con elementos *enhancers* (H3K4me1)<sup>76</sup>. Por lo tanto, además de definir a CTCF como un factor asociado al establecimiento de diferentes tipos de cromatina, lo relaciona directamente con la regulación de la transcripción de los genes.

### **5.1. CTCF y su papel en los dominios génicos**

Históricamente, las secuencias de unión al factor CTCF fueron relacionadas con elementos barrera, también conocidos como *insulators*. Los *insulators* son considerados como elementos de regulación a los cuales se unen factores de transcripción cuya función es la de proteger contra la expansión de la cromatina represora y bloquear el efecto de un *enhancer* (Figura 11)<sup>108</sup>. Como se

mencionó en el apartado de elementos genéticos de regulación, esta función ha sido atribuida principalmente a la presencia de CTCF<sup>27</sup>. El ejemplo más estudiado es el *insulator* del dominio  $\beta$ -globina de pollo (cHS4)<sup>25</sup>; donde estudios han demostrado que CTCF y USF1 protegen contra la heterocromatina constitutiva río arriba a los genes  $\beta$ -globina (Figura 3)<sup>22,111</sup>. Además, se ha demostrado que la presencia de CTCF en este elemento es responsable de la actividad de bloqueo de *enhancer*, llevando a la protección de los genes  $\beta$ -globina de señales externas al dominio<sup>112,113</sup>.

El dominio  $\beta$ -globina de humano y ratón se localizan en una región cromosómica caracterizada por la presencia de cromatina represora. En ambas especies, el dominio se encuentra enmarcado por la unión de CTCF en sus extremos<sup>114,115</sup>. Esto supondría un papel de CTCF como un elemento barrera que enmarca a los genes globina (como sucede en el pollo)<sup>31,116</sup>. Por otro lado, en años recientes se ha involucrado a CTCF en la organización tri-dimensional del genoma al interior del núcleo a través de la formación de asas de cromatina<sup>108</sup> (Figura 12). Mediante la aplicación de la técnica de captura conformacional de cromosomas (3C, por sus siglas en inglés) se ha demostrado que los sitios de unión a CTCF en el dominio  $\beta$ -globina pueden interactuar físicamente entre ellos, llevando a la formación de asas de cromatina que favorecen contactos entre los elementos de regulación distales contenidos en el dominio (*insulators* y *enhancers*) y los promotores de los genes (Figura 13)<sup>31,117</sup>.

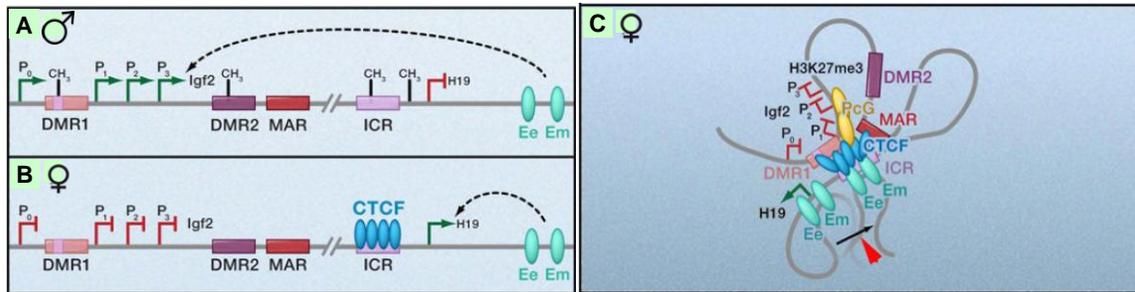


**Figura 13.** Dominio  $\beta$ -globina de ratón y su regulación mediada por CTCF. En este dominio se han descrito varios sitios de hipersensibilidad al corte por la DNasa I (HSs) los cuales corresponden con *enhancers* eritroides (HS1-HS6) que actúan sobre los genes  $\beta$ -globina ( $\gamma$ , *bh1*, *b1* y *b2*). Con la técnica de 3C, se determinó que los diferentes sitios de unión a CTCF, incluyendo el del *insulator* 3' (3'-HS1), establecen contactos físicos que entre ellos promoviendo la entrada de los genes que deben ser *trans*-activados por los *enhancers* dependiendo de la etapa eritroide. Tomada y modificada de<sup>31,117</sup>.

Otro ejemplo clásico donde la unión de CTCF contribuye a una actividad de bloqueo de *enhancers* y formador de asas de cromatina es el dominio de los genes improntados *H19/Igf2*. Es preciso mencionar que la impronta génica tiene que ver con la expresión mono-alélica de cierto grupo de genes, llevando a que estos se expresen únicamente a partir del cromosoma paterno o el materno<sup>118</sup>. En el caso del dominio de genes *H19/Igf2*, existe una región controladora de la impronta (DMR, por sus siglas del inglés: Differential Methylation Region), que se localiza entre los genes *H19* y *Igf2* (Figura 14). La presencia de esta secuencia DMR determina que el gen *H19* es el que se transcribe en el alelo materno, mientras que el gen *Igf2* se expresa en el alelo paterno<sup>119</sup>. La manera en que la

secuencia DMR dicta este patrón alelo específico de expresión, está mediada por la unión de CTCF a la región DMR de manera dependiente de su estado de metilación en el ADN: la unión de CTCF a la región DMR depende de un estado no metilado, asociado al alelo materno causando el bloqueo de los *enhancers* distales localizados río abajo del gen *H19* y, por lo tanto, impidiendo la activación del gen *Igf2* (Figura 14)<sup>118</sup>. En el alelo paterno, CTCF no es capaz de ejercer su efecto de *insulator* como bloqueador de *enhancer* dado que la región DMR se encuentra metilada impidiendo así su unión (Figura 14)<sup>118</sup>. Además, se sabe que existe la formación diferencial de asas de cromatina entre los *enhancers*, los promotores y el DMR que es dependiente de CTCF. Esta formación diferencial de asas contribuye también a la expresión mono-alélica de los genes *Igf2/H19*<sup>120</sup>.

Estos ejemplos dejan en claro la función que CTCF cumple en términos de la regulación a nivel de dominios génicos, tanto en el establecimiento de marcas epigenéticas como en la estructuración tri-dimensional del genoma al interior del núcleo. La existencia de este tipo de factores, conceptualmente permite dos tipos de enfoques en el estudio de la regulación del genoma; 1) CTCF puede ser entendido como un regulador maestro de la organización del genoma y, 2) juega un papel muy importante en la regulación de los genes o grupos de genes a nivel local.



**Figura 14.** Modelo de regulación del dominio de genes improntados *Igf2/H19*. **A)** En el alelo de origen paterno, la región DMR se encuentra metilada, impidiendo la unión de CTCF lo cual permite que los *enhancers* del dominio (óvalos verdes) puedan *trans-activar* al gen *Igf2*. **B)** Por el contrario, en el alelo materno la región DMR se encuentra desmetilada permitiendo la unión de CTCF y llevando al bloqueo de la acción de los *enhancer* sobre el gen *Igf2* y favoreciendo la *trans-activación* del gene *H19*. **C)** Modelo que describe la formación diferencial de asas de cromatina para la expresión de los genes *Igf2* y *H19* de manera alelo específica. Tomada y modificada de<sup>121</sup>.

## 5.2. CTCF y la metilación del ADN

Como se ha descrito, son varias las funciones atribuibles a CTCF como lo son las de *insulator* y organizador tri-dimensional del genoma, no obstante, poco se sabe de la manera en que CTCF se puede unir en ciertas regiones del genoma. No todos los motivos potenciales de unión a CTCF en el genoma se encuentran ocupados por este factor en las células, además su patrón de unión puede variar entre los tipos celulares<sup>105</sup>.

La secuencia consenso de unión a CTCF contiene CpGs por lo que es una secuencia de ADN potencialmente metilable. La consecuencia de tal metilación es la pérdida en la unión de CTCF *in vitro*<sup>122</sup>, inclusive se ha observado *in vivo* para el caso del DMR del alelo paterno de locus *Igf2/H19* (Figura 13)<sup>118</sup>. Por tal motivo,

la manera en que CTCF puede regular la estructura de la cromatina y la expresión de los genes dependerá del estado de la metilación del ADN en los sitios de unión para dicho factor. De hecho, se ha descrito que la presencia de CTCF puede tener influencia en la metilación del ADN mediante su asociación en un complejo con dos enzimas involucradas en la metilación: la poly(ADP-ribosa) polimerasa 1 (PARP1) y la metil-transferasa de ADN 1 (DNMT1). En esta asociación, CTCF activa a la PARP1, la cual puede entonces adicionar grupos ADP-ribosa a la enzima DNMT1 causando su inactivación y, por consecuencia, manteniendo libre de metilación a las CpGs<sup>123,124</sup>.

De hecho, se ha postulado que CTCF puede regular el establecimiento de los patrones de metilación del ADN<sup>122,125</sup> y que alteraciones en la abundancia de esta proteína modifican la metilación de en regiones promotoras de genes importantes para el ciclo celular y la diferenciación<sup>126,127</sup>. Tal es el caso de los genes supresores de tumores, Retinoblastoma o *Rb* y *Brca1*, donde se ha descrito que en condiciones normales, la unión de CTCF a la región promotora de estos genes mantiene un estado libre de metilación promoviendo su expresión; por el contrario, en células cancerígenas, la ausencia de CTCF induce una hipermetilación de los promotores y consecuentemente el silenciamiento de estos genes controladores del ciclo celular<sup>128,129</sup>.

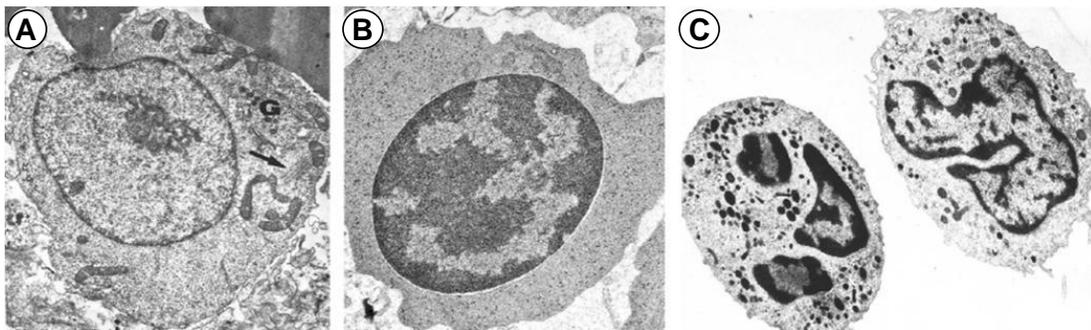
### **5.3. Resumen de las funciones del factor nuclear CTCF**

En la regulación de los programas de expresión génica existe una interdependencia entre la estructura de la cromatina, los elementos genéticos de regulación y los factores de transcripción. El factor multifuncional CTCF es capaz de establecer diferentes estructuras de la cromatina, además de contribuir a la organización espacial del genoma. A un nivel local, la presencia de CTCF puede controlar la expresión diferencial de los genes mediante la formación de asas de cromatina y participando en el establecimiento de patrones de metilación del ADN. Es por ello, que el estudio de la participación de CTCF y sus interacciones con otros factores es primordial para el entendimiento de los procesos fisiológicos y moleculares del genoma en eucariontes. Un aspecto que también es abordado en el presente trabajo.

### **6. La regulación transcripcional en el espacio nuclear**

Como se ha descrito, son múltiples los niveles en que la transcripción de los genes puede ser regulada, desde el reconocimiento de las secuencias reguladoras por los factores de transcripción, el establecimiento de diversos tipos de cromatina y de diferentes marcas epigenéticas como la metilación-desmetilación del ADN. Los diferentes patrones de expresión presentes en cada tipo celular depende, en parte, de variaciones en la estructura de la cromatina y marcas epigenéticas particulares a cada tipo celular. La característica más general para entender la activación y el silenciamiento de los genes es la

formación de estructuras tales como la eucromatina y la heterocromatina, respectivamente. De hecho, cuando diferentes núcleos celulares en interfase son observados mediante la microscopía electrónica de transmisión, la distribución y abundancia de estos dos estados de la cromatina se presenta muy distinta entre tipos celulares diferentes (Figura 15). Esto fuertemente sugiere que la compartimentalización y distribución de la cromatina al interior del núcleo puede tener un efecto en la regulación de los programas de expresión. A continuación se describen diferentes evidencias que sugieren la alta organización y la formación de sub-compartimientos nucleares en la regulación de los vinculados al genoma en los eucariontes.



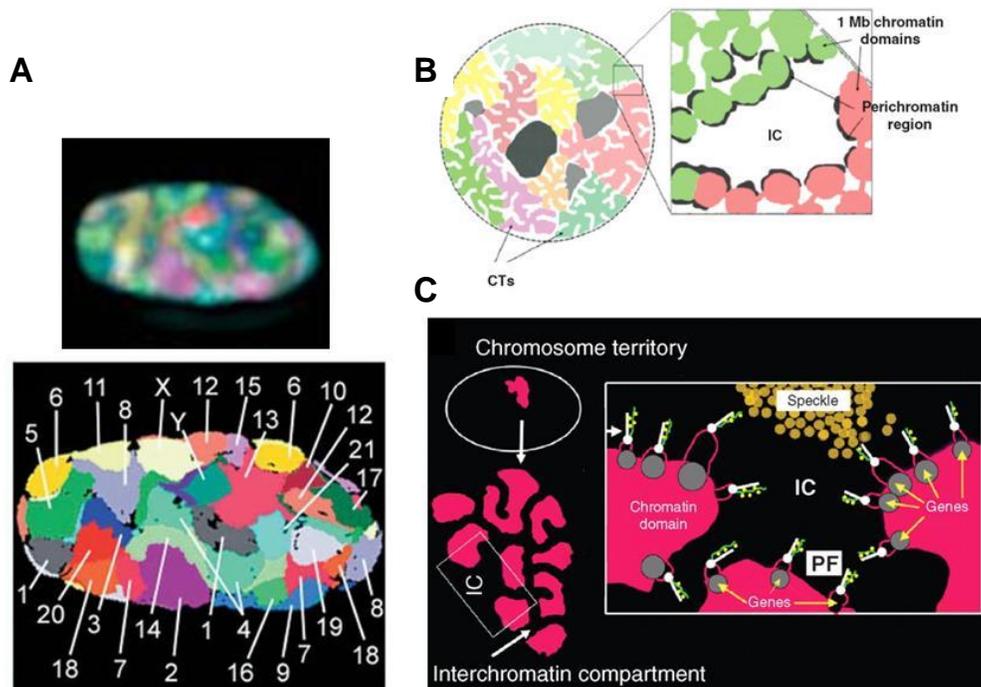
**Figura 15.** Micrografías electrónicas de células hematopoyéticas. **A)** Célula pro-eritroblástica; **B)** Célula eritroblástica tardía y **C)** Célula monocito. Es de notar las diferencias en la arquitectura nuclear entre estos tipos celulares marcadas por las regiones electrón-densas. Tomada de<sup>130</sup>.

### 6.1. Territorios cromosómicos

El acomodo de los cromosomas a través del espacio nuclear no es homogéneo.

Mediante la tinción específica de cromosomas y observación en microscopio

óptico se ha observado que cada cromosoma ocupa un región particular y discreta (Figura 16A)<sup>131</sup>. Ante tal constatación, se propuso que los cromosomas ocupan lugares finitos al interior del núcleo denominados territorios cromosómicos (TCs)<sup>131</sup>. Más aun, se sabe que existe un relación entre la densidad génica presente por cromosoma y su posición en el núcleo, esto es: cromosomas con mayor riqueza génica se localizan en un espacio central en el área nuclear y viceversa<sup>132</sup>. Por ello, además de conformar un TC, existe un posicionamiento espacial en el núcleo por parte de los cromosomas, sin embargo, aun se desconoce el papel de este ordenamiento en la regulación de los genes y sobre todo la manera en que este se establece a lo largo de la diferenciación celular<sup>133</sup>.



**Figura 16.** Organización de la cromatina en Territorios Cromosómicos (TC). **A)** Marcaje individual de cromosomas mediante sondas de ADN acopladas a distintos fluoróforos en células humanas (arriba). Esquema de la reconstrucción indica la posición de cada cromosoma (abajo). **B)** Modelos de TCs y ICs en la regulación transcripcional de los

genes. En la región peri-cromatiniana de los TCs (CTs) y hacia la IC se posicionan los genes a ser activados. **C)** En los canales ICs se encuentran las maquinarias que permiten la transcripción y procesamientos de los genes los cuales son excluidos de los TCs mediante la formación de asas de cromatina. Tomadas y modificadas de<sup>134</sup>.

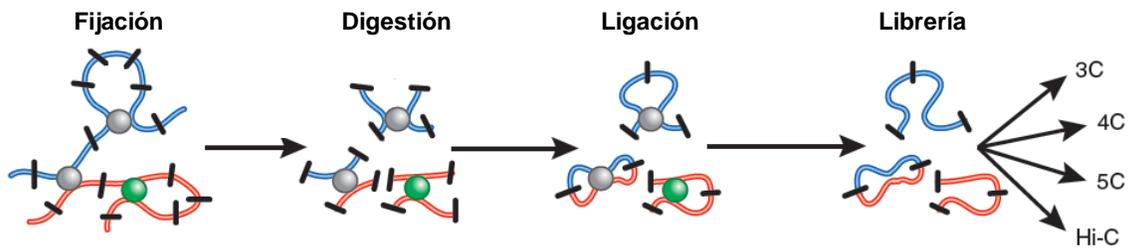
Este arreglo de la cromatina en TCs ha sido ampliamente aceptada como un principio básico en la organización nuclear<sup>134</sup> (Figura 16). En términos de la regulación de los genes se han propuesto modelos en los cuales se postula que la transcripción puede ocurrir en compartimientos intercromatínicos (IC), es decir entre los TCs (revisado en<sup>133,134</sup>). El IC es una red espacial continua en 3D la cual forma “canales” entre los TCs. Se ha propuesto que los genes con actividad transcripcional se localizan en la superficie de los TCs, en la región peri-cromatiniana, donde la transcripción, el procesamiento o *splicing*, y la duplicación del ADN tienen lugar (Figura 16B). Otros modelos proponen que en los ICs los genes activos se posicionan en estas zonas mediante la formación de asas de cromatina que emergen a partir de los TCs y entran en contacto con los complejos de transcripción y procesamiento solubles en los ICs (Figura 16C; revisado en<sup>133,134</sup>).

## **6.2. Nuevas tecnologías para la visualización del arreglo 3D del genoma**

El modelo de TCs y ICs se ha desarrollado a partir de observaciones bajo la microscopia óptica y electrónica. Así mismo, la formación de asas de cromatina fue propuesta con base a predicciones y observaciones indirectas. No fue sino hasta el 2002 cuando se pudo demostrar a un nivel molecular que la cromatina

puede adoptar diferentes conformaciones, las cuales promueven contactos distales mediante la formación de asas de cromatina. Esto fue posible gracias al desarrollo de una metodología denominada: *captura conformacional de cromosomas* o 3C (por sus siglas).

La técnica de 3C se basa en la fijación de las interacciones presentes en el genoma al interior del núcleo mediante el uso del formaldehído como agente químico fijador (Figura 17). Estas interacciones son aisladas mediante el uso de endonucleasas de restricción quedando en asociación aquellas regiones que estaban en contacto constituyendo una macro-molécula con extremos cohesivos de ADN (Figura 17). Posteriormente, estas interacciones son ligadas intramolecularmente por sus extremos cohesivos constituyendo híbridos entre las regiones del genoma que estuvieron en cercanía<sup>135</sup>. La manera de evidenciar las interacciones puede ser mediante PCR con oligonucleótidos de los candidatos de interacción, micro-arreglos de ADN o secuenciación masiva<sup>135-138</sup>. Como resultado de estas innovaciones, ahora se tiene bien demostrado que, como parte de la regulación transcripcional de los genes, existe un arreglo 3D funcional de los elementos genéticos de regulación al interior del núcleo.



**Figura 17.** Captura Conformacional de Cromosomas (3C). Posterior al protocolo de 3C se pueden analizar las librerías de interacción a partir de diversas técnicas experimentales como PCR, micro-arreglos de ADN y secuenciación masiva. Tomada de<sup>139</sup>.

Mediante la tecnología de 3C, diferentes dominios génicos han sido estudiados en términos de su estructuración espacial, como por ejemplo los dominios  $\alpha$  y  $\beta$  globina de humano, ratón y el pollo<sup>32,108,140</sup>, entre otros (ver ejemplo de  $\beta$ -globina; Figura 13). Todos estos estudios han podido elucidar que los contactos diferenciales entre los elementos genéticos de regulación están en estrecha relación con la manera en que los genes se expresan en tiempo y espacio. Más aun, con el empleo de las técnicas experimentales como lo son el 4C, el 5C y el Hi-C (Figura 17)<sup>137,138</sup>, las cuales emplean la secuenciación masiva de ADN para detectar una amplia variedad de interacciones, se ha encontrado una estrecha correlación entre la organización tri-dimensional del genoma y la regulación de la expresión genética. La importancia en el empleo de estas técnicas radica en el hecho de que permiten comprender el arreglo del genoma completo al interior del núcleo de una manera global.

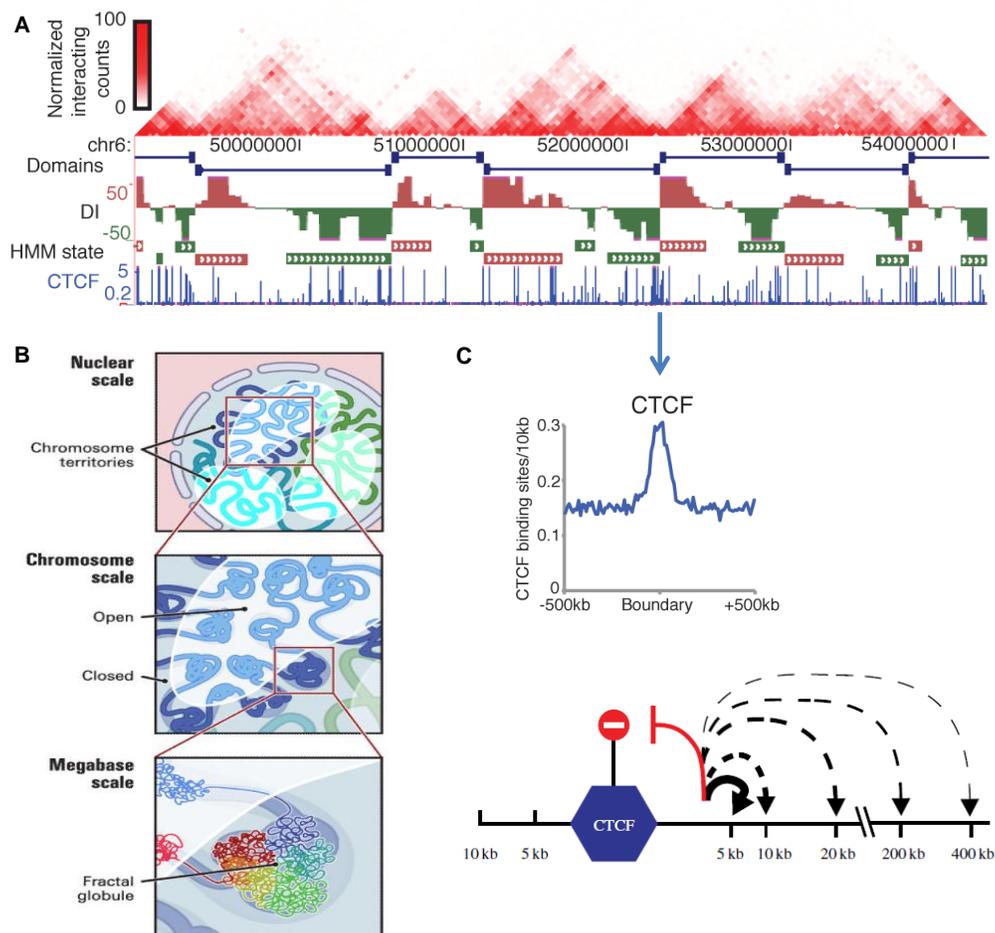
Recientemente, con el empleo de las técnicas de 4C, 5C y Hi-C, se ha podido evidenciar la manera en el que el genoma completo se encuentra

organizado al interior del núcleo de las células<sup>141,142</sup>. Las conclusiones más remarcables de estos estudios dejan ver que las interacciones presentes en el genoma asemejan el modelo de organización en territorios cromosómicos anteriormente propuesto. Además, se pudo elucidar que los cromosomas se estructuran en regiones genómicas de una o más megabases de longitud (Figura 18A). Esta organización de los cromosomas se denominó como *dominios topológicos*. Estos dominios están formados por las interacciones físicas entre las secuencias de ADN contenidas en cada dominio topológico (Figura 18B). En combinación con los estudios de ChIP-seq, también fue posible establecer que CTCF se encuentra enriquecido al interior de estos dominios topológicos, lo cual puede contribuir a la formación de estos dominios al favorecer los contactos en su interior<sup>141</sup> (Figura 18C). Estos dominios se encuentran claramente ‘delimitados’ por la proteína CTCF, lo cual confirma el papel primordial de este factor en la estructuración 3D del núcleo.

Esta manera de organización del genoma se conserva entre las especies puesto que estos estudios fueron realizados en células de humano y ratón donde se obtuvieron resultados similares en cuanto a la organización de los genomas<sup>141,142</sup>. Por tal motivo, resulta de gran relevancia el estudio de las interacciones y procesos que suceden al interior de estos dominios y la importancia de la proteína CTCF no solo a nivel del genoma, sino al interior de dichos dominios. Por otro lado, sería un error atribuirle al factor CTCF la totalidad en la estructuración del genoma, puesto que se ha demostrado que existen otras proteínas involucradas en el arreglo 3D de muchos dominios génicos. Entre estas

se pueden mencionar a los complejos de cohesinas, condensinas, el factor SATB1 y las proteínas que conforman la lámina nuclear, entre otras<sup>32</sup>.

En cuanto al papel en la regulación de la expresión génica, los dominios topológicos además de estar definidos por sus frecuencias de interacción, también poseen características en cuanto al tipo de cromatina. Por ejemplo, un dominio topológico puede estar enriquecido de la marca de represión H3K27me3 y la H3K9me2 y carente de las marcas de cromatina abierta (H3ac y H4K9ac)<sup>141,142</sup>. De esta manera, los genes contenidos en cada dominio serán blancos de activación o represión a la transcripción según sea el caso.



**Figura 18.** Organización del genoma en dominios topológicos. **A)** Mediante la técnica de Hi-C, se pueden determinar los contactos 3D del genoma al interior del núcleo. Las asociaciones entre secuencias se grafican con valores en escala de rojos. Se observa que a lo largo de un cromosoma, el 6 por ejemplo, existen dominios topológicos de organización o asociación (barras azules debajo de las 'pirámides'). Estos dominios topológicos comprenden generalmente 1 Mb donde se favorecen las interacciones entre secuencias contenidas al interior de cada uno de estos (abajo, barras rojas y verdes). **B)** Con el empleo de programas de modelación de estas interacciones, se ha podido deducir que los cromosomas se pliegan en territorios cromosómicos (arriba), constituidos por glóbulos fractales (dominios topológicos de 1 Mb; en medio y abajo). **C)** Mediante ChIP-seq se observa que en los límites de estos dominios topológicos se encuentra frecuentemente unida la proteína CTCF (arriba). El modelo derivado de estas observaciones (abajo), describe que CTCF es capaz de separar dichos dominios y además favorecer los contactos específicos al interior de estos llevando a la regulación de los genes allí contenidos. Figuras tomadas y modificadas de<sup>108,141,142</sup>.

Con el desarrollo una tecnología conocida como *DamID*, el grupo del Dr. Bass Van Steensel en Holanda y colaboradores pudo identificar las secuencias genómicas que preferentemente se encuentran adosadas o en interacción con la periferia nuclear, en específico con la lámina nuclear<sup>143</sup>. Estos experimentos constan de la sobre-expresión y acoplamiento a la periferia nuclear de la proteína Lamina-B modificada para metilar a las secuencias con las que se asocia. El 'mapeo' posterior de las secuencias metiladas por secuenciación masiva permite identificar que regiones del genoma se encontraban en interacción con esta proteína. De esta manera, se pudo establecer que existen dominios de asociación con la lámina nuclear o LADs (por sus siglas en inglés), los cuales constan de regiones genómicas de 1 o más Mb<sup>143</sup>. Estos LADs corresponden mayoritariamente con cromatina cerrada enriquecida con la marca de histona H3K27me3 y en general con una baja densidad génica. De manera interesante, en los bordes de los LADs, además de contener islas CpGs, se encontró con un

alto grado de correlación la unión de CTCF, lo que llevo a proponer a CTCF como un punto delimitador de los LADs<sup>143</sup>. Además, varios de estos LADs delimitados por CTCF, corresponden exactamente con los dominios topológicos descritos por las frecuencias de interacción del genoma (Figura 18, discutido en<sup>108</sup>). Estos resultados dejan en claro que existe un principio de organización y plegamiento 3D del genoma al interior del núcleo y que este se encuentra en estrecha relación con otras estructuras como lo es la lámina nuclear. El papel de CTCF en estos procesos de estructuración parece ser fundamental, sin embargo, aun se desconoce el mecanismo preciso por el cual CTCF cumple esta función así como sus interacciones con la cromatina y otros factores de transcripción.

Además, se ha determinado que la formación de estos dominios puede ser un proceso dinámico a través de la diferenciación celular. Un ejemplo muy claro se puede observar en concreto con los LADs, se sabe que los LADs donde genes inactivos están incluidos en células pluripotentes, pierden asociación con la lamina nuclear (a través de una re-localización) en la diferenciación neural, lo cual lleva a la activación de estos genes<sup>144</sup>.

### **6.3. Resumen de la dinámica nuclear**

La manera en que el genoma se encuentra empaquetado al interior del núcleo no es azarosa. Se ha demostrado que los cromosomas se distribuyen a través de territorios cromosómicos donde la exposición de los genes a la periferia permite su expresión. Por ello, se ha propuesto que la posición que guardan los genes y

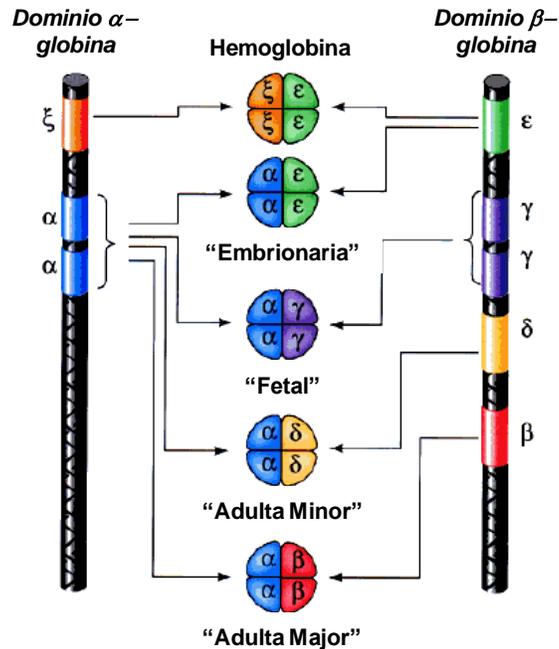
los elementos genéticos de regulación al interior del núcleo contribuye a la ejecución de los programas de expresión específicos de cada celular. El desarrollo de nuevas técnicas experimentales ha proporcionado un acercamiento más detallado al arreglo 3D del genoma al interior. Con esto ha sido posible determinar que la constitución en dominios topológicos dicta en primera instancia el estado de actividad de los genes, por tal motivo es necesario incorporar estudios del arreglo espacial de la cromatina para el entendimiento de la expresión de los genes.

Hasta este punto se han abordado diferentes aspectos involucrados en la regulación de la expresión génica en los eucariontes, desde la secuencia del ADN así como la estructura de la cromatina y el arreglo 3D al interior del núcleo. Como se ha descrito, los dominios multi-génicos resultan de gran interés y son a menudo utilizados como modelos de estudio dado que poseen varios genes cuyos patrones de expresión dependen de la presencia de distintos estados y características de la cromatina. El uso de familias de genes como modelo resulta muy útil para comprender los mecanismos involucrados en la activación de un gen en un tejido o una etapa específica en comparación con otra u otro tejido. Ejemplo de ello son los dominios o familias de genes  $\alpha$  y  $\beta$  globina, los cuales son la parte central de estudio del presente trabajo.

A continuación se abordan aspectos generales y particulares de estos dominios génicos, haciendo énfasis en los mecanismos de regulación epigenética y de estructura de la cromatina del grupo de genes  $\alpha$  y  $\beta$  globina.

## 7. Los dominios de genes globina como modelos de estudio

El transporte y asimilación de oxígeno es un proceso necesario para la vida de muchos organismos. Mediante la producción de moléculas afines a este gas como lo son las hemoglobinas, muchos organismos son capaces de suplir las necesidades de oxígeno a cada una de sus células. Sin embargo, los requerimientos por este gas a través del desarrollo embrionario y de la vida adulta no son homogéneos. Por ello, a lo largo de la evolución, se ha llevado a la generación de diversos tipos de hemoglobinas que presentan diferentes afinidades por el oxígeno y son capaces de suplir las necesidades específicas de cada célula dependiendo de la etapa del desarrollo del organismo y el tejido específico. Para la generación de las diferentes tipos de hemoglobinas, es necesario la expresión coordinada y regulada de múltiples genes contenidos en dos familias: las familias  $\alpha$ - y  $\beta$ -globina. De una manera tal que cada molécula de hemoglobina poseerá tanto una cadena peptídica  $\alpha$ - como una  $\beta$ -globina (Figura 19).



**Figura 19.** Dominios de genes  $\alpha$ - y  $\beta$ - globina. Mediante la expresión transcripcional de estos dominios génicos y su síntesis proteica se pueden generar las diferentes moléculas de hemoglobina las cuales son las responsables de suplir los diferentes requerimientos de oxígeno que se presentan a través del desarrollo de los vertebrados.

Los dominios de genes  $\alpha$  y  $\beta$ -globina proveen dos de los mejores sistemas-modelo para tratar de entender la relación entre los elementos genéticos y la estructura de la cromatina en la regulación de la expresión génica. Y esto se debe a que durante varios años de investigación, se ha observado que la expresión y el silenciamiento de los genes contenidos en estos dominios corresponden con características estructurales particulares de la cromatina y otros factores epigenéticos<sup>145</sup>.

Al nivel de la secuencia del ADN, los genes codificantes para las hemoglobinas se encuentran altamente conservados entre los vertebrados<sup>146</sup>. Sin

embargo, las regiones no-codificantes presentan un menor grado de conservación. En especies como lo son el ratón, el pollo y el humano, se han descrito diversos elementos genéticos de regulación como los son los promotores de los genes, *enhancers*, LCRs, *insulators* y silenciadores, entre otros. Si bien estos elementos de regulación no presentan conservación al nivel de la secuencia del ADN, se han descrito secuencias en estos dominios que presentan actividades de regulación similares entre las especies<sup>147</sup>. Por ello, estos elementos de regulación han sido muy estudiados. No obstante, son menos los trabajos relacionados con el estudio de la cromatina en estos dominios y su papel en la regulación espacio temporal de los genes globina. En el presente trabajo, se aborda, de una manera integral, la relación entre los elementos de regulación y la estructura de la cromatina que permiten la expresión diferencial de los genes  $\alpha$ -globina. Se ha adoptado como modelo biológico al pollo (*Gallus gallus*) ya que los genes  $\alpha$ -globina presentan un patrón de expresión diferencial a través del desarrollo del organismo y la diferenciación eritroide.

### **7.1. Regulación epigenética de los genes globina**

Como se ha mencionado hasta ahora, la intervención de diversos elementos de regulación, la estructura de cromatina y las modificaciones epigenéticas tanto en las histonas (acetilación, metilación, etc.) así como en el ADN (metilación, hidroximetilación, etc.), participan en la regulación de la expresión de los genes en tiempo y espacio. Esto se sabe gracias a que, entre otros, los dominios de genes

$\alpha$ - y  $\beta$ -globina proveen excelentes modelos de estudio para comprender la interrelación entre las secuencias de ADN y las modificaciones epigenéticas. Son diversos los estudios que han llevado a la generación de modelos que tratan de explicar dichos mecanismos de regulación (revisado en<sup>148</sup>). Por ello, es preciso revisar lo que hasta ahora se sabe de la regulación epigenética en estos organismos y con ello generar una visión comparativa con los mecanismos que actúan en otros vertebrados como el pollo.

Tanto en humano como en el ratón, el recambio en la expresión de los genes globina sucede en tres etapas sucesivas a través del desarrollo: 1) etapa embrionaria, 2) etapa fetal y, 3) adulta<sup>149,150</sup>. Es decir, 1) en etapas tempranas del desarrollo las células eritroides primitivas (en el hígado fetal) expresan a los genes globina embrionarios. Posteriormente, 2) en el mismo hígado fetal existe un recambio por células eritroides donde gen embrionario se ve silenciado y los genes fetales y adultos presentan actividad transcripcional. Finalmente, 3) conforme el desarrollo fetal progresa, la hematopoyesis sucede ahora en la médula ósea y las células eritroides solo presentan la expresión de los genes adultos y mantienen silenciados a los genes embrionarios y fetales.

La presencia de factores nucleares en los promotores de los genes  $\alpha$  y  $\beta$ -globina, entre ellos GATA-1, que es un factor eritroide específico GATA-1 considerado “maestro” de la diferenciación ya que además está involucrado en la regulación de diversas vías importantes para la diferenciación eritroide<sup>151,152</sup>, se ha asociado con la formación de una estructura tri-dimensional que facilita la interacción con elementos distales de regulación como lo son los LCRs<sup>13</sup>. Se sabe

que en las diferentes etapas eritroides (embrionaria, fetal y adulta), existen contactos diferenciales con el elemento LCR del dominio  $\beta$ -globina lo que promueve la expresión de los genes de una manera regulada (Figura 12)<sup>117</sup>. Se sabe, por ejemplo, que la asociación de los factores GATA-1 y Ldb1 tanto en los promotores de los genes globina activos como en el LCR fomentan contactos entre dichos elementos de regulación<sup>153,154</sup>. Esto ha llevado a proponer un modelo de regulación donde la unión de dichos factores nucleares (entre otros) lleva a la activación de los genes mediante los contactos distales con el elemento potenciador LCR del dominio.

Aunado a dicho mecanismo de regulación, se ha descrito que la metilación de ADN etapa-específica de los promotores correlaciona con el silenciamiento de los genes llevando al cierre de la cromatina en dichas secuencias de ADN. Dicha estructura de la cromatina impide la unión de los factores nucleares encargados de la activación de los genes, entre ellos GATA-1, Ldb1 y EKLF, entre otros<sup>20,155,156</sup>. Sin embargo, estos estudios dejan abierta la interrogante del papel de la metilación del ADN en los promotores, puesto que se desconoce si es propiamente esta metilación la que interfiere con la unión de dichos factores o es un evento consecuente. Por tal motivo, resulta importante el entender el papel de la metilación del ADN en dicha regulación. En el presente trabajo de tesis se ha abordado el papel de la metilación del ADN y la estructura de la cromatina en la región promotora de los genes  $\alpha$ -globina de pollo.

## 7.2. Competencia por el *enhancer* y silenciamiento autónomo

El modelo actual para la regulación de los genes  $\beta$ -globina de ratón y humano, como se ha mencionado, propone que la interacción física entre los promotores de los genes y el LCR del dominio dicta la actividad transcripcional de cada gen. A este tipo de regulación se le conoce como de “*competencia por el enhancer*”. Esto quiere decir que los promotores de los genes “compiten” por la interacción física con el LCR o *enhancer* del dominio, y en aquellos promotores donde los factores de transcripción antes descritos estén presentes podrán establecer contactos con el LCR (interacciones del tipo proteína-proteína). Este mecanismo de regulación implica que en ausencia de un gen dominante (entiéndase como el que atrae más la acción del *enhancer*) los otros genes se pueden ver mayoritariamente favorecidos por el *enhancer*. Sin embargo, también se ha caracterizado un *silenciamiento autónomo* sobre el gen embrionario (gen  $\varepsilon$  en humano y el gen  $\gamma$  en ratón) el cual involucra la metilación de su promotor y el cierre de la cromatina imposibilitando su interacción con el LCR<sup>155–157</sup>. De hecho, se han descrito secuencias de ADN denominadas silenciadores que se encuentran inmediatamente río arriba de la región promotora de los genes embrionarios. A estas secuencias se les atribuye la adquisición de la metilación del ADN conforme progresa el desarrollo y los genes embrionarios se ven silenciados. Esta metilación es reconocida por complejos de silenciamiento como lo son los reclutados por la proteínas de unión a ADN metilado MBD2<sup>156</sup>. De esta manera, una vez silenciado el gen embrionario (en etapa fetal) o el gen fetal (en la etapa adulta) el *enhancer* solo puede *trans*-activar a los promotores que permanecen

abiertos y con la combinatoria precisa de factores para establecer los contactos físicos.

Además de promover la metilación del ADN de la región promotora del gen embrionario, estas secuencias silenciadoras son capaces de unir factores de transcripción como GATA-1 que en asociación con otras proteínas (FOG, NuRD y Mi2) constituyen un complejo de represión que propician el cierre de la cromatina en la región promotora<sup>155</sup>. Esto ha llevado a proponer el modelo del silenciamiento autónomo de los genes como un proceso de regulación que lleva al recambio de los genes globina a través del desarrollo de los organismos.

Es necesario enfatizar que ambos modelos son muy aceptados en la actualidad y no son excluyentes entre sí. Sin embargo, son necesarios más estudios integrales donde se estudie, por ejemplo, la relación que guarda la metilación de los promotores, la estructura de la cromatina y el efecto en la organización espacial del dominio. Además, se sabe que en los dominios globina de estas especies también se encuentra unido el factor CTCF<sup>25,31,112</sup>. Sin embargo, se desconoce su papel (si lo posee) en el silenciamiento autónomo de los genes y/o en el establecimiento de los contactos distales con el LCR. En el presente trabajo de tesis, se decidió evaluar estos aspectos de una manera integral y considerando el papel de CTCF en la regulación fina de los genes globina tomando como modelo de estudio el desarrollo del pollo y la diferenciación eritroide.

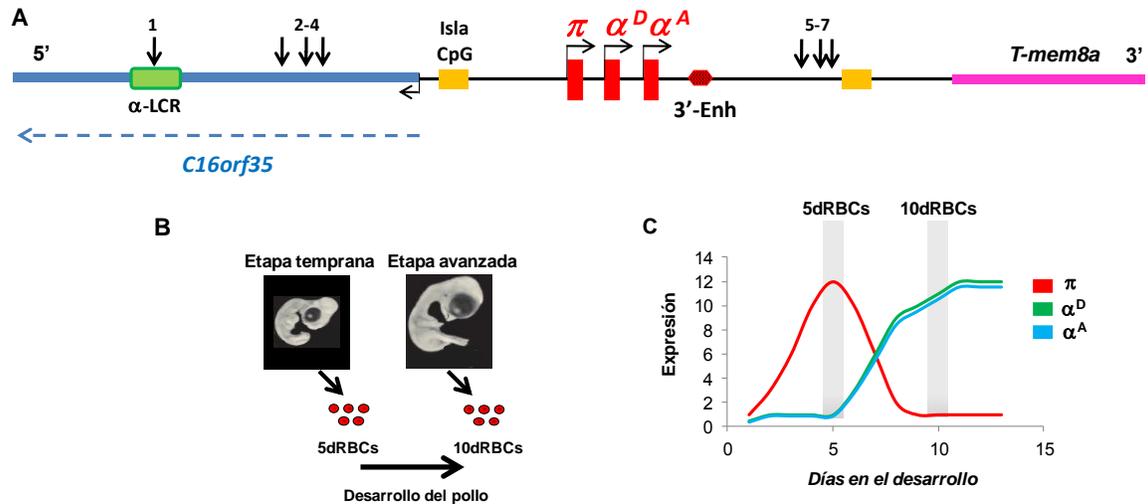
## 8. Antecedentes particulares

### 8.1. El dominio $\alpha$ -globina de pollo

El dominio  $\alpha$ -globina de pollo comprende una región genómica de 40 kb y contiene a tres genes conocidos como:  $\alpha^\pi$ ,  $\alpha^D$  y  $\alpha^A$  (alfa-pi, alfa-D y alfa-A; Figura 20A). En las etapas tempranas del desarrollo del pollo (que comprenden del día 3 al 6; Figura 20B), las células eritroides primitivas aparecen en circulación en el embrión (5dRBCs; del inglés *Red Blood Cells*), y es en estas células donde el gen  $\alpha^\pi$  es activo transcripcionalmente mientras que los genes  $\alpha^D$  y  $\alpha^A$  permanecen inactivos. Por ello, el gen  $\alpha^\pi$  es denominado gen embrionario. A partir del día 7 del desarrollo, el linaje eritroide definitivo o adulto inicia la expresión de los genes  $\alpha^D$  y  $\alpha^A$  (10dRBCs; Figura 20B), y es por ello que son conocidos como los genes  $\alpha$ -globina adultos, mientras que el gen  $\alpha^\pi$  no se expresa en este linaje (Figura 20C). A este recambio en la expresión de los genes globina conforme progresa el desarrollo de los organismos se le conoce como “*switching* o recambio de las globinas” (del inglés *haemoglobin switching*)<sup>148,158</sup>.

Diversos estudios se han dirigido para comprender la participación de los elementos genéticos de regulación en el recambio de las globinas (revisado en<sup>108,148,158</sup>). Se sabe que los promotores de los genes  $\alpha$ -globina se caracterizan por tener una organización modular de sitios de unión al ADN tanto para factores nucleares eritroide-específicos como factores ubicuos. Cada uno de los promotores posee una caja de regulación TATA y por lo general una serie de sitios de unión a los factores YY1, Sp1. Además, las secuencias promotoras

poseen diversos sitios de unión al factor eritroide específico GATA-1, el cual corresponde con un factor nuclear considerado “maestro” de la diferenciación ya que además está involucrado en la regulación de diversas vías importantes para la diferenciación eritroide<sup>151,152</sup>. En 1993, Knezetic y Felsenfeld mediante análisis de células transgénicas que contenían a los promotores de los genes  $\alpha$ -globina de pollo dirigiendo la expresión de un gen reportero (*Luciferasa*), demostraron que la secuencia de los promotores contiene la información necesaria para dirigir la expresión del transgen de manera eritroide específica<sup>159</sup>. Sin embargo, en estas secuencias promotoras no se detectó algún motivo de unión al ADN que reclute a factores que lleve a la expresión etapa específica dentro del linaje eritroide. Es decir, no se encontró la unión de un factor responsable para la expresión temprana del gen embrionario  $\pi$  y la expresión adulta de los genes  $\alpha^D$  y  $\alpha^A$ . Por esta razón, se ha generado la hipótesis de que la estructura de la cromatina y mecanismos de regulación epigenética podrían participar en la regulación de la expresión diferencial de los genes  $\alpha$ -globina. En el presente trabajo de tesis se ha decidido contribuir al conocimiento de la participación de la estructura de la cromatina y los procesos de regulación epigenética en la expresión de los genes  $\alpha$ -globina de pollo.



**Figura 20. Dominio  $\alpha$ -globina de pollo.** **A)** Además de contener a los tres genes que codifican para la cadena alfa de la hemoglobina, el dominio presenta diversas secuencias de ADN cuya función parece estar vinculada a la regulación de la expresión de estos genes. Los DHSs descritos se indican con las flechas, siendo la DHS 1 un probable elemento LCR; las DHSs 2-4 con función de *insulator* dependiente del factor CTCF; las DHSs 5-7 aun no han sido descritas, sin embargo, datos preliminares apuntan a una probable función de *enhancer* e *insulator*. Cabe mencionar que tanto los promotores de los genes, como el elemento enhancer 3'-Enh corresponden con DHSs etapa-eritroide específicas. Las cajas amarillas esquematizan las regiones con alto contenido del di-nucleótido CpG. Este dominio no solo contiene a los tres genes  $\alpha$ -globina; el gen *C16orf35* se encuentra codificado de manera anti-sentido a los globina y su función aun se desconoce. Río abajo del dominio (costado 3') se encuentra presente el gen *T-mem8a*, cuya función se desconoce. **B)** A través del desarrollo del pollo se tiene la presencia de diferentes linajes eritroides en circulación (5dRBCs y 10dRBCs, respetivamente). **C)** Patrón de expresión de los genes  $\alpha$ -globina a través del desarrollo del pollo. Se resalta el recambio de la expresión de los globina entre las células 5dRBCS y 10dRBCs.

Además de los promotores de los genes, en este dominio se pueden distinguir otras regiones con características particulares como los son secuencias de ADN con alto contenido del di-nucleótido CpG o islas CpG (Figura 20A). Sin embargo, no se han abordado estudios para comprender su papel (si presentan)

en la regulación de la estructura de la cromatina del dominio y/o en la expresión de los genes.

Es preciso mencionar que en esta región genómica también se contiene al gen *C16orf35* codificado de manera anti-sentido a los genes  $\alpha$ -globina. Sin embargo, la función de este gen no está relacionada con la vía de las hemoglobinas y su función precisa aun está por determinarse (Figura 20A).

## **8.2. La estructura de la cromatina y la regulación de los genes $\alpha$ -globina de pollo**

A través de la secuencia de ADN de los promotores de los genes  $\alpha$ -globina, se han caracterizado varios sitios de hipersensibilidad al corte con la DNasa I (DHSs; Figura 20A) cuya presencia correlaciona con la actividad transcripcional de los genes globina. Así mismo, las regiones genómicas o codificantes de los genes también se caracterizan por DHSs eritroide específicos. No obstante, se han descrito mas DHSs que no corresponden propiamente con las secuencia de los genes  $\alpha$ -globina. Estudios pioneros del grupo del Dr. Gary Felsenfeld y colaboradores, describieron la presencia de varios DHSs en el dominio  $\alpha$ -globina de pollo. En la región 5' del dominio  $\alpha$ -globina se describió un DHS constitutivo (a 20 kb río arriba del gen  $\pi$ ; Figura 20A), que al evaluar su actividad reguladora en genes reporteros y transfecciones en células de pollo, se observó que actúa como un *enhancer* o LCR (Recillas-Targa, *datos no publicados*). Sin embargo, la función precisa de estas secuencias de ADN no se ha definido y es el objeto de diversos

estudios que actualmente están en curso en el laboratorio del Dr. Félix Recillas-Targa.

Por otra parte, a 14 kb río arriba del gen  $\pi$  se han descrito una serie de tres DHSs eritroide específicas (Figura 20A)<sup>160</sup>. Ensayos de transgenes en células eritroides de pollo determinaron que esta región posee una actividad de tipo *insulator* pues presentan una acción de bloqueo de *enhancer*<sup>161</sup>. Al evaluar su papel de manera *in vivo*, se determinó que estas podrían corresponder con el *insulator* que delimita al dominio  $\alpha$ -globina en su costado 5' pues además se encontró que su actividad de bloqueo de *enhancer* y de delimitador de cromatina es dependiente de la unión del factor CTCF a esta región<sup>162</sup>. No obstante, aun son necesarios más estudios para comprender el papel de estos elementos en la regulación de la expresión diferencial de los genes  $\alpha$ -globina.

De manera interesante, se ha descrito que este dominio presenta un estado de cromatina relajado o abierto en diferentes etapas de la diferenciación eritroide (células eritroblásticas y eritrocitos)<sup>148,158</sup>. Particularmente, al estudiarse el estado de las modificaciones post-traduccionales de las histonas de cromatina abierta (acetilación de las histonas H3 y H4 (H3ac y H4ac, respectivamente), se observó su enriquecimiento a través de varias regiones del dominio<sup>162</sup>. Esta distribución de la H3 y H4ac se presenta hasta los elementos DHSs 5 y 7 (Figura 20A), lo que los sugiere como los extremos del dominio. Otra implicación es que el dominio pudiera estar abierto en las etapas eritroides permitiendo la intervención fina y coordinada de los diferentes elementos de regulación en él contenidos para la regulación específica de los genes  $\alpha$ -globina<sup>162</sup>.

En resumen, el dominio  $\alpha$ -globina de pollo posee diversos elementos de regulación y características en su cromatina que se han visto involucrados en su regulación a través del desarrollo del organismo. La caracterización de diversos DHSs sugiere que existen varios elementos de regulación que pudieran participar en la regulación de los genes. Históricamente, uno de estos DHSs presente en el costado 3' de los genes ha sido ampliamente estudiado dado su conservación en otras aves como el pato. Este corresponde con un *enhancer* cuyas funciones y características se describen a continuación.

### **8.3. El *enhancer* 3' del dominio $\alpha$ -globina**

Dentro de los DHSs caracterizados en el dominio  $\alpha$ -globina de pollo, el DHS presente en el costado 3' a 1 kb río abajo de los genes globina (Figura 20A) ha sido de particular atención por su papel regulador de la expresión de estos genes (revisado en<sup>145,148</sup>). Este corresponde con un elemento tipo *enhancer*, conocido como 3'-Enh, el cual puede potenciar o aumentar la actividad transcripcional de los genes  $\alpha$ -globina de pollo<sup>163-165</sup>. Cuando el 3'-Enh fue colocado en presencia de cada uno de los promotores de los genes  $\alpha$ -globina controlando la expresión del gen reportero de la *Luciferasa*, se observó la potenciación de la transcripción exclusivamente en células eritroides<sup>165,166</sup>. Por ello se considera que este elemento debe estar involucrado en la expresión de los genes en el linaje eritroide. Se sabe que dicho *enhancer* presenta motivos de ADN que son reconocidos por el factor GATA-1, y cuya actividad potenciadora de la transcripción

es dependiente de la unión de dicho factor al *enhancer*<sup>163,165</sup>. La presencia del factor GATA-1 en el 3'-Enh podría estar involucrado en conferirle su actividad de una manera eritroide específica ya que es solo en este linaje donde el *enhancer* puede *trans*-activar a los genes  $\alpha$ -globina<sup>165,166</sup>. La unión del factor GATA-1 en el 3'-Enh participa en el remodelado en la estructura de la cromatina en esta región ya que al alterar la unión de dicho factor se observó la pérdida de la DHS y con ello el cierre de la cromatina en dicha región<sup>163</sup>.

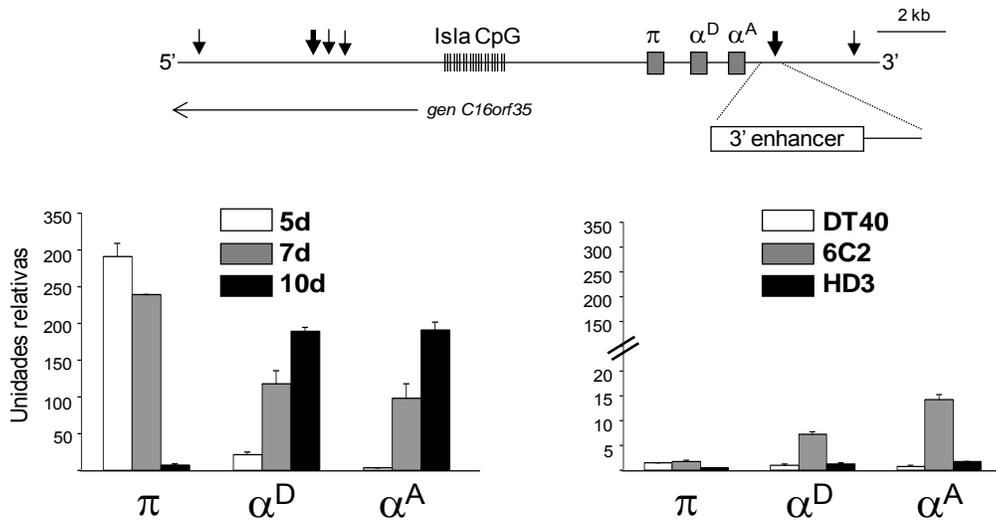
Cabe mencionar que la actividad de este *enhancer* sobre los genes globina ha sido evaluada mediante su presencia en transgenes (*Luciferasa*), por lo cual se desconoce la manera en que actúa en el contexto *in vivo* sobre sus genes blanco. Además, se desconoce la manera en que el 3'-Enh promueve la expresión selectiva de cada gen en las diferentes etapas del desarrollo del pollo. Por ello, en el presente trabajo se decidió estudiar el estado tridimensional del dominio  $\alpha$ -globina, en especial la conformación de la cromatina entre dicho *enhancer* y los promotores de los genes  $\alpha$ -globina como una aproximación para comprender la contribución de este elemento en la expresión diferencial de los genes.

La presencia del 3'-Enh y de otros DHSs sugieren que el mecanismo de regulación del dominio  $\alpha$ -globina de pollo podría encajar en los modelos aceptados de regulación en mamíferos (competencia por el *enhancer* y silenciamiento autónomo). De hecho, se ha estudiado el papel del silenciamiento epigenético mediado por la metilación del ADN en este dominio y se ha descrito que existe un silenciamiento autónomo similar al descrito en el humano y el ratón.

Sin embargo, se desconoce el papel del *enhancer* y otros elementos de regulación en dicho mecanismo de regulación.

#### **8.4. Silenciamiento autónomo del gen embrionario $\pi$ $\alpha$ -globina de pollo**

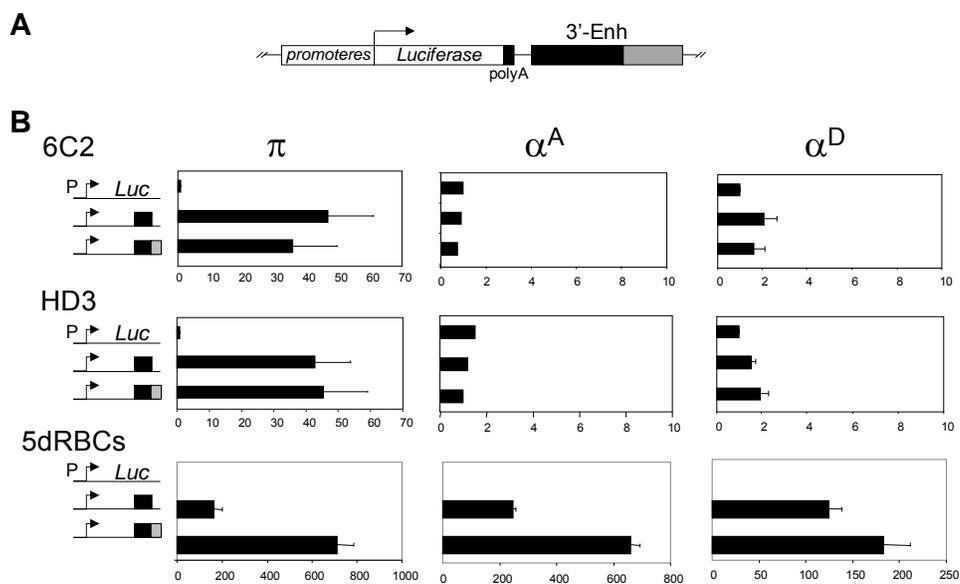
Como un primer acercamiento para entender el papel de la regulación epigenética en la expresión diferencial de los genes  $\alpha$ -globina de pollo, fue el realizado por el Dr. Héctor Rincón-Arano durante el desarrollo de su doctorado en el laboratorio del Dr. Félix Recillas-Targa<sup>165,166</sup>. En primera instancia, se determinó el patrón de expresión preciso de los tres genes  $\alpha$ -globina a través de diferentes etapas del desarrollo de pollo (marcadas por el recambio de las globinas). Además, se incluyó un modelo de células inmortalizadas en diferentes etapas eritroides de pollo (linaje eritroide adulto) con el objetivo de tener una mejor comprensión de la regulación de estos genes en la diferenciación eritroide (Figura 21). Con ello se determinó que los genes presentan expresión diferencial exclusivamente en células terminalmente diferenciadas (RBCs) y no así en las etapas tempranas de la diferenciación eritroide. Lo que sugiere que en las etapas previas a la diferenciación terminal (DT40, 6C2 y HD3; Figura 21) existe un mecanismo de silenciamiento de los genes  $\alpha$ -globina.



**Figura 21.** Expresión diferencial de los genes  $\alpha$ -globina a través del desarrollo del pollo y la diferenciación eritroide. Arriba se esquematiza el dominio  $\alpha$ -globina. Mediante ensayos de RT-PCR semi-cuantitativa se determinó el patrón preciso de expresión en cultivos primarios de eritrocitos colectados de embriones de pollo en diferentes etapas de progresión del desarrollo (5d, 7d y 10d). Así mismo, a través de cultivos de células arrestadas en diversas etapas de la diferenciación eritroide (DT40, 6C2 y HD3). La expresión relativa fue determinada utilizando los niveles de expresión de la  $\beta$ -actina (gen de expresión constitutiva) como referencia. Tomada y modificada de<sup>166</sup>.

Para conocer la contribución de las secuencias promotoras de los genes en dicho patrón de expresión se analizó posteriormente la capacidad de cada uno de los promotores para dirigir la expresión etapa específica de los sus genes  $\alpha$ -globina. Para ello, se clonaron a dichos promotores río arriba del gen reportero de la *Luciferasa* y dichas construcciones fueron transfectadas en diferentes etapas de la diferenciación eritroide de pollo y en cultivos primarios de eritrocitos de embriones de pollo en desarrollo (Figura 22). El sistema de diferenciación eritroide que se emplea consta de cultivos de células inmortalizadas de pollo llamadas 6C2 que corresponde a una etapa conocida como: *unidad formadora de colonias* o

*pro-eritroblástica*, y HD3 (*precursor eritroide o eritroblástico*). En dichas células, ninguno de los genes  $\alpha$ -globina presenta expresión (Figura 21). Como se puede apreciar en la Figura 21, los tres promotores tuvieron la capacidad de dirigir la expresión del gen reportero en células terminalmente diferenciadas de una etapa temprana del desarrollo (5dRBCs), donde el gen endógeno  $\pi$  es el que presenta expresión y no así lo genes adultos  $\alpha^D$  y  $\alpha^A$ . Lo cual indica que las secuencias promotoras no son suficientes para dictar el patrón de expresión diferencial de estos genes.



**Figura 22.** Actividad de los promotores  $\alpha$ -globina en transgenes de la *Luciferasa*. **A)** esquema de las construcciones de ADN donde se incluyó por separado cada una de las secuencias promotoras de los genes  $\alpha$ -globina y el elemento 3'-Enh. Este elemento fue incluido ya que se sabe que es necesario para promover la actividad de los promotores en este tipo de ensayos<sup>165</sup>. **B)** Niveles de actividad transcripcional de los promotores mediante la actividad de la *Luciferasa* (unidades de *Luciferasa*). Tomada y modificada de<sup>166</sup>.

Un aspecto muy relevante de este resultado fue que en las células arrestadas en las diferentes etapas de la diferenciación eritroide (6C2 y HD3), exclusivamente la construcción de ADN que incluye al promotor del gen embrionario  $\pi$ , fue activa en estas células (Figura 22). Lo que indica que este promotor tiene la capacidad de ser activo independientemente de la etapa del desarrollo del pollo y de la diferenciación eritroide. Esto permitió generar la hipótesis de que el promotor del gen embrionario  $\pi$  es más promiscuo (menor grado de selectividad) y pudiera necesitar de otros elementos de regulación y/o mecanismo epigenéticos para silenciarlo en las etapas adultas del desarrollo donde no presenta expresión.

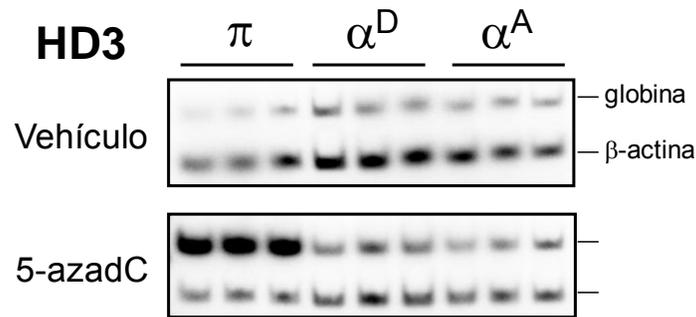
Esta hipótesis encaja en el modelo de silenciamiento autónomo para la regulación de los genes  $\alpha$ -globina. Para comparar dicha hipótesis, posteriormente fueron realizados estudios de ChIP para conocer la presencia *in vivo* de diversos factores de transcripción en los promotores de los genes. Los promotores de los genes adultos  $\alpha^D$  y  $\alpha^A$  presentaron una mayor complejidad en cuanto a la unión de factores de transcripción (unión de los factores GATA-1, Sp1, YY1, NF-1) en comparación con el gen embrionario  $\pi$  que sólo mostró la unión del factor eritroide GATA-1 y del factor ubicuo Sp1<sup>166</sup>. Esta relativa “baja complejidad” del promotor  $\pi$  permite inferir su promiscuidad a ser *trans*-activado en cualquier etapa del desarrollo del pollo, por ello la hipótesis de que un silenciamiento autónomo debe ser dirigido a este promotor para impedir su actividad en etapas tempranas de la diferenciación eritroide y en estadios adultos del desarrollo del pollo.

## 8.5. Silenciamiento autónomo del gen embrionario y metilación del ADN

Como se describió en el humano y en el ratón, el silenciamiento autónomo de los genes embrionarios  $\beta$ -globina está mediado por la presencia de elementos de ADN río arriba de los promotores de dichos genes<sup>155</sup>. A estas secuencias se les atribuye la expansión de una cromatina represora en etapas adultas donde dichos genes están silenciados. La manera en que esta cromatina represora puede extenderse se sabe que está mediada por la presencia de metilación de las CpGs en estas secuencias y en los promotores de los genes<sup>156</sup>. Con esta idea en mente, se analizó el papel de la metilación del ADN en los genes  $\alpha$ -globina de pollo. El primer acercamiento fue la utilización de un inhibidor de la metilación, la 5-aza-2-deoxycitidina (5-azadC), que es un análogo no metilable de la citosina. Al tratar al cultivo celular HD3, donde ninguno de los tres genes son activos, se indujo la activación transcripcional exclusiva del gen embrionario  $\pi$  (Figura 23)<sup>166</sup>. Este hecho, recalca la idea de que el gen embrionario es mayoritariamente silenciado y regulado por la metilación del ADN a diferencia de los genes adultos donde la inhibición de la metilación no indujo su activación (Figura 23).

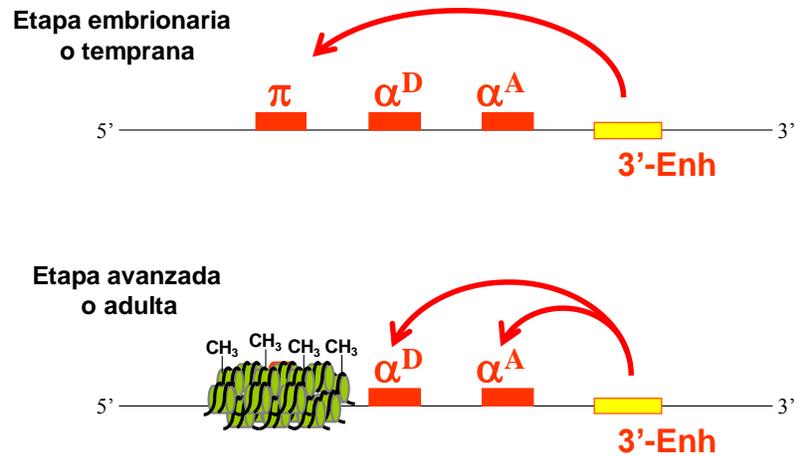
Para corroborar que existe presencia de metilación en el promotor del gen embrionario el segundo acercamiento fue la realización de experimentos con endonucleasas de restricción cuya actividad es sensible a la presencia de metilación (mCpG) en la secuencia de ADN de corte. Al tratar el ADN genómico proveniente de células eritroides de diferentes etapas del desarrollo del pollo y de los cultivos celulares de la diferenciación eritroide, con dichas endonucleasas se reveló la presencia de metilación de ADN en el promotor del gen  $\pi$  en las diferentes

etapas de la diferenciación y en las células eritroides de etapas adultas del desarrollo<sup>166</sup>. Lo cual indica que la metilación del ADN puede llevar al silenciamiento etapa-específica del gen embrionario.



**Figura 23.** Expresión del gen embrionario mediante inhibición de la metilación del ADN. A) Las células eritroblásticas HD3 fueron tratadas con el agente inhibidor de la metilación 5-azadC, y los niveles de expresión de los tres genes  $\alpha$ -globina fueron analizados mediante ensayos de RT-PCR semi-cuantitativa utilizando como normalizador la expresión del gen  $\beta$ -actina. Mediante electroforesis de los productos de la RT-PCR (PCR dúplex del gen globina y  $\beta$ -actina), se observan dos bandas donde la intensidad de la mismas es un indicativo de la abundancia del transcrito y por ello de la expresión del gen. Tomada y modificada de<sup>166</sup>.

Con base en estos en estos resultados, se postuló un modelo de regulación de los genes  $\alpha$ -globina de pollo a través del desarrollo y de la diferenciación eritroide (Figura 24). En dicho modelo, se acentúa el modelo de silenciamiento autónomo del gen embrionario como un mecanismo que permite la acción coordinada de los elementos de regulación (en particular del 3'-Enh) sobre los promotores de los genes  $\alpha$ -globina.



**Figura 24.** Modelo de regulación de los genes  $\alpha$ -globina de pollo. La metilación del ADN ( $\text{CH}_3$ ) y cierre de la cromatina presente en las etapas avanzadas del desarrollo del pollo sobre el promotor de gen  $\pi$ , ejerce un silenciamiento autónomo de dicho gen permitiendo la acción del elemento 3'-Enh sobre los genes adultos  $\alpha^D$  y  $\alpha^A$ . Dicha intervención del 3'-Enh lleva a la actividad transcripcional de los genes adultos en esta etapa del desarrollo del pollo. Tomada y modificada de<sup>166</sup>.

## 8.6. Resumen de la regulación epigenética de los genes $\alpha$ -globina de pollo

Los dominios de genes globina de pollo, al igual que sus homólogos en mamíferos, ofrecen un modelo de estudio muy versátil para la comprensión de los mecanismos de regulación epigenética en la regulación de los genes. La manera de actuar de dichos mecanismos ha llevado a postular dos modelos de regulación de dichos genes; el de competencia por el *enhancer* y el del silenciamiento autónomo. Sería un error asumir que estos modelos son excluyentes entre sí puesto que existen evidencias que apoyan la acción de cada uno de ellos en la regulación de los genes globina. Como se ha revisado para el caso particular del dominio  $\alpha$ -globina de pollo, la metilación del ADN presente en el gen embrionario es un evento muy importante para dictar la expresión diferencial de estos genes.

Sin embargo, se desconoce la participación precisa del elemento distal 3'-Enh y de otros más que llevan al patrón de expresión diferencial de los genes. Por ello, se necesitan de más estudios que puedan encontrar las intersecciones entre dichos modelos de regulación y que permitan el mejor entendimiento del papel de la regulación epigenética en la expresión de los genes globina. Con esto se podría generar una referencia de estudio para otros grupos de genes cuya actividad es diferencial a lo largo de la vasta diversidad de tejidos y células especializadas que constituyen a los organismos multicelulares.

Por tal motivo, en el presente trabajo de tesis se abordaron diversos aspectos de la regulación epigenética del dominio  $\alpha$ -globina de pollo para tener una mayor comprensión de su mecanismo de regulación. En particular, nos preguntamos de qué manera el silenciamiento autónomo es establecido exclusivamente en el gen embrionario  $\pi$ . La intervención de los factores de transcripción en dicho silenciamiento. Y de manera más novedosa, la consecuencia que el silenciamiento autónomo tiene en la conformación tridimensional del dominio y la expresión de los genes. La presente tesis doctoral tiene además la intención de generar una visión más integral de la regulación epigenética incorporando aspectos ambos dos modelos de regulación descritos para los genes globina.

#### IV. Planteamiento del problema.

El grupo de genes  $\alpha$ -globina de pollo poseen un patrón de expresión diferencial a través del desarrollo del organismo y de la diferenciación eritroide. El gen embrionario  $\pi$  es activo a la transcripción en las etapas tempranas del desarrollo del pollo, mientras que los genes adultos  $\alpha^D$  y  $\alpha^A$  presentan actividad en etapas avanzadas. Para la correcta ejecución de este programa de expresión génica, se han descrito diversos elementos genéticos como lo son los promotores de los genes y el elemento 3'-Enh. Además, se ha propuesto que mediante un mecanismo de silenciamiento autónomo dirigido al gen embrionario  $\pi$  se puede obtener este patrón de expresión diferencial de los genes  $\alpha$ -globina de pollo. Sin embargo, se desconoce la manera en que dicho silenciamiento se establece, el cómo se delimita y el efecto que tiene en la conformación espacial del dominio. Por tal motivo, en el presente trabajo de tesis se ha decidió determinar de manera sistemática el establecimiento de dicho silenciamiento autónomo, su extensión y el papel del elemento distal 3'-Enh en las diferentes etapas del desarrollo del pollo y en la diferenciación eritroide.

## V. Hipótesis

El silenciamiento autónomo del gen embrionario  $\pi$  está dirigido por la metilación del ADN en su región promotora y secuencias aledañas, atrayendo la actividad de proteínas que llevan al cierre de la cromatina. Dicho silenciamiento es delimitado por la unión de factores de transcripción constituyendo un sub-dominio de silenciamiento que es específico al gen embrionario. Como consecuencia, la conformación tri-dimensional del dominio favorecerá la actividad potenciadora del elemento 3'-Enh sobre los genes adultos  $\alpha^D$  y  $\alpha^A$  mediante una interacción física entre ellos.

## VI. Objetivo general

Se desea conocer la manera en que el silenciamiento autónomo se establece en la región del gen embrionario  $\pi$  y los cambios en el arreglo tri-dimensional del dominio que permiten la expresión diferencial de los genes  $\alpha$ -globina. Así mismo, la participación de los factores de transcripción en dicho proceso y las modificaciones post-traduccionales de las histonas.

## VII. Objetivos Particulares

- 1) Determinar la distribución precisa de la metilación del ADN en la secuencia promotora de los genes  $\pi$  y  $\alpha^D$  así como de la región genómica intergénica entre estos genes.
- 2) Evaluar la presencia de factores de transcripción que contribuyan al establecimiento/demarcación de la metilación del ADN en la región del gen embrionario  $\pi$ .
- 3) Determinar el estado de la cromatina y de las modificaciones post-traduccionales de las histonas presentes en los promotores de los genes  $\alpha$ -globina en etapas tempranas y avanzadas del desarrollo del pollo y en la diferenciación eritroide.
- 4) Determinar la conformación de la fibra de cromatina entre el los promotores de los genes  $\alpha$ -globina y el elemento 3'-Enh en etapas tempranas y avanzadas del desarrollo del pollo.

## VIII. Estrategia experimental

### 1. Linajes eritroides primitivo y definitivo en el desarrollo del pollo

Para poder cubrir la serie de objetivos planteados en el presente trabajo de tesis, se ocupó como modelo animal al pollo (*Gallus gallus*) para la obtención de los eritrocitos en las diferentes etapas del desarrollo embrionario. En particular, se ha descrito que de manera similar a otros vertebrados, las células eritroides que expresan a los genes globina son el resultado de un proceso de eritropoyesis (diferenciación eritroide) que se divide en dos linajes principales: 1) el linaje eritroide primitivo y, 2) el linaje eritroide adulto. Estos se consideran linajes independientes puesto que cada uno de ellos es el producto de la diferenciación de un precursor eritroide distinto.

Las primeras células eritroides primitivas en el embrión de pollo, aparecen en las islas sanguíneas en el área vascular aproximadamente a las 36 horas después de la gestación. Las células primitivas constituyen las únicas células eritroides del embrión hasta el día 5, cuando las células primitivas se recambian por las definitivas<sup>167</sup>. Las células eritroides definitivas progresivamente van reemplazando a las células primitivas y para el día 8 del desarrollo sólo se observan los eritrocitos definitivos o también conocidos como adultos o maduros<sup>167</sup>.

Con base a estos antecedentes se decidió coleccionar los eritrocitos de embriones de pollo a los días 5 y 10 del desarrollo (5dRBCs y 10dRBCs, respectivamente) como un recurso experimental para estudiar los cambios

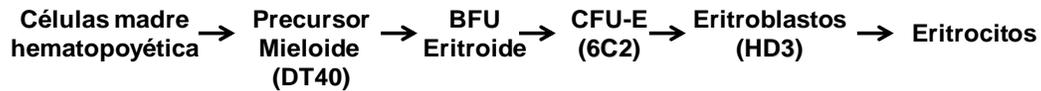
epigenéticos asociados al recambio en la expresión de los genes globina que sucede entre estos dos linajes eritroides. Además, trabajos previos desarrollados en este laboratorio han establecido cuantitativamente los niveles de expresión de los genes  $\alpha$  y  $\beta$ -globina tanto en las 5dRBCs como en las 10dRBCs<sup>145,162,166</sup>.

En las subsiguientes secciones de la presente tesis se hará referencia a los eritrocitos colectados al día 5 del desarrollo del pollo como 5dRBCs o eritrocitos primitivos y a las células colectadas al día 10 del desarrollo como 10dRBCs o eritrocitos definitivos.

## **2. Diferenciación eritroide definitiva en el pollo**

En la transcripción de genes que se expresan en células terminalmente diferenciadas, se ha descrito que diversos procesos de regulación epigenética pueden ocurrir en etapas previas en la diferenciación celular. Estos pueden ser modificaciones post-traduccionales en las histonas y cambios en los patrones de metilación/desmetilación del ADN. Por ello, en el presente estudio se decidió incluir el análisis de diversos eventos de regulación epigenética que pueden ocurrir durante la diferenciación eritroide del pollo. Se conocen diferentes líneas celulares de medula ósea de pollo transformadas con el virus AMV (del inglés, *Avian Myeloblastosis Virus*) y que recapitulan diversas etapas celulares en la diferenciación eritroide de pollo<sup>168</sup>. Estas células poseen la capacidad de auto-renovación y mantener un estado no diferenciado dentro del linaje eritroide<sup>169</sup>. Las células que se utilizaron en el presente estudio son: una línea que corresponde

con un precursor linfoide-mieloide conocida como DT40. Una línea que corresponde con un precursor eritroblástico formador de colonias (CFU-E) denominada 6C2 y una línea eritroblástica que puede ser inducida a la diferenciación terminal eritroide llamada HD3<sup>170</sup> (Figura 25).



**Figura 25.** Diferenciación eritroide de pollo. Se esquematiza la progresión en las diferentes etapas de la diferenciación eritroide y se indican las diferentes líneas celulares transformadas en el presente estudio.

Mediante la colección de sangre circulante de los embriones de pollo en distintas etapas del desarrollo (5dRBCs y 10dRBCs) y la utilización de las líneas celulares de pollo arrestadas en diferentes etapas de la diferenciación eritroide, se puede establecer un sistema de experimental de estudio. En este sistema fue posible abordar distintos aspectos experimentales para comprender el papel de la regulación epigenética en el patrón de expresión diferencial de los genes  $\alpha$ -globina a través del desarrollo del pollo y de la diferenciación eritroide.

## IX. Materiales y Métodos

### 1. Cultivos celulares

La línea celular 6C2 se mantuvo en medio  $\alpha$ -MEM (GIBCO) suplementado con 10% (v/v) de suero fetal bovino (SFB), 2% (v/v) de suero de pollo (ChS), 1 mM de Hepes (pH 7.2) y 50  $\mu$ M de  $\beta$ -mercaptoetanol. La línea celular linfóide de pollo DT40 (célula no eritroide) fue mantenida en medio DMEM (GIBCO) suplementado con 8% de suero fetal bovino, 2% de suero de pollo y 1% de penicilina/estreptomicina (P/S). Los cultivos de la línea celular HD3 de pollo, la cual corresponde a eritroblastos arrestados mediante la infección del virus termo-sensible AEV<sup>169,170</sup>, fueron mantenidos en medio DMEM que contiene 8% SFB, 2% ChS y 1% P/S.

Para los experimentos de diferenciación e inducción de las células HD3, se empleó el agente inhibidor de la proteína cinasa C ISO-H-7 (20  $\mu$ M, Sigma) en el medio de cultivo de las células HD3 con la adición de 10 mM Hepes pH 8.0; la incubación se realizó por 48-72 hrs en una incubadora con una atmósfera de 1% de CO<sub>2</sub> a 42° C lo cual inactiva al virus AEV y propicia la diferenciación terminal de esta línea celular<sup>170</sup>. La eficiencia de diferenciación fue evaluada mediante la tinción de la hemoglobina producida y la actividad transcripcional de los genes globina (ver adelante). Para la tinción de la hemoglobina, las células fueron tratadas con bencidina en medio ácido (0.5% de bencidina (Sigma), 0.3% de H<sub>2</sub>O<sub>2</sub> y 1M de ácido acético)<sup>166,171</sup>.

## **2. Obtención de los eritrocitos de pollo**

Los eritrocitos en circulación de embriones de pollo fueron obtenidos mediante punción de la vena apical de embriones de pollo incubados durante distintos días del desarrollo del organismo (*Gallus gallus domesticus*, Alpes, Puebla, México) con el fin de obtener células rojas de las etapas del desarrollo requeridas (5d y 10dRBCs). Las RBCs fueron colectadas en una solución de PBS y 10% de EDTA como agente anti-coagulante.

## **3. Inmunoprecipitación de cromatina (ChIP)**

En cada ensayo de ChIP, un número de 1 a  $2 \times 10^6$  células fue utilizado para cada anticuerpo a ser evaluado, de acuerdo a trabajos previos<sup>128,163,166</sup>. Las células fueron lavadas dos veces con PBS suplementado con 2% de SFB. Para el entrecruzamiento de las proteínas con el ADN, las células fueron re-suspendidas en una solución de PBS más 1% de formaldehído de concentración final durante 10 min a temperatura ambiente. Para detener el entrecruzamiento, se agregó glicina a una concentración final de 125 mM. Posteriormente, las células son lavadas dos veces con PBS frío suplementado con 1 mM PMSF. Las células fueron re-suspendidas en buffer de lisis (1% SDS, 5 mM EDTA, 50 mM Tris-HCl, pH 8.1 mas inhibidores de proteasas) e incubados por 10 min en hielo. Esta cromatina fue fragmentada mediante sonicación con el fin de obtener fragmentos de ADN en el rango de 200 a 800 pb (este rango es corroborado por electroforesis en gel de agarosa al 1%). Las muestras fueron diluidas 10 veces con la solución de dilución (1% Tritón X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl, pH 8.1 mas inhibidores de proteasas). Con el fin de reducir el ruido de fondo, la solución fue

incubada con perla de proteína A-agarosa (Santa Cruz (SC), Biotechnology) previamente bloqueadas (PBS suplementado con 0.5% BSA, 100 mg/ml de ADN de esperma de salmón incubadas al menos por 12 hrs a 4°C). Las perlas fueron removidas por una centrifugación a 800 g por 5 min a 4 °C. La muestra se dividió en alícuotas de 1 ml (conteniendo cromatina proveniente de 1 a  $2 \times 10^6$  células) y fueron incubadas con el anticuerpo de interés a 4°C en un agitador giratorio durante toda la noche. Posteriormente, 20 ml de perlas de proteína A-agarosa pre-bloqueadas fueron agregadas e incubadas por 4 hrs a 4°C. Las perlas fueron recuperadas por una centrifugación de a 800 g por 5 min. Las perlas fueron lavadas una vez con cada uno de los siguientes buffers: Paro I (0.1% SDS, 1% Tritón X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl, pH 8.1), Paro II (0.1% SDS, 1% Tritón X-100, 2 mM EDTA, 500 mM NaCl, 20 mM Tris-HCl, pH 8.1), Paro III (0.25 M LiCl, 1% NP40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.1) y TE (10 mM Tris HCl, pH 8, 2 mM EDTA), cada uno de los amortiguadores suplementados con inhibidores de proteasas. Las perlas fueron incubadas en 1% de SDS, 0.1 M NaHCO<sub>3</sub> y Proteinasa K (Sigma) a una concentración final de 0.1 mg/ml. La fracción soluble fue recuperada y limpiada con el sistema de purificación para PCR Minelute (Qiagen) o bien mediante el uso de fenol-cloroformo-iso-amílico y precipitación con etanol al 100%.

Los siguientes anticuerpos fueron utilizados: anti-MBD2 N-18 (SC-9397); anti-MeCP2 (Upstate; 07-013); anti-CTCF de pollo<sup>161</sup>; anti-Histona H3 acetilada (06-599, Millipore), anti-Histona H4 acetilada (06-866, Millipore), anti-H3K4me2 (07-030, Millipore), anti-H2A.Z (ab18263, Abcam), y anti-IgG de conejo (12-370, Millipore).

## 4. Métodos de evaluación de los experimentos de ChIP

### 4.1. Análisis semi-cuantitativo radioactivo

Con el material resultante de los experimentos de ChIP de las proteínas MBD2, MeCP2, CTCF y las diferentes modificaciones post-traduccionales en las histonas, se realizaron PCRs dúplex en presencia de [ $\alpha$ - $^{32}$ P] dCTP y los fragmentos amplificados fueron separados en un gel de acrilamida al 6% en buffer TBE al 0.25X. Las PCRs dúplex contienen tanto los oligonucleótidos para la región de estudio (promotor de los genes  $\pi$  y  $\alpha^A$ ), así como el de una región negativa a la presencia del factor o histona en cuestión (promotor del gen  $\beta$ -actina). De tal manera que al ser cargadas en los geles de acrilamida pueden evidenciarse simultáneamente ambos productos de amplificación, eliminando así los sesgos provenientes de la variación entre reacciones de PCR y su cargado. El gel fue secado y expuesto a una pantalla sensible a radiactividad y el registro se hizo en un sistema STORM. Los archivos fueron analizados con el software Image-Quant3. Con el fin de hacer el cálculo semi-cuantitativo, al valor registrado para la región de estudio se le resta el valor proveniente de la región control, esto fue calculado tanto para el material proveniente de anticuerpo de estudio o el anti-IgG. Posteriormente, se dividió el valor del anticuerpo resultante entre el correspondiente del anti-IgG. La siguiente fórmula representa los cálculos mencionados:  $\text{Enriquecimiento} = \frac{\text{Anticuerpo} (V_e - V_c)}{\text{anti-IgG}(V_e - V_c)}$ ; donde  $V_e$  = valor de la región de estudio, y  $V_c$  = valor de la región control.

## 4.2. Análisis cuantitativo mediante PCR en tiempo Real

Los experimentos de ChIP del factor CTCF así como de las histonas H3 y H4 acetiladas fueron evaluados de manera cuantitativa mediante el uso de la técnica de PCR en tiempo real. Para ello, el ADN obtenido de los ChIPs fue amplificado mediante PCR utilizando los oligonucleótidos enlistados en la Tabla 2. Para estas reacciones de PCR se utilizó la Taq polimerasa de ADN Platinum (Invitrogen) en presencia del fluoróforo de intercalación en el ADN *SYBR Green I* (Sigma) a una dilución final de 4X. Los valores de amplificación fueron detectados utilizando el equipo Step-One™ (Applied Biosystems). La manera en que se determinó el enriquecimiento en las diferentes regiones de estudio fue mediante la adaptación del método de  $\Delta\Delta CT$  y es expresado en la siguiente fórmula:  $\text{Enriquecimiento} = 0.5^{(\text{Anticuerpo}(\text{Cte}-\text{Ctc}) / 0.5^{(\text{anti-IgG}(\text{Cte}-\text{Ctc}))})$ , donde Cte= valor de la región de estudio, y Ctc= valor de la región control. Los experimentos de 5hMeDIP fueron analizados mediante este mismo método.

## 5. Ensayo de interferencia del ARNm de MeCP2

Para la reducción de los niveles de abundancia de la proteína MeCP2 se realizaron experimentos de interferencia de la traducción del ARNm del gen MeCP2. Para ello, se sintetizaron oligo-ribonucleótidos de interferencia (ARNi) conocidos como Stealth siRNAs a través de la compañía Invitrogen. El diseño de estos ARNi se realizó de acuerdo a las instrucciones de dicha compañía y utilizando como referencia la secuencia del transcrito de MeCP2 disponible en el sitio web NCBI, número de acceso Y14166<sup>172</sup>. Las células HD3 y dif-HD3 fueron

transfectadas con los ARNi mediante lipofección utilizando el reactivo Lipofectamine 2000 (Invitrogen). Una mezcla de 500 nM de cada ARNi fueron transfectados en un número de  $5 \times 10^5$  células, 24 hrs después se realizó una segunda ronda de transfección. 24 hrs posteriores a la segunda transfección las células fueron colectadas para el análisis de la expresión de los genes globina y western-blot. Las secuencias de ARN de los ARNi es la siguiente: (353): 5'-CCGGAAGUACGACGUCUAUCUCAU-3'; 5'-AUGAGAUAGACGUCGUACUCCCGG-3', y (410): 5'-'UGGACCUGAUCGCGUACUUCGAGAA- 3'; 5'-UUCUCGAAGUACGCGAUCAGUCCA- 3'. Los siguiente ARNi fueron utilizados como control: (C353): 5'-CCGAAGUACGACGUCUAUCUGGCAU-3' y 5'-AUGCCAGAUAGACGUCGUACUUCGG-3'.

## **6. Extracción de ARN y análisis mediante RT-PCR y PCR semi-cuantitativa**

Las células fueron re-suspendidas en Trizol (Invitrogen) y el ARN fue aislado siguiendo las instrucciones de la compañía. El ARN fue disuelto en agua libre de RNasas y cuantificado utilizando el equipo Nano-drop. Para la generación del ADN complementario (ADNc), la reacción de transcripción reversa se realizó utilizando el sistema de Transcripción Reversa (Promega). 1  $\mu$ g de RNA fue mezclado con *Oligo-DT* en un volumen de reacción de 20  $\mu$ l.

Para el análisis de la expresión de los genes en los experimentos de interferencia de la proteína MeCP2, se realizaron amplificaciones por PCR fueron realizadas utilizando la ADN polimerasa Hot-Start (Qiagen) en presencia de [ $\alpha$ -

$^{32}\text{P}]\text{dCTP}$ . La reacción de PCR fue diseñada para obtener de manera reproducible condiciones lineales de amplificación. Los fragmentos fueron separados en un gel de acrilamida al 6% en TBE al 0.25X. El gel fue secado y expuesto a una pantalla sensible a radiactividad y el registro se hizo en un sistema STORM. Los archivos fueron analizados con el software ImageQuant3. Con el fin de hacer un cálculo semi-cuantitativo, se realizó una PCR dúplex considerando dos pares de oligonucleótidos (Tabla 2): un par que corresponde al gen de interés y otro para un gen constitutivamente expresado ( $\beta$ -actina). Los oligonucleótidos para amplificar los genes  $\alpha$ -globina fueron diseñados sobre la secuencia de empalme de los exones para asegurar el análisis de ARN mensajero maduro y evitar la amplificación de secuencias de ADN genómico.

## **7. Análisis de expresión mediante RT-PCR y PCR en tiempo real**

Para el análisis de expresión del gen  $\alpha^D$  en los experimentos de diferenciación de las células HD3, se ocupó el método de RT-PCR y PCR en tiempo real. La obtención del ARN y su posterior conversión a ADNc se realizó de la manera anteriormente descrita. Sin embargo, en estos experimentos el ARN fue retro-transcrito utilizando *random primers*. Para la PCR cuantitativa en tiempo real, se utilizó la ADN polimerasa Platimun (Invitrogen) en presencia de *SYBR-Green I* (Sigma) disuelto en DMSO (4X dilución final). Las reacciones de PCR se realizaron en el sistema de detección StepOne<sup>TM</sup> (Applied Biosystems). Los niveles de expresión relativa del gen  $\alpha^D$  fueron determinados utilizando como

referencia los niveles de abundancia del transcrito del gen ribosomal 28S (Tabla 2). El método de análisis  $\Delta\Delta\text{CT}$  fue utilizado (Applied Biosystems).

## **8. Western blot**

Las células transfectadas con los ARNi contra MeCP2 ( $5 \times 10^5$ ), fueron centrifugadas a 2,300 g a 4°C. La pastilla celular se re-suspendió con 100 ml de buffer de lisis RIPA (150 mM de NaCl, 1% Triton X-100 (Sigma), 0.5 % deoxicolato de sodio, 0.5% de EDTA y 50 mM de HEPES) que contenía inhibidores de proteasas (50 mg/ml Aprotinina, 25 mg/ml Leupeptina, 25 mg/ml Pepstatina, y 2 mM PMSF). La mezcla se incubó en hielo durante 10 min y posteriormente se centrifugó a 11,000 g durante 5 min a 4 °C. La concentración de las proteínas se determinó por el método de Bradford.

Para la separación de proteínas de extractos celulares totales, se realizaron geles de poliacrilamida desnaturalizantes. Los geles al 10% de acrilamida se prepararon mediante la mezcla de 1.25 ml de solución de corrida (48 mM de HCl, 3 M Tris base, pH= 8.9), 2.5 ml de la solución de acrilamida al 40 % (38.96 g de acrilamida (Invitrogen), 1.04 g de bis-acrilamida (Invitrogen) en agua des-ionizada), 100 ml de 10% dodecil sulfato de sodio (SDS), 100 ml de 10% persulfato de amonio (APS), 10 ml de TEMED y llevando todo a 10 ml con agua des-ionizada. Para obtener el gel concentrador (3.5% de acrilamida) se mezclaron 1.25 ml de buffer concentrador (500 mM Tris base pH= 6.7 en agua des-ionizada), 0.75 ml de acrilamida 40%, 100 ml de 10% SDS, 100 ml de 10% APS, 10  $\mu\text{l}$  de TEMED y llevando todo a 10 ml con agua des-ionizada. Las

muestras de extractos proteicos son hervidas durante 5 min a 100°C y cargadas en los geles.

Para las inmunodetecciones, las proteínas fueron transferidas del gel a una membrana de di-fluoruro de polivinilideno (PVDF). La membrana de PVDF con las proteínas transferidas, se lavó dos veces con PBS y se dejó en una solución bloqueadora (5% leche Carnation libre de grasas, 0.5% BSA y 0.05 % Tween-20 en PBS) una hora a TA. Posteriormente, la membrana se incubó durante 1 h a temperatura ambiente con el anticuerpo anti-MeCP2 diluido en la

Solución bloqueadora. La membrana se lavó cinco veces con PBS y se deja incubando durante 1 hr a temperatura ambiente con el anticuerpo secundario (anti-IgG de conejo marcado con peroxidasa) diluido en la solución bloqueadora. La membrana se lava cinco veces con PBS y se transfiere a un recipiente que contenga el substrato quimio-luminiscente “SuperSignal” (PIERCE) durante 5 min. Finalmente, la membrana se expone por 10 min o más a una placa fotográfica para rayos X. Para normalizar la abundancia de la proteína, se utilizo la abundancia de la proteína  $\beta$ -actina mediante su inmunodetección en la misma membrana bajo el protocolo descrito.

## **9. Análisis de la metilación del ADN mediante conversión con Bisulfito de Sodio**

De 1 a 3 mg de ADNg fueron convertidos mediante el tratamiento con bisulfito de sodios. Brevemente, El ADN digerido fue desnaturalizado a 95°C por 5 min,

enfriado en hielo por 2 min e incubado con 0.3 M de NaOH (volumen final: 20  $\mu$ l) a 37°C por 15 min. Para proteger de oxidación a la mezcla de bisulfito a utilizar, el ADN fue tratado con una solución de hidroquinona (0.5 M). Posteriormente, se adiciono la solución de bisulfito de sodio (pH 5.0; [1.7 M]) con incubación en obscuridad a 50 °C. durante toda la noche. El ADN fue limpiado con columnas de purificación Wizard (Promega) de acuerdo a las indicaciones de la compañía. El ADN purificado fue incubado con NaOH a un concentración final de 0.3 M por 30 min. a 37° C. Posteriormente, se utilizó el sistema de purificación de ADN ya mencionado para eliminar el exceso de NaOH. Para la amplificación del ADN convertido con bisulfito de sodio, se diseñaron oligonucleótidos especiales (Tabla 2) los cuales poseen la característica de hibridar específicamente con el ADN convertido (en la sustitución de las citosinas por uracilos como resultado del tratamiento). Para las PCRs se utilizaron tanto las polimerasas de ADN *Hot-Start* (Qiagen) y *Platinum* (Invitrogen).

Los productos de PCR fueron purificados mediante geles de agarosa y su extracción con el kit *QiaexII Gel Extraction* de acuerdo a las instrucciones de la compañía (Qiagen). EL ADN fue sub-clonado en el vector de clonación y secuenciación pGEM-T-Easy (Promega). A partir de la transformación de los fragmentos incluidos en el vector, se seleccionaron de 10 a 12 colonias para su análisis por secuenciación Sanger.

## **10. *Dot-Blot* de ADN**

Con la intención de evaluar la abundancia de la marca en el ADN de 5hmC tanto en las células RBCs, así como en las líneas celulares de pollo, se empleó la

técnica de *dot-blot* como se describe a continuación. El ADNg proveniente de estas células y de los controles positivos de ESC y cerebro de ratón fueron cargados en una membrana HybondN<sup>+</sup> (GE, Healthcare). Posteriormente, el ADN fue fijado a la membrana mediante su exposición a luz UV en el equipo Stratalinker 1800 (Stratagene). La membrana fue bloqueada con una solución buffer PBS-leche al 5% durante una hora a temperatura ambiente. Posteriormente, la membrana fue incubada con el anticuerpo anti-5hmC (Active Motif 39769; dilución 1:10,000). La membrana fue lavada con PBS-Tween (0.05 %). Se utilizó un anticuerpo secundario acoplado a la HRP para la reacción de quimioluminiscencia (Thermo Scientific; dilución 1:5,000) y su exposición en films foto-sensibles.

## 11. Plásmidos

Lo plásmidos utilizados fueron generados a partir de la amplificación por PCR de las regiones genómicas correspondientes al promotor  $\alpha^D$  (300 pb) y la región intergénica 5' río arriba a este promotor (aproximadamente 1 kb). En la tabla 2 se enlistan los oligonucleótidos y su secuencia de ADN utilizados en las reacciones de PCR. Para las construcciones que contienen intactos al promotor  $\alpha^D$  y la región intergénica ( $p\alpha\Delta$  y  $\alpha DIG$ , respectivamente), se amplificaron por PCR las secuencias de ADN del promotor  $\alpha^D$  ( $p\alpha D$ ) y la del promotor mas la región intergénica ( $\alpha DIG$ ). Cada uno de estos fragmentos amplificados se sub-clonaron en un vector de clonación comercial (pGEM-T-easy; Promega) donde quedan flanqueados por diversos sitios de cortes a endonucleasas para su posterior

introducción en vectores de expresión. Posteriormente, cada fragmento fue sub-clonado por separado en la región 5' del MCS del vector de para análisis de actividad transcripcional pEGFP-1 (Clontech) utilizando los sitios de corte AgeI y HindIII. Para la construcción plasmídica que contiene la eliminación de los sitios de unión a CTCF ( $\alpha\Delta C1$ ), se generaron por separado dos fragmentos de ADN de la siguiente manera: se amplificó por PCR la región inmediatamente río debajo al motivo de unión a CTCF-1 y hasta la secuencia 3' donde se incluye el TSS del gen  $\alpha^D$  (Figura 40). De igual manera, se amplificó por PCR la región inmediatamente río arriba a cada motivo de unión a CTCF-1 hasta una secuencia 5' en la región intergénica con el gen  $\pi$ . Ambos fragmentos se sub-clonaron en el vector de clonación pGEM-T-easy. Finalmente, los fragmentos tanto río arriba como río abajo al sitio de unión a CTCF-1 fueron introducidos en el vector de expresión y análisis de la transcripción pEGFP-1 utilizando los sitios de corte AgeI y HindIII.

## **12. Transfecciones estables y tratamiento con fármacos.**

Las construcciones plasmídicas fueron introducidas en la línea celular HD3 mediante la transfección estable de 1  $\mu\text{g}$  de cada plásmido *linearizado* en las células HD3 mediante el uso de Lipofectamine plus 2000 (Invitrogen) bajo condiciones de selección de las células resistentes al antibiótico Geneticina (0.8 mg/ml; Calbiochem). El porcentaje de células fluorescentes, donde la proteína GFP es expresada, fue determinado mediante citometría de flujo utilizando parámetros previamente descritos para la línea celular HD3<sup>162,173</sup>.

Para los tratamientos farmacológicos de las células transfectadas con los plásmidos anteriormente descritos, el inhibidor de desacetilasas de histonas, Tricostatina A (TSA), fue utilizado en las células HD3 transfectadas con los transgenes a 2.5 ng/ml en medio de cultivo por 24 h. El inhibidor de la metilación del ADN, 5-aza-2'-deoxicitidina, fue utilizado a 3  $\mu$ M por tres días, realizando cambios diarios de dicho reactivo<sup>166</sup>. Los análisis de expresión de la proteína GFP se realizaron mediante FACS como se ha descrito anteriormente<sup>128,162,173</sup>.

### **13. Captura Conformacional de Cromosomas (3C)**

Las células RBCs fueron colectadas de embriones de pollo en una solución de buffer PBS-EDTA 10% (0.5M pH 8) en hielo. Las células se centrifugaron a 1100 RPM a T.A. y se re-suspendieron en buffer PBS-SFB 10%. La fijación de las células se realizó con una solución de PBS-SFB (10%)-Formaldehído (1%) por 10 min. a T.A. en rotación moderada. La fijación fue detenida con al adicionar glicina a una concentración final de 0.15 mM e inmediatamente puesta en centrifugación a 300g por 10 min. a 4° C. Las células se lavan una vez con PBS frío. Para la obtención de los núcleos, las células se re-suspendieron en 1-10 mL de buffer de lisis frío (50mM TRIS, 150mM NaCl, 5mM EDTA, 0.5% NP-40, 1% Tritón-100) durante 40 min. La obtención de los núcleos se verifica en el microscopio. Posteriormente, los núcleos se centrifugan a 700g durante 10 min a 4° C. y el buffer de lisis es removido. Los núcleos se lavan con 1 mL de PBS frío por cada experimento de 3C (10-20x10<sup>6</sup> células) y se distribuyen en alícuotas de 1 mL en tubos *ependorf*. Las muestras se centrifugan a 800g durante 2min a 4° C, el PBS es removido y los núcleos pueden ser almacenados mediante su congelación

inmediata en nitrógeno líquido y puestos a -80 C. Los núcleos son re-suspendidos en 500  $\mu$ l de del buffer de la endonucleasa de elección [1.35X] (New England Biolabs) y centrifugados a 800g a T.A. por 2 min. Para la digestión con la endonucleasa de interés, los núcleos se re-suspendieron en 500  $\mu$ l de 1.2X de buffer de digestión, se calientan a 37° C. durante 5 min y se les agrega 15  $\mu$ l de SDS al 10% y se incuban por una hora a 37° C en agitación moderada. Para secuestrar el SDS, se agregan 75  $\mu$ l de Tritón-X-100 al 20% y se dejan en agitación a 37° C. por una hora. La digestión se realiza con la adición de 200 U de la enzima de elección durante 2-4 hrs a 37° C. en agitación moderada. Posteriormente, se agregan de 200 a 400 U de la enzima y se dejan las muestras toda la noche en las mismas condiciones. Para evaluar la digestión, se toma una muestra de 10  $\mu$ l y se revierte la fijación por 2hr a 65° C. con Proteinasa K. El patrón de digestión del ADN así obtenido es verificado en gel de agarosa al 0.6%. Adicionalmente, con el ADN de esta muestra se realizan PCRs en tiempo real para evaluar cuantitativamente la eficiencia de digestión (ver más adelante). La enzima se inactiva mediante la incubación de la muestra a 65° C. (si es el caso para la enzima de acuerdo a las especificaciones de la compañía) por 20 min. La muestra es diluida para fomentar una ligación intra-molecular al llevarla a 7 ml en presencia de 1X del buffer de Ligacion (NEB). Se adicionan 400 U de T4 Ligasa (NEB) y se incuba a 16° C. toda la noche. La eficiencia de ligación es determinada mediante una muestra de 100  $\mu$ l donde se ha revertido la fijación a 65° C y Proteinasa K en un gel de agarosa al 0.6%. Una vez verificada la ligación, se realiza la limpieza y extracción del ADN (ahora nombrado ADN-3C). Para ello, se

revierte el entre-cruzamiento de la muestra mediante la adición de 30  $\mu$ l l de Proteinasa K (10mg/mL) y su incubación a 65° C. en baño maría toda la noche. La extracción del ADN-3C se realiza con fenol-cloroformo y precipitación con etanol al 100%. Se realizan dos lavados con etanol al 70%. La muestra puede ser re-suspendida en 100-200  $\mu$ l de H<sub>2</sub>O. Se analizan las frecuencias de interacción mediante PCR en tiempo real (ver más adelante).

Para evaluar cuantitativamente la eficiencia de digestión anteriormente citada, se diseñaron oligonucleótidos que se encuentran flanqueando al sitio de restricción de interés (en los elementos de probable interacción) y se realizan experimentos de PCR en tiempo real con SYBR-GREEN I como se ha descrito anteriormente. Utilizando el ADN purificado durante el paso de digestión con la endonucleasa y un ADNg no digerido como control, se realizan los siguiente cálculos<sup>174</sup>: % digestión =  $100 - 100/2^{((Ct_R - Ct_C)D - (Ct_R - Ct_C)UND)}$ ; donde D= muestra digerida; UND= ADNg no-digerido; Ct<sub>R</sub>= valor de Ct de amplificación de la región de estudio; Ct<sub>C</sub>= valor de CT de la amplificación de una región control donde no existe corte a la endonucleasa.

Para el análisis de las frecuencias de interacción, se diseñó en la región del elemento 3'-Enh la siguiente sonda Taqman<sup>TM</sup> (Applied Biosystems) acoplada al fluoróforo FAM-6: 5'- 6-FAM/ACTCAGCAC/ZEN/CTGGCAGGTTTTAC- 3'-IABkFQ (IDT Technologies), Dicha sonda Taqman es capaz de hibridar con el ADN que es amplificado por PCR en tiempo real utilizando el oligonucleótido ancla Enh-DpnII-R (Tabla 2) y en combinación con alguno de los otros oligonucleótidos en los elementos de probable interacción (en los promotores de los genes  $\pi$  y  $\alpha^D$ ;

Tabla2). Se utilizó la ADN polimerasa *Platinum* (Invitrogen) para los experimentos de PCR en tiempo de real utilizando las combinatorias de oligonucleótidos del elemento 3'-Enh (Reverse-oligonucleótido constante) y en combinación con los otros oligonucleótido de las regiones de interés (forward). El cálculo de la frecuencia de interacción se realizó de acuerdo a publicaciones previas<sup>171,174,175</sup>.

#### 14. Lista de Oligonucleótidos utilizados en los experimentos de PCR

<b>Tabla 2. Oligonucleótidos utilizados</b>	
<b>Primer</b>	<b>Secuencia (5'- 3')</b>
<b>Análisis de la metilación del ADN</b>	
PiBisE-F	CCTGGAATTCGTTTGGTGTGTATATTTGTGAA
PiBisE-R	CCTGGAATTCTACTAACTCAAACCTCCCTTCAA
3PiBisE-F	CCTGGAATTCTAGGGTGGATTTAGTGAATTTTA
3PiBisE-R	CCTGGAATTCTCCCTAACCCAATATAAAAACC
IntGR1-2F	CCTGGAATTCATTGGGTTAGGGATATGTATTTA
IntGR1-2R	CCTGGAATTCAAATATTCACCTTATCCAAACCC
IntGBisE-R2F	CCTGGAATTCCTTAGTGTGGTATTTTTTTGA
IntGBisE-R2R	CCTGGAATTCCTCACTAAAAAACAAAAATATAAC
IntGBisE-R3F	CCTGGAATTCGAATAAGGTAGTAAGTTTAGGTT
IntGBisE-R3R	CCTGGAATTCAAAAAAATCCCAAACAACCCC
PaDBisE-F	CCTGGAATTCAGGGTAGGATAGGGTAGGGTA
PaDBisE-R	CCTGGAATTCCTTCTCCCAAACCTACTAAAT
IntGBis-R4F	GTTATATTTTTGTTTTTTTAGTGAG
IntGBis-R4R	AACCTAACTTACTACCTTATTC
BS- <i>GFP</i> -R1	CCCTTACTCACCATAATAAC
BS- <i>GFP</i> -R2	ACCAAATAAACACCACCCC
<b>Análisis de Expresión</b>	
aD-F	GCTCTGACTAGGATGTTCCACC
aD-R	CTGCGACAACAGCTTGAAATTG
28s-F	CCTAGCAGCCGACTTAGAACTG
28s-R	GCTCAACAGGGTCTTCTTTCC
BActinF2	CCAGACATCAGGGTGTGATG
BActinR	GAACACGGTATTGTCACCAACTGG
<b>Ensayos de CHIP y 5hMeDIP</b>	
FII-F	GGTTGAAGAAAAGAAGCAGGC

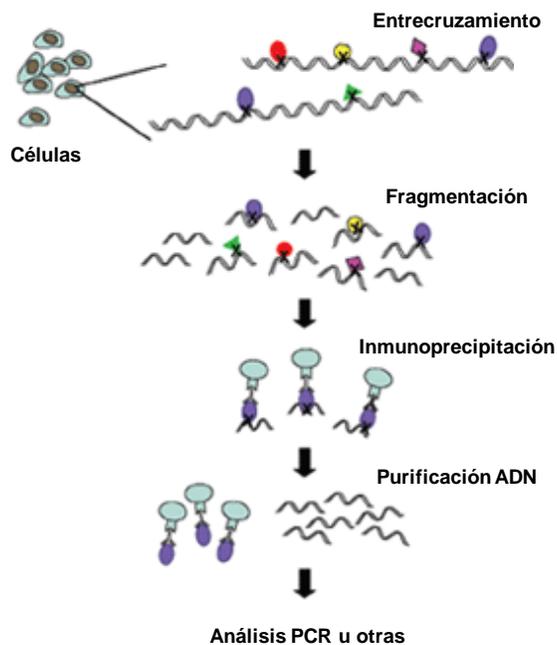
FII-R	CGTTCAGACGAAAGCGATCC
H1-F	GGAAGTAGTTGGCAAGGTCCTCT
H1-R	TCTTCTGCCCTGCCCCGTAT
aDprom-F	CTGCTCTCAGCATTGCACAG
aDprom-R	GGGCTGGTGGCTGGTGG
aPiprom-F	AGAATCACACAGCCACAAATC
aPiprom-R	GTGCTTCCTGAGATGGTAAC
EnhCTCF-F	CAGGAACCATCAGCACTTG
EnhCTCF-R	CAGCTCTTTATCAGCCACC
pGF-CVQ-R	CTCCTCGCCCTTGCTCAC
<b>Plásmidos</b>	
InG-BglII-F	AGATCTGCACTACAAGGTGCACTGC
aD2-AgeI-R	ACCGGTGGTGGCAGCTGGTGGGG
DCTCF1-HindIII-R	AAGCTTGCCTGCTGCCCTGCC
DCTCF1-HindIII-F	AAGCTTCGCTGAGGCACCGCCGCTC
DCTCF2-HindIII-F	AAGCTTCAGGGTGCAGCCTTGATG
DCTCF2-HindIII-R	AAGCTTCTTGGCCTCGTGGTGCTG
EnhNheI-2F	CCTGGCTAGCGAGTGGGCAGACAGGCTG
EnhNheI-R	CCTGGCTAGCCGTGTGGCGAGGAATGAAG
<b>Captura Conformacional de Cromosomas (3C)</b>	
PiDpn-3CF2	AGAATCACACAGCCACAAATC
PiDpn-3CR	GTGCTTCCTGAGATGGTAAC
adDpn-3CF2	ACAGGACAGTGGCTGCCAA
adDpn-3CR2	CTGCCAGTTGAGTTGACCC
EnhDpn-3CF2	GGGCAGCAGATAGCCTCG
EnhDpn-3CR2	TCTCAAACCTCCATCTCTGC
ERCC3Dpn-3CF	CAGAAGCGTTGTATCCATTTG
ERCC3Dpn-3CR	TCAAACATCTTCCTCAGGC
ACTB1-3CF	GTGTTACCGCTGGCCCTG
ACTB1-3CR	CAGCCCAACCTCCCGCTA

## X. Resultados

### 1. Reclutamiento de proteínas de unión a ADN metilado en el silenciamiento autónomo del gen embrionario $\pi$ .

Como se ha descrito en la sección de antecedentes particulares, en la etapa temprana del desarrollo el gen embrionario presenta actividad transcripcional, mientras que los genes adultos  $\alpha^D$  y  $\alpha^A$  permanecen silenciados (Figura 20C). En etapas tardías o avanzadas del desarrollo del pollo, el gen  $\pi$  es silenciado y los genes adultos son activos transcripcionalmente. La manera en que el gen embrionario  $\pi$  es inactivo a la transcripción fue descrita como un silenciamiento autónomo, es decir, dirigido al gen embrionario<sup>166</sup>. El estudio de la metilación del ADN en los promotores de los genes permitió determinar que el gen  $\pi$  es silenciado autónomamente por la presencia de la metilación del ADN<sup>166</sup>. Sin embargo, se desconoce el mecanismo por el cual la metilación del ADN en promotor del gen  $\pi$  puede llevar a su silenciamiento. Dentro de los mecanismos de silenciamiento mediados por la metilación del ADN se ha descrito la acción de proteínas con dominio de unión a ADN metilado o MBDs (del inglés Methyl Binding Domain) cuyo efecto es el de fomentar un estado en la cromatina no permisible a la transcripción<sup>39,45,47</sup>. Por tal motivo, se decidió estudiar el papel que juega la metilación del ADN en el silenciamiento del gen  $\pi$  mediado por la presencia de proteínas MBDs como lo son MBD2 y MeCP2 las cuales han sido ampliamente estudiadas en otros *loci*<sup>48,83</sup>.

Para ello se ocupó la técnica conocida como inmunoprecipitación de la cromatina o ChIP, la cual se basa en el enriquecimiento de las secuencias de ADN a las cuales se encuentra unido determinado factor o proteínas mediante el uso de anticuerpos específicos (Figura 26)<sup>71</sup>. Cabe mencionar que el ChIP es una técnica de análisis *in vivo* de las interacciones entre el ADN y las proteínas pues las células son sujetas a fijación o entrecruzamiento mediante el uso de agentes fijadores como lo es el formaldehído. De esta manera se entrecruzan químicamente las interacciones ADN-proteína permitiendo su subsecuente inmunoprecipitación y cuantificación del ADN mediante técnicas como la PCR, hibridación en micro-arreglos de ácidos nucleicos, secuenciación y otras.



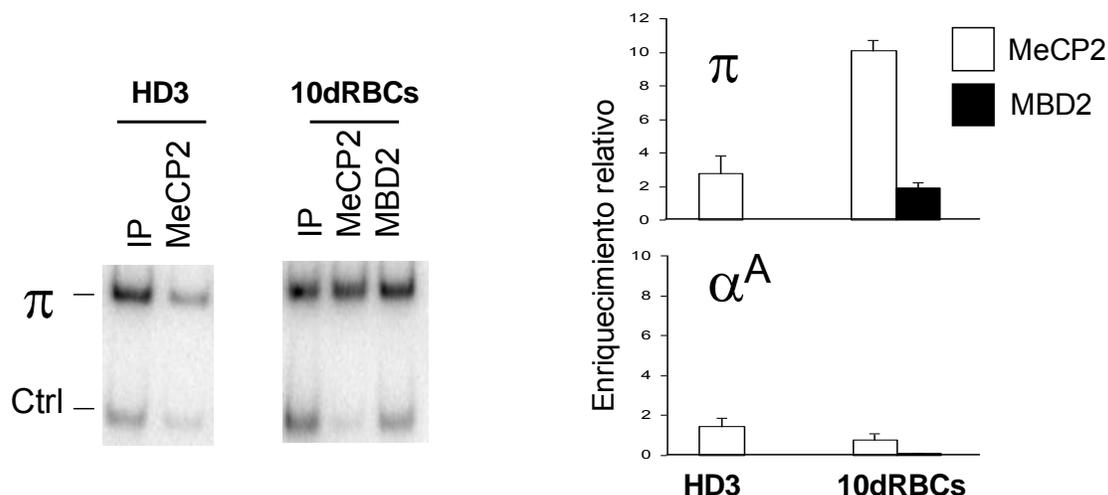
**Figura 26.** Esquema general del protocolo de inmunoprecipitación de la cromatina o ChIP. En este tipo de ensayos las células son fijadas con agentes químicos como el formaldehído (entrecruzamiento) de manera que las proteínas de la cromatina (figuras moradas, amarillas, rojas y verdes) permanecen unidas al ADN en los pasos subsecuentes del protocolo. Previo a la inmunoprecipitación con anticuerpos específicos, la cromatina es fragmentada mediante sonicación. De esta manera los anticuerpos

pueden ser precipitados al ser acoplados con perlas afines a estos. El ADN es purificado de los complejos ADN-proteína-anticuerpo-perla y es cuantificado mediante técnicas como la PCR u otras.

Para evaluar la presencia de las proteínas MeCP2 y MBD2 en los promotores de los genes  $\alpha$ -globina se utilizó el protocolo de CHIP en eritrocitos colectados de embriones de pollo a los días 10 del desarrollo y en la línea celular HD3, estadios donde el gen  $\pi$  se encuentra silenciado autónomamente por metilación del ADN (10dRBCs). Mediante el uso de oligonucleótidos específicos a cada uno de los promotores  $\alpha$ -globina se pudo evaluar el enriquecimiento de estas proteínas con la técnica de PCR semi-cuantitativa dúplex radioactiva (Figura 27). De esta manera, fue posible determinar que la proteína MeCP2 se encuentra específicamente unida al promotor del gen embrionario  $\pi$  en la etapa del desarrollo donde se encuentra silenciado (10dRBCs). Así mismo, al evaluar la presencia de este factor en la línea celular eritroblástica HD3, donde los tres promotores  $\alpha$ -globina se encuentran metilados, se observó un enriquecimiento de la proteína MeCP2 en el gen  $\pi$ . La unión de MeCP2 es específica al promotor del gen embrionario puesto que no se observó su presencia en el gen adulto  $\alpha^A$ . Este resultado pone de manifiesto la idea de un silenciamiento autónomo dirigido al promotor del gen  $\pi$  donde se involucra a la metilación del ADN y la presencia de proteínas de represión como los son las MBDs.

Como se ha mencionado anteriormente, el reclutamiento de proteínas que reconocen la metilación del ADN sugiere su participación en la conformación de una cromatina que no es permisible al reclutamiento de factores de transcripción y

con ello la expresión génica<sup>39</sup>. Dado que la proteína MeCP2 está presente en las etapas donde el gen  $\pi$  se encuentra silenciado, es posible involucrar a esta proteína en el silenciamiento autónomo de este gen.



**Figura 27.** Presencia de la proteína MeCP2 en el promotor del gen embrionario. Ensayos de ChIP en etapas avanzadas del desarrollo del pollo (10dRBCs) y en la línea celular eritroblástica HD3. Para determinar cuantitativamente la presencia de MeCP2 en este tipo de ensayos, se empleó la PCR semi-cuantitativa dúplex donde se utiliza como valor de enriquecimiento nulo o ruido de fondo la amplificación de una región de ADN donde la proteína en cuestión no presente unión (Ctrl), en este caso una región libre de metilación del ADN. La barra de error fue calculada a partir de dos ensayos independientes.

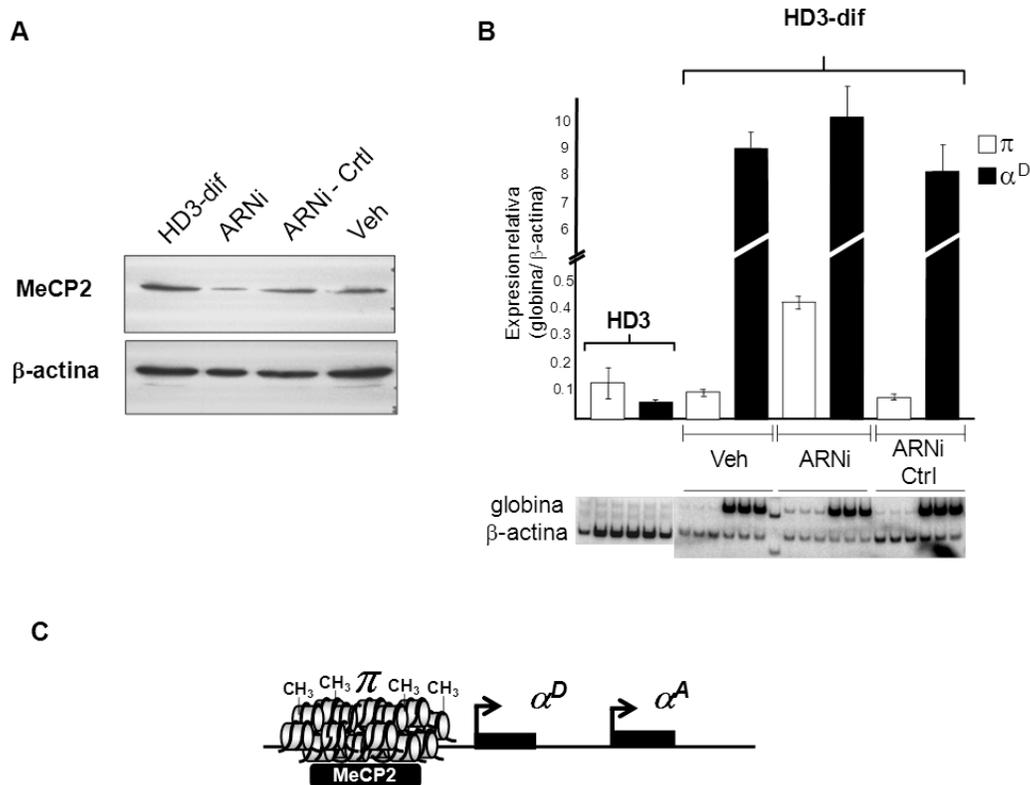
### 1.1. MeCP2 participa en el silenciamiento autónomo del gen embrionario $\pi$ .

Para conocer el papel que tiene la presencia de la proteína MeCP2 en el promotor del gen  $\pi$  en las etapas donde este se presenta metilado e inactivo a la transcripción, se decidió estudiar el efecto que tendría la reducción en los niveles de esta proteína el silenciamiento del gen embrionario  $\pi$ . La estrategia que se empleó para la reducir los niveles de la proteína consiste en la introducción de oligonucleótidos de ARN que hibridan con la secuencia del ARN mensajero del

gen *MeCP2* llevando a la interferencia en su traducción y con ello una disminución en la síntesis de esta proteína. Esta tipo de técnica es conocida como ARN de interferencia y los oligonucleótidos empleados son conocidos como ARN interferentes (ARNi).

Una serie de tres ARNi fueron diseñados para reconocer o hibridar específicamente con el ARNm de la proteína MeCP2, así mismo fueron diseñados un par de oligonucleótidos control los cuales no presentan complementariedad con ningún ARNm descrito en el pollo. Para evaluar la posible acción represora de esta proteína, los ARNs interferentes fueron introducidos en las células HD3 donde el promotor del gen embrionario  $\pi$  se encuentra metilado y con la unión de la proteína MeCP2. La ventaja que ofrecen estas células es que se encuentran arrestadas en una etapa eritroblástica previa a la etapa eritroide terminal donde se presenta la expresión de los genes adultos  $\alpha^D$  y  $\alpha^A$  mientras que el gen  $\pi$  permanece reprimido por metilación del ADN<sup>166,170</sup> (Figura 25). Estas células pueden ser inducidas a la diferenciación eritroide terminal mediante la incubación a 42°C, lo cual inactiva al virus AMV (del inglés, *Avian Myeoloblastosis Virus*) que las mantiene arrestadas en esta etapa<sup>168,170</sup>. Así mismo, es necesario el uso de fármacos que modulen las vías de señalización celular que promueven la diferenciación eritroide como lo es el ISO-H-7, el cual es un inhibidor de la Proteína Kinasa C<sup>169-171</sup>. Con este sistema de células HD3 inducibles a la diferenciación (HD3-dif), se tienen niveles elevados de expresión de los genes adultos  $\alpha^D$  y  $\alpha^A$  y el gen embrionario  $\pi$  permanece silenciado (Figura 28).

Empleando sistema de células HD3-dif, fueron transfectados los ARNi mediante la técnica de lipofección celular. Para verificar que los ARNs introducidos tienen la capacidad de disminuir los niveles en la abundancia de la proteína MeCP2, se realizaron experimentos de *western blot* que permiten evaluar cuantitativamente la abundancia de las proteínas. Como resultado, se observó una disminución de la proteína MeCP2 en las células que fueron transfectadas con los ARNi diseñados para hibridar con el transcrito de MeCP2 (Figura 28A). Las células tratadas con los ARNi control no presentaron disminución de la proteína lo que corrobora la especificidad de en el experimento.



**Figura 28.** Participación de MeCP2 en el silenciamiento del gen embrionario. **A)** Las células HD3 fueron inducidas a la diferenciación eritroide (HD3-dif) y transfectadas con ARNi para reducir los niveles de abundancia de la proteína MeCP2. El ensayo de western blot muestra que exclusivamente los ARNi diseñados contra el ARNm de dicha proteína

tiene la capacidad de reducir sus niveles. Por su parte, ni la lipofección sin ARNi (Veh) ni los controles de no interferencia (ARNi-Ctrl) tienen efecto en los niveles de la proteína MeCP2. **B)** Niveles de expresión de los genes  $\pi$  y  $\alpha^D$  en las células HD3 y HD3-dif. En la parte inferior se muestra un gel de poli-acrilamida representativo las RT-PCRs radioactivas dúplex. Las barras de error provienen de dos experimentos independientes. **C)** Dominio de represión epigenética sobre el gen embrionario  $\pi$  el cual está constituido por la presencia de metilación del ADN y la acción de la proteína MeCP2.

Una vez verificada la reducción de la proteína MeCP2 en las células HD3-dif, fueron evaluados los niveles de expresión de los genes  $\alpha$ -globina mediante RT-PCR semi-cuantitativa dúplex radioactiva utilizando como referencia la expresión constitutiva del gen  $\beta$ -actina de pollo (Figura 28B). En las células donde se indujo la disminución de la proteína MeCP2 se pudo observar, de manera reproducible, una activación transcripcional discreta del gen embrionario  $\pi$ . Dicha activación transcripcional es dependiente de la disminución de la proteína ya que en las células tratadas con el ARNi control dicho gen permanece silenciado. De esta manera se puede determinar que la presencia de la proteína MeCP2 es necesaria para mantener silenciado al gen embrionario  $\pi$  en las etapas adultas de la diferenciación eritroide donde los genes  $\alpha^D$  y  $\alpha^A$  presentan actividad transcripcional.

En conclusión, se puede vincular a la metilación del ADN que se presenta en el promotor del gen  $\pi$  y al reclutamiento de la proteína de unión a ADN metilado MeCP2 en el silenciamiento del gen embrionario en las etapas eritroides adultas o definitivas. Por su parte, la presencia de MeCP2 es necesaria para fomentar un ambiente de represión transcripcional puesto que dicha proteína, además de su dominio de unión a ADN metilado, posee un dominio de represión

(TRD, del inglés: *Transcriptional Repression Domain*)<sup>176</sup>. Dicho TRD se sabe tiene la capacidad de interactuar con proteínas y complejos asociados a la represión transcripcional como los son las HDACs<sup>176</sup>.

Con esta serie de resultados se ha podido determinar que, como parte del silenciamiento autónomo del gen embrionario  $\pi$  dirigido por la metilación del ADN, es necesario el reclutamiento de proteínas que pudieran fomentar un estado represor de la cromatina en la región promotora de dicho gen. Este tipo de cromatina es exclusiva de las etapas adultas de la diferenciación eritroide permitiendo la expresión diferencial de los genes  $\alpha$ -globina de pollo (Figura 28C). Por ello, a la presencia de dicha metilación en el ADN y la acción de la proteína de represión MeCP2 la hemos denominado como un *dominio epigenético de represión*. Dicho dominio se encuentra presente en la región genómica del promotor del gen embrionario y se desconoce su distribución en las regiones genómicas aledañas a este (Figura 28C).

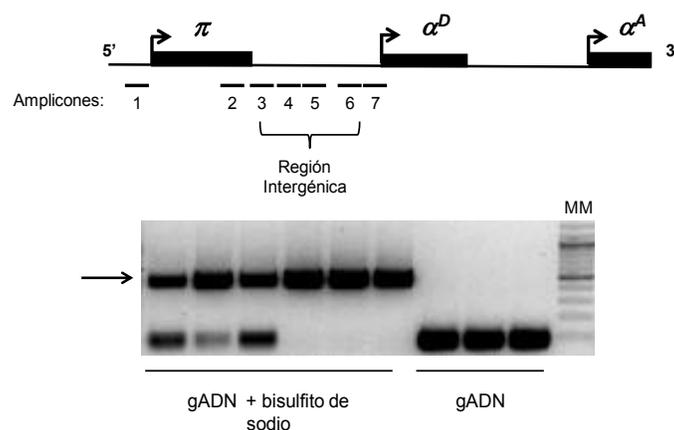
De manera interesante, se sabe que la metilación del ADN es capaz de expandirse a las regiones aledañas y con ello los complejos y proteínas asociadas con esta<sup>42,49,118</sup>. Por ello, resulta necesario comprender la manera en que el *dominio de represión epigenética* del gen embrionario  $\pi$  es delimitado a esta región genómica y no interfiere con la expresión de los genes aledaños  $\alpha^D$  y  $\alpha^A$ .

## 2. Distribución de la metilación del ADN en la región genómica del gen embrionario $\pi$ .

Con la intención de caracterizar la distribución y delimitación del *dominio epigenético de represión*, se decidió analizar a detalle la metilación del ADN presente en la región promotora del gen  $\pi$ , el cuerpo del gen, la región intergénica 3' hacia el gen  $\alpha^D$  así como el promotor del gen adulto  $\alpha^D$  (Figura 29). De esta manera, si se conoce el estado en la metilación en estas regiones y en las diferentes etapas del desarrollo, se puede determinar la distribución de dicho dominio de represión. Con la técnica conocida como transformación de ADN por bisulfito de sodio es posible determinar el estado de metilación (metilada o no metilada) de cada CpG contenida en el genoma en cuestión. El tratamiento del ADN genómico con el agente químico bisulfito de sodio induce una oxidación de todas las citosinas presentes cuando no se encuentran metiladas convirtiéndolas en una base de estructura similar al uracilo (CpN > UpN). Dicha conversión es independiente al tipo de di-nucleótido (CpG o CpN). Por el contrario, la presencia de metilación en el di-nucleótido CpG ( $^m\text{CpG}$ ) interfiere con el proceso de oxidación de la citosina y esta permanece sin conversión ( $^m\text{CpG} > ^m\text{CpG}$ )<sup>177</sup>. De esta manera, mediante la amplificación por la técnica de PCR de este ADN así tratado utilizando oligonucleótidos específicos a la(s) región(es) de interés y posterior secuenciación, se puede determinar el estado en la metilación. Esto es, después de la secuenciación las citosinas que no se encuentran metiladas se evidenciarán con adenina o timina (dependiendo de la hebra de ADN

secuenciada) y por su parte las citosinas que se encuentran metiladas permanecerán como citosina o guanina.

Para conocer la distribución de la metilación del ADN en las diferentes etapas del desarrollo del pollo, fueron colectados eritrocitos en circulación de embriones de pollo al día 5 y 1° del desarrollo (5dRBCs y 10dRBCs, respectivamente). El ADN genómico (ADNg) fue obtenido de estas células mediante un protocolo estándar de extracción. Se utilizaron aproximadamente 3 microgramos ( $\mu\text{g}$ ) de ADN genómico para ser tratado con el bisulfito de sodio. Fueron diseñados 7 diferentes pares de oligonucleótidos para amplificar mediante PCR las regiones genómicas anteriormente mencionadas (Figura 29). Los productos de PCR fueron purificados mediante extracción en geles de agarosa y clonados en un vector plasmídico que contiene secuencias de hibridación para oligonucleótidos de secuenciación Sanger. Fueron secuenciadas y analizadas al menos 10 clones correspondientes a cada región de estudio.



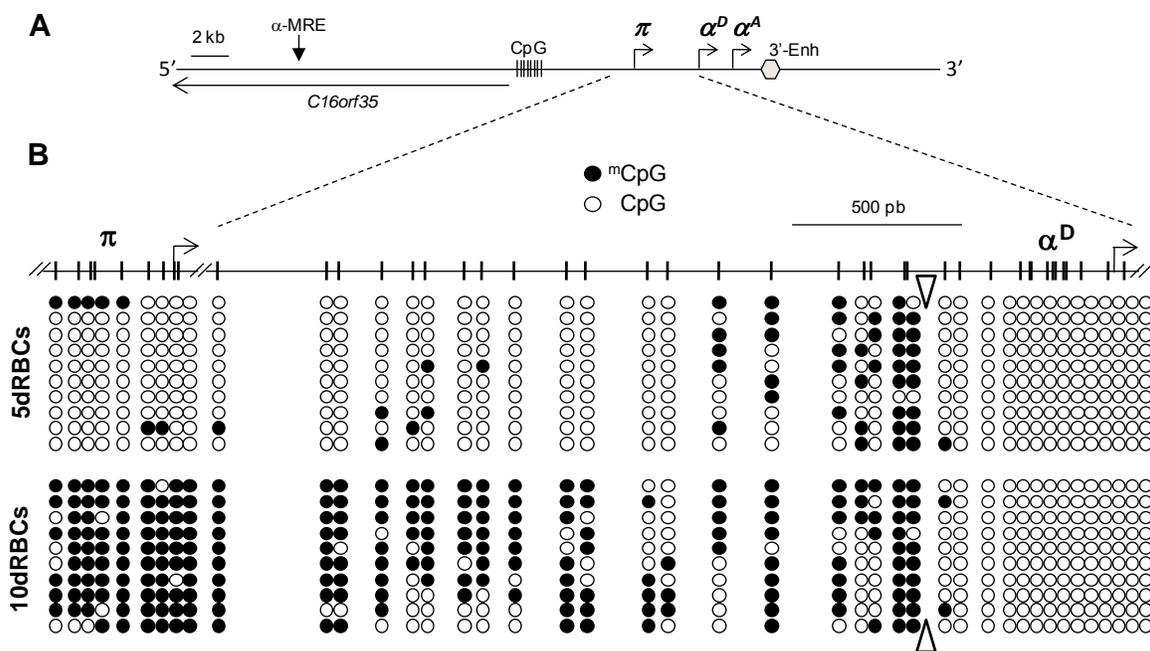
**Figura 29.** Regiones genómicas para el análisis de metilación del ADN por bisulfito de sodio. Se diseñaron siete pares de oligonucleótidos específicos para amplificar la región promotora del gen  $\pi$  (1), el cuerpo del gen (2), la región intergénica 3' (3-6) y el promotor

del gen  $\alpha^D$ . Como se muestra en el gel en la parte inferior, estos oligonucleótidos son específicos para la amplificación de ADN<sub>g</sub> modificado con bisulfito de sodio dado que no es posible su hibridación en el ADN no tratado (ADN<sub>g</sub>). La flecha señala el producto de amplificación específico de la región 2. MM: marcador de peso molecular.

Como resultado, se observó que los eritrocitos al día 5 del desarrollo del pollo (5dRBCs) no presentan metilación en las regiones promotoras de los genes  $\pi$  y  $\alpha^D$  (Figura 30B). En lo que se refiere a la región intergénica, esta se encontró al igual desprovista de metilación en el ADN, con la excepción de una región muy notoria a 700 pb río arriba del sitio de inicio de la transcripción (TSS; del inglés, Transcriptional Start Site) del gen  $\alpha^D$ . De manera interesante, en los eritrocitos de etapas avanzadas del desarrollo (10dRBCs), se observó un estado de hipermetilación del ADN en la región promotora del gen  $\pi$ , el cuerpo de este gen y la región intergénica. Dicha hipermetilación se extiende desde el promotor del gen  $\pi$  hasta una zona a 700 pb río arriba del TSS del gen  $\alpha^D$ ; coincidente con la zona de metilación presente en los 5dRBCs (triángulos; Figura 30B).

Esta serie de resultados permiten concluir que: 1) en los 5dRBCs, donde exclusivamente el gen  $\pi$  es activo a la transcripción, no se presenta metilación del ADN en ninguno de los promotores analizados, por lo que el silenciamiento del gen  $\alpha^D$  en estas etapas del desarrollo no se encuentra mediado por metilación en el ADN. Sin embargo, existe metilación de CpGs en una zona de la región intergénica a 700 bp río arriba del TSS. 2) en los 10dRBCs, etapa donde el gen  $\pi$  se encuentra silenciado y los genes adultos activos a la transcripción, la metilación del ADN se presenta más allá del promotor  $\pi$  hacia la región

intergénica con el gen  $\alpha^D$  con una marcada transición en la zona a 700 bp río arriba del TSS. De esta manera, fue posible determinar que el dominio de silenciamiento epigenético del gen  $\pi$  se extiende hasta la región intergénica cercana al promotor del gen  $\alpha^D$ , donde ocurre una transición en el estado de metilación. Estos resultados están de acuerdo con las propuestas anteriormente planteadas de un silenciamiento epigenético autónomo dirigido exclusivamente al gen embrionario como parte del mecanismo para la expresión diferencial de los genes globina<sup>20,155,166</sup>.



**Figura 30.** Distribución de la metilación del ADN en la región genómica del gen  $\pi$  y  $\alpha^D$ . **A)** Esquema del dominio  $\alpha$ -globina de pollo donde se indica la región la región de estudio (líneas punteadas). **B)** En el esquema ampliado (indicado por las líneas punteadas), se muestran cada una de las CpGs (líneas cruzadas) contenidas entre el promotor del gen  $\pi$  y el promotor del gen  $\alpha^D$ . El estado de metilación de cada CpG se representa con los círculos. Los círculos cerrados (negros) corresponden con CpGs metiladas (<sup>m</sup>CpG) y los círculos abiertos con CpGs no metiladas. La zona de transición de un estado metilado a uno no metilado está indicada por los triángulos.

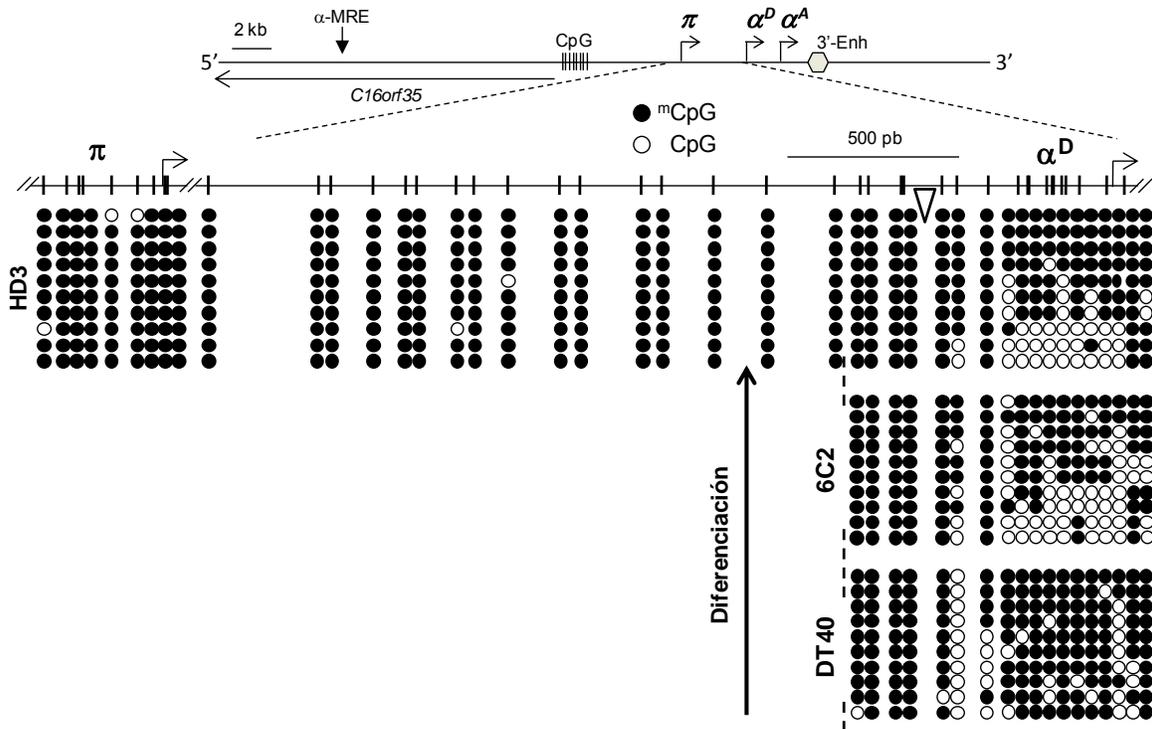
Un aspecto muy relevante de estos resultados es el descubrimiento de una zona de transición de la metilación del ADN (a 700 pb del TSS del gen  $\alpha^D$ ), la cual podría indicar la formación de sub-dominios de regulación en el dominio  $\alpha$ -globina. Es decir, para la expresión diferencial de este grupo de genes a través del desarrollo se ha encontrado el establecimiento de un silenciamiento epigenético (sub-dominio de metilación) que abarca al gen embrionario y el cual se extiende río abajo en la región intergénica con el gen  $\alpha^D$ . La presencia de dicho sub-dominio sugiere la acción de mecanismos moleculares que demarcan o delimitan la extensión de dicho silenciamiento epigenético. Sin embargo, se desconoce el la manera por la cual dicho sub-dominio de metilación puede ser regulado.

## **2.2. Distribución de la metilación del ADN en la diferenciación eritroide.**

Un aspecto relevante en el estudio de la regulación epigenética es el conocer los procesos y mecanismos que permiten la ejecución de los programas de expresión de los genes a través de la diferenciación de las células. Los eritrocitos adultos o 10dRBCs, donde se presenta el sub-dominio de metilación que conlleva al silenciamiento del gen embrionario, son el resultado de un proceso de diferenciación denominada adulta o definitiva. Por tal motivo, el conocer la manera en que se establece el sub-dominio de metilación en etapas previas a la diferenciación terminal (expresión de los genes adultos), podría ayudar a

comprender los mecanismos implicados en la ejecución de dichos programas de expresión diferencial de los genes.

Para conocer la manera en que la metilación se presenta en etapas previas a la expresión de los genes adultos, se decidió estudiar el estado en la metilación del ADN en las líneas celulares eritroblásticas adultas. En estas etapas ninguno de los genes  $\alpha$ -globina es activo a la transcripción<sup>166</sup>. Para ello, fue obtenido el ADNg de las líneas celulares DT40 (precursor linfoide-mieloide), 6C2 (pro-eritroblástica) y HD3 (eritroblástica). El ADNg fue sujeto al protocolo de bisulfito de sodio, amplificación por PCR y secuenciación, de la misma manera que para los análisis anteriores. Como se muestra en la Figura 31, se observó un estado de hipermetilación del ADN en la zona de transición incluyendo al promotor del gen  $\alpha^D$  en las células DT40 y 6C2. En las células eritroblásticas HD3, donde fueron analizadas todas las regiones de estudio, se pudo evidenciar una hipermetilación del ADN que se extiende desde el promotor del gen  $\pi$  hasta el TSS del gen  $\alpha^D$ . De esta manera, se pudo establecer que durante la diferenciación eritroide definitiva, se presenta un estado de hipermetilación que abarca a los promotores y la región intergénica y es coincidente con la inactividad transcripcional de los genes.



**Figura 31.** Distribución de la metilación del ADN a través de la diferenciación eritroide definitiva. Esquema del dominio  $\alpha$ -globina de pollo donde se indica la región la región de estudio (líneas punteadas). En el esquema ampliado (indicado por las líneas punteadas), se muestran cada una de las CpGs (líneas cruzadas) contenidas entre el promotor del gen  $\pi$  y el promotor del gen  $\alpha^D$ . El estado de metilación de cada CpG se representa con los círculos. Los círculos cerrados (negros) corresponden con CpGs metiladas ( $^m$ CpG) y los círculos abiertos con CpGs no metiladas. La zona de transición de un estado metilado a uno no metilado está indicada con el triángulo.

En resumen, se ha podido caracterizar que en los estados de diferenciación celular previos a la expresión de los genes  $\alpha$ -globina, se presenta un estado de hipermetilación del ADN el cual se extiende desde el promotor del gen  $\pi$  hasta el TSS del gen  $\alpha^D$  correlacionando con la inactividad transcripcional de los genes. En etapas terminales de la diferenciación, donde hay expresión diferencial de los genes (RBCs), existe un sub-dominio de metilación que se

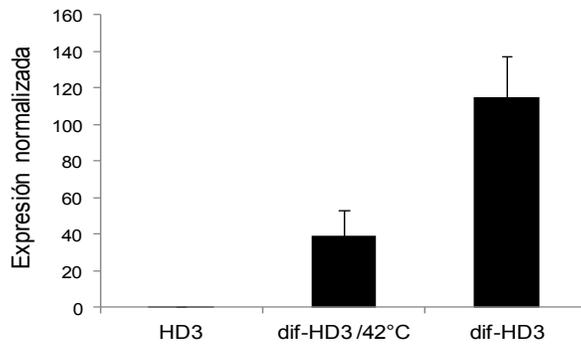
extiende hasta la región intergénica el cual participa en el silenciamiento epigenético del gen embrionario (10dRBCs). Con estos resultados, se han podido formular las siguientes propuestas: en las etapas previas a la expresión de los genes adultos (células DT40, 6C2 y HD3) la hipermetilación del ADN participa en el silenciamiento de los tres genes  $\alpha$ -globina. Para el caso de la etapa adulta terminalmente diferenciada (10dRBCs), donde los genes adultos son transcripcionalmente activos, la metilación se distribuye a lo largo del gen  $\pi$  llevando a su silenciamiento y es acotada o delimitada hasta 700 pb río arriba del TSS del gen  $\alpha^D$ , constituyendo así un sub-dominio de metilación el cual es necesario para la expresión regulada de los genes adultos. Esta serie de propuestas implican la acción de un mecanismo molecular dirigido al establecimiento del sub-dominio de metilación sobre el gen  $\pi$  y la región intergénica río abajo.

### **2.3. Desmetilación del ADN en el dominio $\alpha$ -globina**

Para corroborar el papel que tiene la formación de un sub-dominio de metilación en la expresión regulada de los genes adultos, se decidió estudiar los procesos que ocurren para el establecimiento de dicho sub-dominio. Para ello, fue necesario emplear un sistema celular donde se pudiera inducir experimentalmente la expresión de los genes adultos y estudiar el estado de la metilación relacionado con la inducción de la transcripción. Dicho sistema celular corresponde con las células HD3, donde se ha caracterizado la hipermetilación de la región de estudio. Las células HD3 pueden ser inducidas a la diferenciación

eritroide terminal mediante la inactivación del virus aviar (AMV) con el cual fueron arrestadas en una etapa eritroblástica<sup>170</sup>. La inactivación del AMV se logra con la incubación de las células a 42°C. lo cual conlleva a la expresión de los genes globina  $\alpha^D$  y  $\alpha^{A166}$ . De esta manera, utilizando el protocolo de bisulfito de sodio se puede evaluar el estado en la metilación del ADN en las células donde se ha inducido la expresión de los genes adultos.

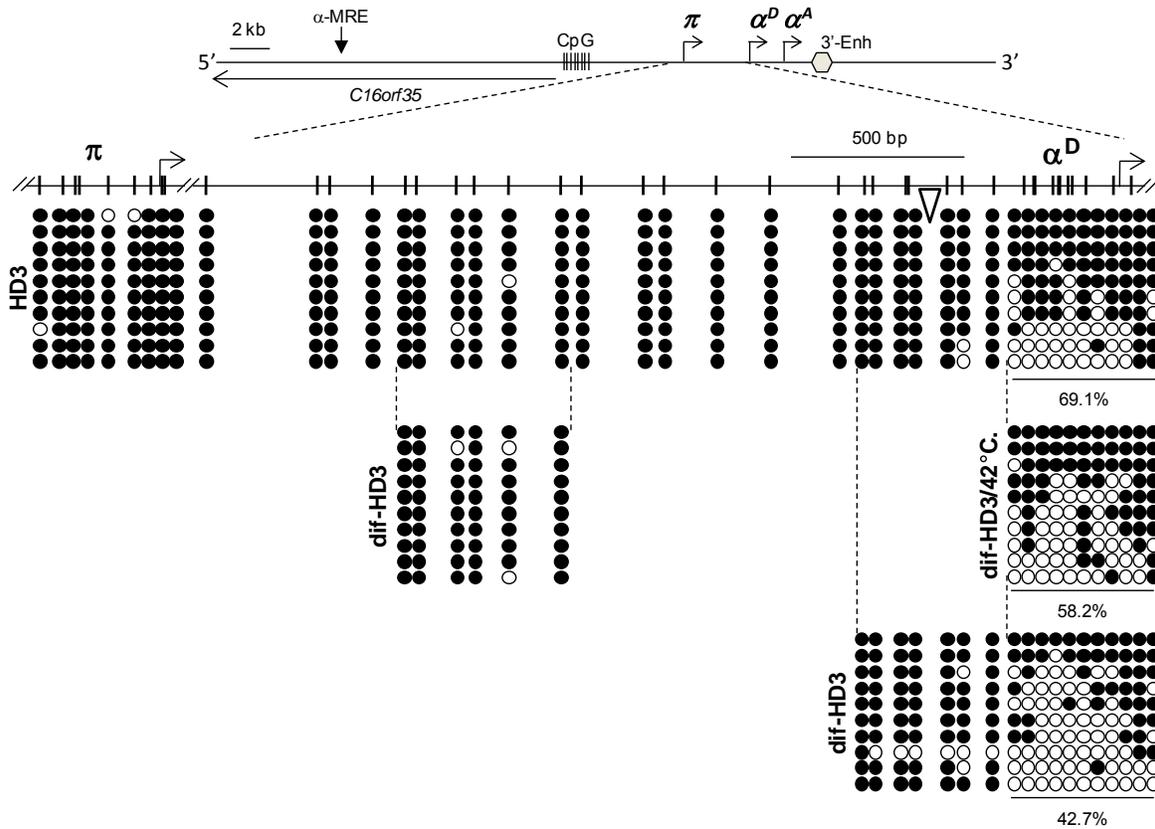
Las células HD3 fueron incubadas a 42°C durante 48 hrs (dif-HD3/42°C) y se extrajo el ADN genómico así como ARN para corroborar el estado de expresión de los genes. Como se muestra en la figura 32, en las células dif-HD3/42°C el gen  $\alpha^D$  es activo a la transcripción como consecuencia de la inducción a la diferenciación por temperatura. Al analizar el estado de metilación se pudo evidenciar una disminución en los niveles de metilación en la región del promotor  $\alpha^D$  en comparación con las células HD3 no inducidas donde el promotor se encuentra hipermetilado (Figura 33). Esta disminución en la metilación pudo ser cuantificada en términos de porcentaje de metilación presente siendo que en las células HD3 se presenta un 69.1% de metilación en el promotor  $\alpha^D$  y en las células dif-HD3/42°C corresponde con un 58.2%. De esta manera fue posible vincular a la desmetilación del ADN con la regulación en la expresión del gen adulto  $\alpha^D$ .



**Figura 32.** Niveles de expresión del gen  $\alpha^D$  en la inducción a la diferenciación. La expresión del gen  $\alpha^D$  fue determinada mediante RT-PCR en tiempo real. La expresión del gen fue normalizada usando como referencia la abundancia de transcrito del gen ribosomal 18S. Las barras de error provienen de tres experimentos independientes.

Para corroborar la relación entre la disminución de la metilación del ADN y la activación del gen  $\alpha^D$ , se decidió estudiar los niveles de metilación en un contexto de mayor inducción a la diferenciación de las células HD3 (dif-HD3). Esto es posible ya que el tratamiento de estas células con un inhibidor de la proteína Kinasa-C, conocido como ISO-H7, se puede inducir un mayor grado de diferenciación y con ello niveles más elevados de expresión del gen  $\alpha^{D170,171}$ . El mayor grado de inducción fue evaluado mediante el análisis de transcripción del gen  $\alpha^D$  (Figura 32). En las células dif-HD3 se detectó una disminución aun mayor en los niveles de metilación en el promotor del gen  $\alpha^D$  (42.7%) en comparación con las células dif-HD3/42°C (58.2%) y HD3 (69.1%). Dicha reducción es exclusiva a la región del promotor  $\alpha^D$  ya que el análisis de una zona río arriba en la región intergénica no presentó cambios en los niveles de metilación (Figura 33). Considerando el incremento en la actividad transcripcional del gen  $\alpha^D$  y la pérdida en metilación del ADN fue posible concluir que la desmetilación que se presenta

en esta región es necesaria para la regulación de los genes  $\alpha$ -globina a través de la diferenciación eritroide.



**Figura 33.** Distribución de la metilación del ADN en células inducidas a la diferenciación eritroide. Distribución de la metilación del ADN a través de la diferenciación eritroide definitiva. Esquema del dominio  $\alpha$ -globina de pollo donde se indica la región la región de estudio (líneas punteadas). En el esquema ampliado (indicado por las líneas punteadas), se muestran cada una de las CpGs (líneas cruzadas) contenidas entre el promotor del gen  $\pi$  y el promotor del gen  $\alpha^D$ . El estado de metilación de cada CpG se representa con los círculos. Los círculos cerrados (negros) corresponden con CpGs metiladas ( $^m$ CpG) y los círculos abiertos con CpGs no metiladas. La zona de transición de un estado metilado a uno no metilado está indicada con el triángulo.



ha no hay presencia de dichos factores en el promotor del gen  $\alpha^D$ , por lo que la metilación en estas células podría interferir con su reclutamiento al promotor. Sin embargo, estos factores se encuentran presentes en etapas donde dicho promotor se encuentra desmetilado (dif-HD3 y 10dRBCs) y activo transcripcionalmente<sup>166</sup>. Por tal motivo es posible vincular a la metilación-desmetilación del ADN como un mecanismo.

#### **2.4. Desmetilación activa del ADN en la diferenciación eritroide**

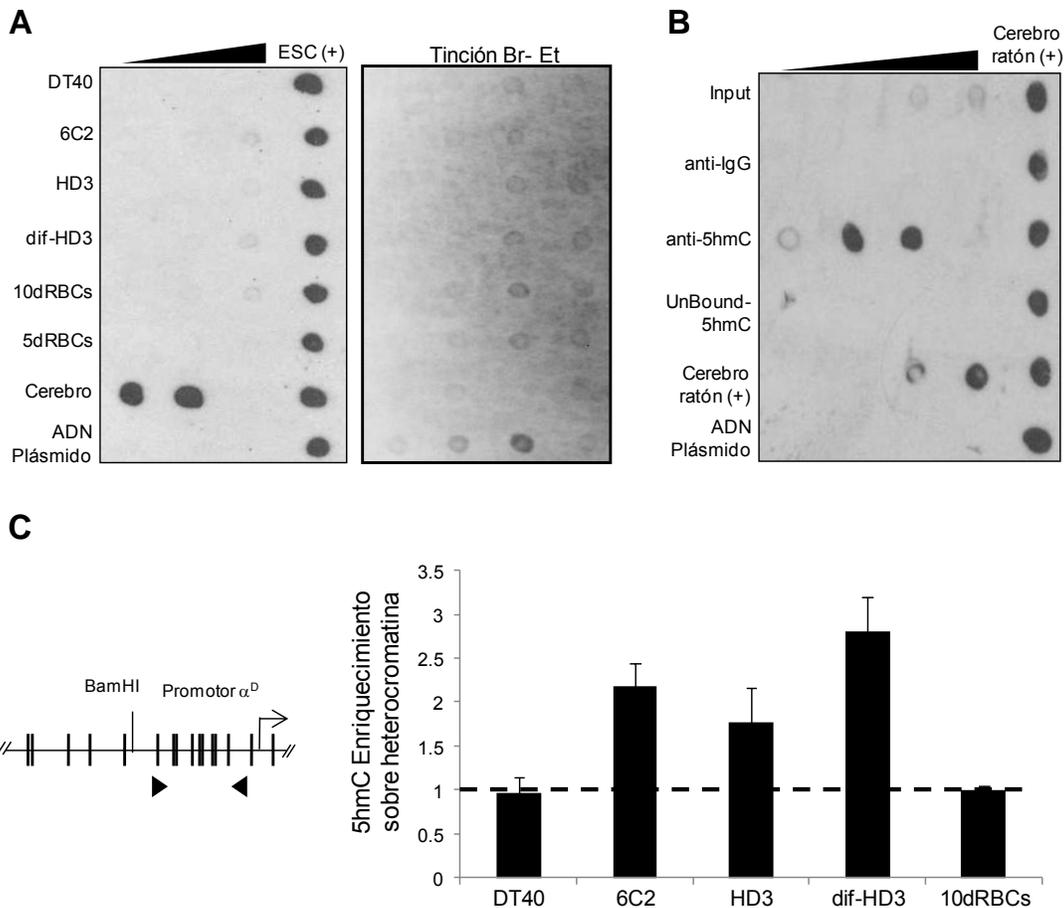
Una vez correlacionada la presencia de la metilación del ADN y el su delimitación para la activación de los genes adultos, se decidió abordar los posibles mecanismos que participan en la distribución de la metilación del ADN. Un aspecto muy importante es que esta desmetilación que se presenta con la inducción a la diferenciación y activación del gen  $\alpha^D$ , ocurre específicamente en la región donde se presenta la transición en el sub-dominio de metilación del gen  $\pi$ . Por tal motivo, es posible considerar la existencia de uno o varios mecanismos involucrados en el establecimiento y delimitación de dicho subdominio. Dentro de los mecanismos conocidos que participan en la modulación del estado de metilación en los promotores de los genes se encuentra la desmetilación del ADN mediada por la presencia de 5hmC<sup>62,63,65</sup>. La adición de un grupo hidroxilo en CpG metiladas (<sup>5hm</sup>CpG) puede llevar a la incorporación de una citosina no metilada por mecanismos relacionados con la reparación del ADN de manera dependiente o no de la división celular<sup>69</sup>.

Por tal motivo, decidimos evaluar la presencia de 5hmC asociada con la desmetilación en el promotor del gen  $\alpha^D$ . Como un primer acercamiento, fueron evaluados los niveles globales de 5hmC en células correspondientes a las diferentes etapas del desarrollo del pollo (RBCs) y a través de la diferenciación eritroide (células DT40, 6C2, HD3 y dif-HD3). Para ello, el ADNg proveniente de dichas células fue cargado en membranas de hibridación para la inmunodetección de la 5hmC utilizando un anticuerpo que es específico para dicha modificación, una técnica conocida como dot-blot de ADN. Como se muestra en la Figura 35, fue posible detectar la presencia 5hmC en las células del linaje eritroide. Sin embargo, los niveles que se presentan son muy reducidos en comparación con los controles positivos que corresponde con ADNg de células madre y tejido nervioso de ratón<sup>67</sup>. Esta ligera presencia de 5hmC en las células eritroides de pollo sugiere que su papel puede ser más el de promover una desmetilación del ADN localizada a ciertas regiones del genoma.

Como segundo abordaje, se decidió analizar la presencia de 5hmC en la región del promotor  $\alpha^D$  en las etapas donde se presenta la desmetilación del ADN. Para ello, se utilizó una técnica conocida como 5hMeDIP la cual consta de la inmunoprecipitación del ADN 5-hidroxi-metilado (validaciones en la Figura 35B). El ADNg de las células eritroides de pollo fue reducido y homogeneizado en fragmentos de menor tamaño (rango de 200 a 800 pb) para facilitar su inmunoprecipitación. Posteriormente, este ADNg es incubado con un anticuerpo que reconoce específicamente la 5hmC. De esta manera, los fragmentos de ADN que poseen la marca de 5hmC pueden ser enriquecidos mediante la precipitación

del complejo ADN-anticuerpo utilizando perlas de afinidad (validación en la Figura 35B). Los niveles de enriquecimiento del promotor  $\alpha^D$  fueron evaluados mediante PCR en tiempo real utilizando oligonucleótidos específicos a dicho promotor. Como resultado, se observó que existe un enriquecimiento de la 5hmC en la región promotora del gen  $\alpha^D$  (Figura 35C) en las etapas donde la hay presencia de metilación del ADN (6C2 y HD3). Dicho enriquecimiento es mayor en las células dif-HD3 coincidente con la reducción en los niveles de metilación y activación del gen  $\alpha^D$  (Figura 32 y 33; respectivamente). De manera importante en las células DT40, que corresponden con un estado linfoide-mieloide en la hematopoyesis del pollo, no se detectó la presencia de la 5hmC en el promotor  $\alpha^D$  no obstante de su estado hipermetilado; un aspecto que valida la especificidad del experimento. Con este resultado se ha podido concluir que, la incorporación de la 5hmC en región promotora del gen  $\alpha^D$  puede contribuir al establecimiento del subdominio de metilación. Además, este resultado es el primer antecedente de la participación de la 5hmC en la regulación de la expresión diferencial de los genes globina.

En conclusión, se ha determinado que en las etapas adultas del desarrollo del pollo y en la diferenciación eritroide definitiva, existe un sub-dominio de metilación que se extiende a lo largo del promotor y cuerpo del gen  $\pi$  participando en su silenciamiento. Dicho sub-dominio de metilación es delimitado y acotado mediante una desmetilación activa del ADN que involucra a la 5hmC en la región promotora del gen  $\alpha^D$ . Esta desmetilación del ADN es necesaria para la expresión regulada de los genes adultos.



**Figura 35.** Análisis de la marca 5hmC en las células eritroides del pollo. **A)** Determinación de la abundancia global de la 5hmC en los diferentes contextos celulares. Se muestran las membranas del ensayo de *dot-blot*. La membrana del lado izquierdo corresponde a la incubación con el anticuerpo específico anti-5hmC. Como controles positivos del experimento se incluyeron muestra de ADNg proveniente de cerebro de ratón y pollo; células troncales embrionarias de ratón mESCs. Del lado derecho se muestra la membrana teñida con bromuro de etidio para evidenciar el cargado del ADN. **B)** Validación de la técnica de 5hMeDIP mediante ensayos de *dot-blot*. Se cargaron cantidades crecientes de ADN de cerebro de pollo proveniente de la incubación con el anti-5hmC y con el anti-IgG (control negativo). Tanto en A y en B, las muestras de ADN plásmido provienen de bacteria donde no existe la marca de 5hmC, por ello se considera como un control negativo. **C)** 5hMeDIP y PCR en tiempo real de la región promotora del gen  $\alpha^D$ . En el esquema se muestran los oligonucleótidos utilizados (puntas de flechas). El ADNg previo a su fragmentación por sonicación fue digerido con la endonucleasa de restricción BamHI, de esta manera se analizó el enriquecimiento específico al promotor y no así de las regiones aledañas. Las líneas cruzadas representan a las CpGs. Se grafican los niveles de enriquecimiento en el promotor  $\alpha^D$  por encima de una región desprovista de 5hmC que corresponde con heterocromatina constitutiva<sup>61</sup> (línea

horizontal discontinua). Las barras de error fueron generadas a partir de dos experimentos independientes.

### **3. Estructura de la cromatina en la regulación de la expresión diferencial de los genes $\alpha$ -globina.**

Una vez determinada la distribución de la metilación del ADN, el estudio fue enfocado a comprender la participación de la estructura de la cromatina en la expresión diferencial de los genes  $\alpha$ -globina de pollo. Como se ha revisado, el factor multifuncional CTCF se ha descrito como un regulador maestro de la estructura de la cromatina y de la organización espacial al interior del núcleo<sup>108,121</sup>. Por tal motivo, el análisis de la distribución del factor CTCF podría ayudar a comprender la manera en que la estructura de la cromatina puede influenciar la expresión de los genes  $\alpha$ -globina. Como un primer acercamiento, se decidió estudiar la distribución del factor CTCF en el dominio  $\alpha$ -globina mediante la técnica de inmunoprecipitación de la cromatina (ChIP) acoplada a secuenciación masiva de alta resolución o ChIP-seq. Este tipo de experimentos permiten conocer la distribución de los factores en cuestión en todo el genoma con una resolución de hasta 20 pb<sup>71,106</sup>. La aplicación de este tipo de técnicas fue posible gracias a una colaboración con el Dr. J. Luis Gómez-Skarmeta en Sevilla, España, y colaboradores, enfocada a comprender el papel del factor CTCF en la regulación de los genes a través del desarrollo embrionario de los vertebrados. Una vez establecida dicha colaboración se decidió analizar la distribución de

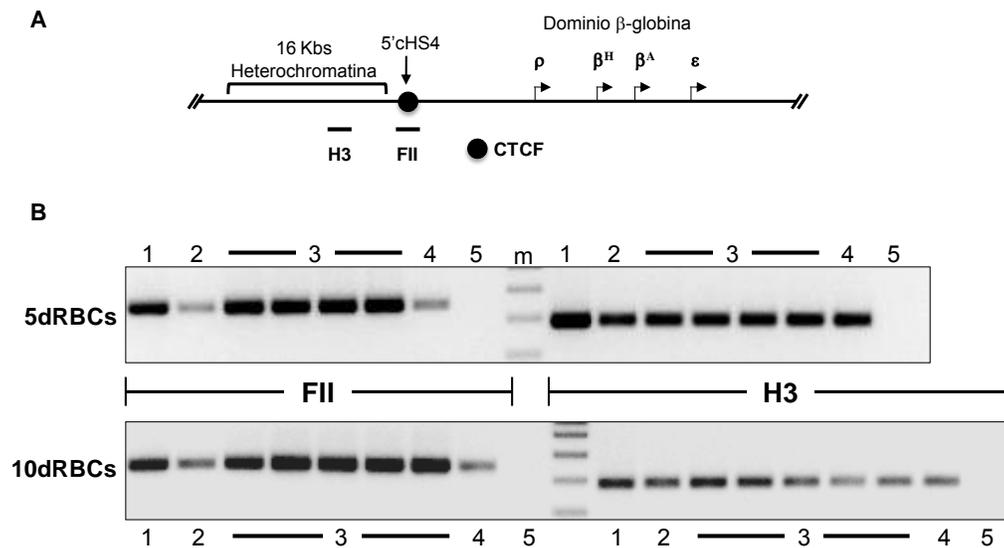
CTCF en los eritrocitos en diferentes etapas del desarrollo del pollo (5dRBCs y 10dRBCs, respectivamente).

### **3.1. Distribución del factor CTCF en el dominio $\alpha$ -globina a través del desarrollo**

Para los experimentos de ChIP-seq se empleo un anticuerpo generado en el laboratorio el cual es altamente específico para reconocer la proteína CTCF de pollo<sup>161</sup>. Se utilizaron 5dRBCs y 10dRBCs, etapas donde existe expresión diferencial de los genes  $\alpha$ -globina. Inicialmente, los experimentos de ChIP-seq fueron validados analizando el enriquecimiento del factor en una región *insulator* del dominio  $\beta$ -globina de pollo conocida como 5'cHS4 cuya función *insulator* depende de la presencia de CTCF<sup>27,113</sup>. El sitio de unión de CTCF dentro del insulator 5'cHS4 es una secuencia conocida como FII (del inglés, Footprint II). Utilizando oligonucleótidos específicos para la secuencia FII se observó el enriquecimiento de CTCF en la región FII (Figura 36B). En contraste, cuando se analizó una región de heterocromatina constitutiva (H3) que se sabe no es reconocida por el factor CTCF, no se encontró enriquecimiento de CTCF (Figura 36B). Con estos resultados se pudo verificar la especificidad del anticuerpo y la calidad adecuada del ChIP para ser analizado mediante secuenciación masiva.

Una vez verificada la especificidad y calidad de los experimentos de ChIP, estos fueron enviados para su secuenciación con cobertura de todo el genoma de pollo. Para la secuenciación fue necesario la incorporación de adaptadores a los

fragmentos de ADN presentes en las muestras de ChIP empleando un kit de la compañía Illumina. Estos adaptadores son empleados para la amplificación del material y su secuenciación. La secuenciación de este material se realizó en un equipo conocido como *Illumina Genome Analyzer platform II*. Los datos de secuenciación fueron analizados con un programa que permite obtener las coordenadas genómicas de las secuencias obtenidas y la abundancia de cada una de ellas. Para conocer la distribución de CTCF en el genoma del pollo, estos datos con coordenadas fueron cargados en una plataforma de visualización del genoma llamada *Genome Browser (UCSC, de la Universidad de Santa Cruz, EU)*. Para conocer los sitios de enriquecimiento de CTCF que se encuentran por encima del ruido de fondo del experimento, se empleó un algoritmo conocido como *peak calling* el cual permite identificar y asignar “picos” de enriquecimiento.



**Figura 36.** Validación de los experimentos de ChIP-seq del factor CTCF. **A)** En el dominio  $\beta$ -globina de pollo se ha caracterizado al elemento 5'cHS4 como un *insulator* cuya actividad depende de la presencia de CTCF. FII y H3 corresponde a las regiones de análisis en los ensayos de PCR. **B)** PCRs de los ensayos de ChIP utilizando como

templado el ADN proveniente de: (1) ADN input; (2) anti-IgG; (3) anti-CTCF (cuatro y cinco experimentos independientes en 5dRBCs y 10dRBCs, respectivamente); (4) Beads (perlas de inmuno-afinidad) y (5) H<sub>2</sub>O o control de PCR sin ADN. Cuando se utilizan los oligonucleótidos que reconocen a la región FII, existe una mayor amplificación del material proveniente del CHIP con el anticuerpo de CTCF (3), la cual sobrepasa la amplificación de los controles negativos (2), (4) y (5). Para el caso de la región H3, el ADN inmunoprecipitado es igual o menos enriquecido que los controles negativos. m; marcador de peso molecular.

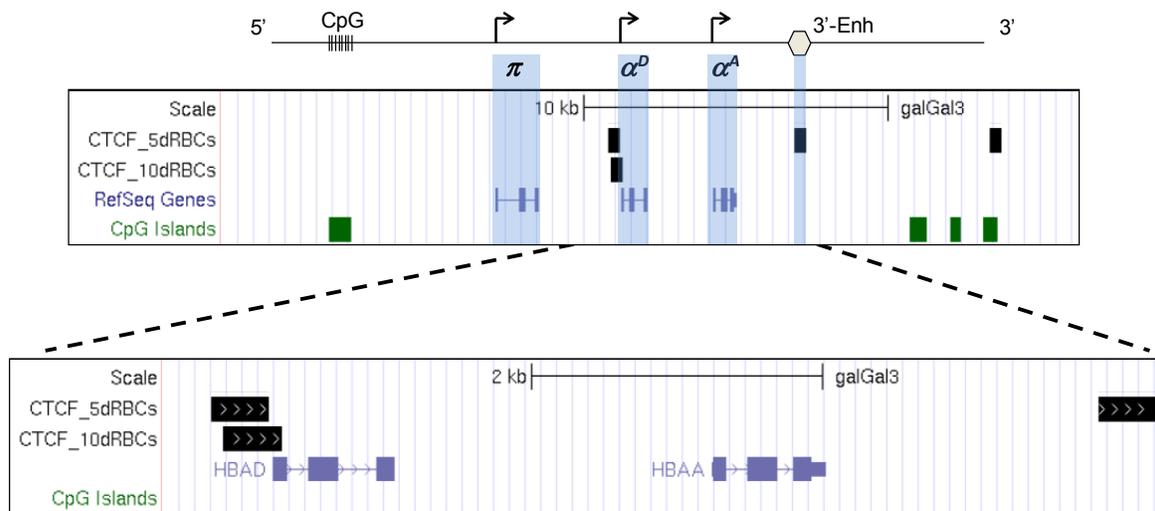
Con esta serie de análisis se pudo determinar que existe una amplia distribución de CTCF a lo largo del genoma del pollo tanto en los eritrocitos de 5 y 10 días de desarrollo (Tabla 3). Sin embargo, este factor se encuentra mayormente presente en los 10dRBCs lo cual representa un aspecto relevante para futuras investigaciones.

Tipo celular	5dRBCs	10dRBCs
Picos de CTCF (n)	5830	21,358
Picos constitutivos (n)		3433

**Tabla 3.** Número de sitios de unión del factor CTCF en el genoma de pollo. Estos números fueron calculados después de la aplicación de algoritmo denominado *peak calling*.

Una vez identificadas de las coordenadas de los picos de unión del factor CTCF, se decidió estudiar a un mayor detalle la distribución CTCF en el dominio  $\alpha$ -globina de pollo a través de las diferentes etapas del desarrollo. Para ello, los datos generados de CHIP-seq normalizados por el algoritmo de *peak calling* fueron introducidos en la plataforma de visualización *Genome Browser*. Como se muestra en la Figura 37, fue posible detectar la presencia de sitios de unión al

factor CTCF en algunas regiones del dominio tanto en los eritrocitos de 5 y 10 días. Para el caso de los 5dRBCs, se pudo observar que el factor se encuentra unido *in vivo* en la región del promotor del gen  $\alpha^D$ , así como en la región 3' río abajo que corresponde con el elemento *enhancer* 3'-Enh. En los 10dRBCs, CTCF fue localizado en la región del promotor  $\alpha^D$ . La presencia de CTCF en el promotor  $\alpha^D$  en ambas etapas del desarrollo indica que pudiera ser un sitio constitutivo de unión a dicho factor.



**Figura 37.** Distribución del factor CTCF en el dominio  $\alpha$ -globina. Utilizando la plataforma de visualización del genoma de pollo (*Genome Browser*, UCSC; *galGal3-2006*) fue posible la localización de los picos de unión al factor CTCF. Arriba el esquema muestra la posición de los tres genes  $\alpha$ -globina y el elemento 3'-Enh. Los recuadros de abajo representan una ampliación genómica del cromosoma 14 de pollo donde se localiza el dominio. Dentro de estos recuadros, las cajas negras corresponden con los sitios detectados de unión del factor CTCF. Los elementos azules representan a los genes globina donde las cajas azules corresponden con los exones. En recuadro en la parte inferior muestra una ampliación donde se evidencia la unión de CTCF en la región promotora del gen  $\alpha^D$  en ambos contextos del desarrollo del pollo.

Un aspecto muy relevante de estos resultados es el hecho de que el factor CTCF se encontró asociado en la región promotora del gen  $\alpha^D$  lo cual podría indicar su participación en la regulación de la expresión de dicho gen. Además, este sitio de unión a CTCF se encuentra en cercanía con la región de transición en la metilación del ADN presente en los 10dRBCs y en las células dif-HD3 (Figuras 30 y 33, respectivamente). Se ha descrito que la presencia de CTCF en la regiones promotoras de genes humanos como *p53* y *Retinoblastoma (Rb)* se encuentra asociada a la regulación de los genes y además participa en su protección contra la expansión de la metilación del ADN de las zonas aledañas<sup>126,128</sup>. Más aún, estudios de la distribución global de CTCF y otras marcas de cromatina han asociado la presencia de CTCF con zonas de transición entre diferentes tipos de cromatina, estados en la metilación del ADN y la organización topológica del genoma<sup>29,76,108,110,121,141</sup>. Por tal motivo, la presencia de CTCF en esta región del promotor  $\alpha^D$  y la transición de la metilación nos indica su probable papel en la delimitación del sub-dominio de metilación y la regulación en la expresión diferencial de los genes  $\alpha$ -globina.

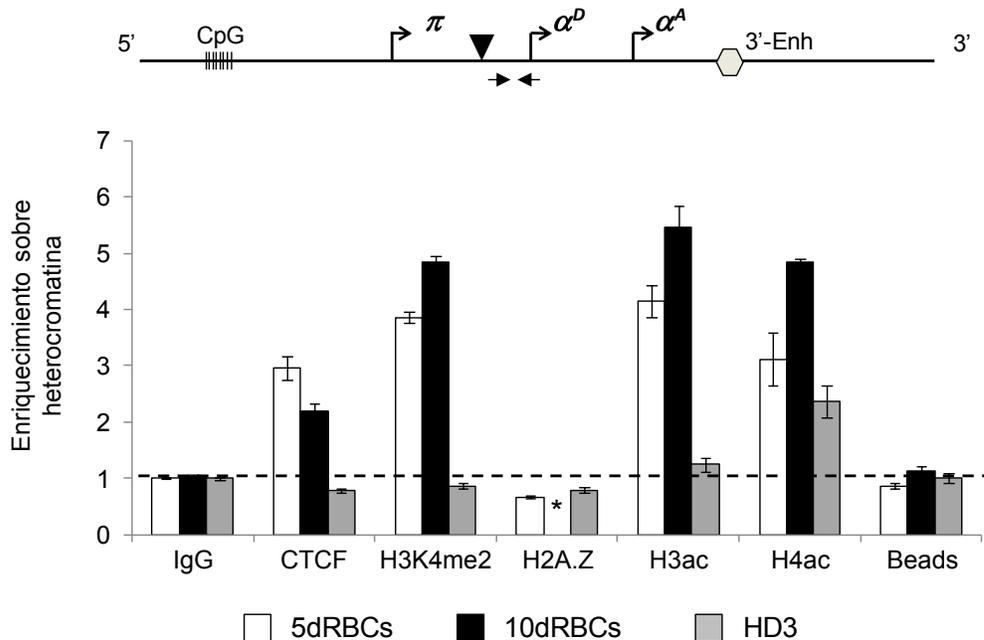
Es de mencionarse que gracias a estos experimentos de ChIP-seq de CTCF en el pollo, se pudo contribuir a una publicación alterna al presente trabajo. En conjunto con datos de ChIP-seq de CTCF generados en células de humano y de ratón, fue posible evidenciar múltiples sitios de unión a CTCF conservados evolutivamente entre los vertebrados y que además, son blanco de alteraciones en diversas enfermedades relacionadas al desarrollo de los organismos<sup>104</sup>.

### 3.2. Presencia de CTCF en la región de transición del sub-dominio de metilación.

Una vez determinada la presencia de CTCF en la región de la transición del sub-dominio de metilación, el estudio fue dirigido a comprender el estado en la estructura de la cromatina en esta región y su relación con la presencia del factor CTCF. Para ello, se decidió realizar experimentos de ChIP utilizando diversos anticuerpos específicos para distintas modificaciones post-traduccionales en la región amino-terminal de las histonas. Particularmente, para el análisis de las marcas en las histonas asociadas a cromatina abierta se emplearon anticuerpos para detectar la H3K4me2 (asociada a promotores y *enhancers* activos), la acetilación en las histonas H3 y H4 (H3ac y H4ac, respectivamente) y la variante de histona H2A.Z asociada a TSS y *enhancers*. Así mismo, se repitieron los ChIPs utilizando el anticuerpo de contra CTCF para corroborar los datos obtenidos a partir del ChIP-seq. El estudio fue dirigido a comprender los cambios en las diferentes etapas en el desarrollo (5 y 10dRBCs) y la diferenciación eritroide (HD3 y dif-HD3).

En los ChIPs realizados en los eritrocitos colectados a 5 y 10 días de desarrollo se pudo reproducir el enriquecimiento de CTCF en el promotor  $\alpha^D$  observado en los datos de ChIP-seq (Figura 38). En lo que respecta a las diferentes modificaciones post-traduccionales o marcas en las histonas, se pudo determinar que estas se presentan de manera constitutiva en el promotor  $\alpha^D$  con la excepción de la H2A.Z la cual no fue detectada en ninguna etapa del desarrollo.

De esta manera, se pudo establecer que la presencia de CTCF coincide con el establecimiento de una cromatina abierta.

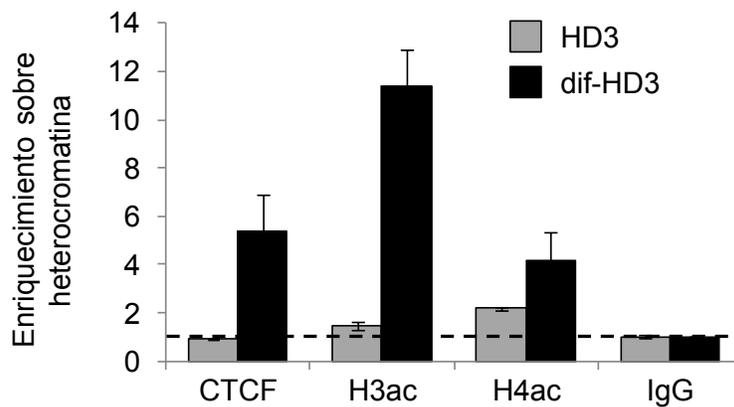


**Figura 38.** Análisis *in vivo* de la cromatina en el promotor  $\alpha^D$ . Mediante ensayos de ChIP se determinó la abundancia la cromatina abierta y la presencia del factor CTCF en la región del promotor  $\alpha^D$ . En el esquema de arriba se muestra la región analizada (flechas) y el triángulo invertido representa la cercanía con la región de transición de la metilación del ADN. Estos resultados fueron normalizados utilizando una región de heterocromatina como control de ruido de fondo (H3 en dominio  $\beta$ -globina; Figura 36) y dividido por los valor de los controles negativos IgG y *Beads*. De tal manera que la línea horizontal discontinua representa el nivel de ruido de fondo y todos los valores por arriba de esta se consideran enriquecimiento. La barra de error fue calculada a partir de dos o más experimentos independientes. El asterisco indica que no sea realizado el experimento.

En el caso de las células HD3, se observó que el factor CTCF no presenta unión en el promotor  $\alpha^D$  (Figura 38) coincidente con su estado de hipermetilación en el ADN y de inactividad transcripcional de dicho gen (Figuras 31 y 32; respectivamente). En lo que respecta a la estructura de la cromatina, no se observó la presencia de la H3K4me2 ni de la H3ac lo que pudiera ser indicativo

de una cromatina cerrada o no permisible a la transcripción (Figura 38). Sin embargo, se observó la presencia de la marca H4ac en esta región, lo que sugiere un cierto grado de apertura en esta zona. En su conjunto, estos resultados sugieren que la presencia de CTCF en el promotor de  $\alpha^D$  así como de su coexistencia con una cromatina abierta podrían participar en el establecimiento y delimitación del sub-dominio de metilación para la correcta expresión de los genes  $\alpha$ -globina adultos.

Para estudiar el papel de CTCF en la delimitación de la metilación del ADN y la regulación del gen  $\alpha^D$ , se decidió evaluar su presencia *in vivo* en las células HD3 inducidas a la diferenciación (dif-HD3). En esta condición de inducción se ha descrito la desmetilación activa del ADN en la región del promotor  $\alpha^D$  y su delimitación río arriba de este (Figuras 32-35). Mediante la aplicación de CHIP se observó que el factor CTCF se encuentra presente en el promotor  $\alpha^D$  en las dif-HD3 y no así en las células no inducidas HD3 (Figura 39). Además, al evaluarse las marcas de cromatina abierta, H3 y H4ac, se observó que estas se enriquecen en el promotor en las células inducidas a la diferenciación. Con estos resultados se puede vincular la presencia de CTCF en el promotor  $\alpha^D$  con: 1) la activación del gen  $\alpha^D$ , 2) en el establecimiento de una cromatina permisible a la transcripción y 3) en la creación de una zona de transición para delimitar el sub-dominio de metilación proveniente del gen embrionario  $\pi$ .



**Figura 39.** Asociación *in vivo* del factor CTCF en la inducción a la diferenciación. Se grafican los ChIPs de CTCF y las marcas de cromatina abierta. Los datos fueron normalizados de la misma manera que se ha descrito para la Figura 38.

En resumen, se ha determinado que el factor CTCF posee múltiples sitios de unión a lo largo del dominio  $\alpha$ -globina de pollo. Uno de ellos corresponde con una región genómica río arriba del TSS del gen  $\alpha^D$  (a 200 pb) y se encuentra presente en etapas tempranas y avanzadas del desarrollo del pollo (5 y 10 días). Coincidente con la presencia de CTCF en esta región, se observa un enriquecimiento de marcas de cromatina abierta sugiriendo su reclutamiento por la acción de CTCF. La presencia de CTCF y de la cromatina abierta en esta región, están también en coincidencia con las etapas donde se establece y delimita un sub-dominio de metilación que mantiene silenciado al gen embrionario  $\pi$  en las etapas adultas del desarrollo. Por tal motivo, resulta relevante el estudio del papel de CTCF en esta región de transición.

#### **4. Papel funcional de CTCF en el promotor del gen $\alpha^D$ .**

Una vez determinada la presencia del factor CTCF en promotor  $\alpha^D$ , el estudio fue enfocado a comprender su papel en esta zona de una manera funcional. Para ello, se decidió eliminar el sitio de unión presente en esta región y evaluar los efectos en términos de la actividad transcripcional del promotor  $\alpha^D$ , la metilación del ADN y la estructura de la cromatina. La eliminación del sitio fue preferida sobre la reducción de los niveles de la proteína CTCF ya este factor presenta diversos sitios de unión en el dominio lo que podría complicar la interpretación de los resultados. La estrategia seleccionada para este propósito fue la generación de células transgénicas a una serie de construcciones plasmídicas de ADN donde se incluye al promotor  $\alpha^D$  gobernando la actividad transcripcional de un gen reportero que codifica para proteína verde fluorescente (GFP). La versatilidad de esta estrategia es que las construcciones pueden ser modificadas en la región del promotor y evaluar el efecto de dicha alteración en términos de expresión de la GFP. Además, estas construcciones pueden ser integradas establemente en el genoma de las células, lo cual permite estudiar el efecto de la cromatina en la regulación de dichos transgenes.

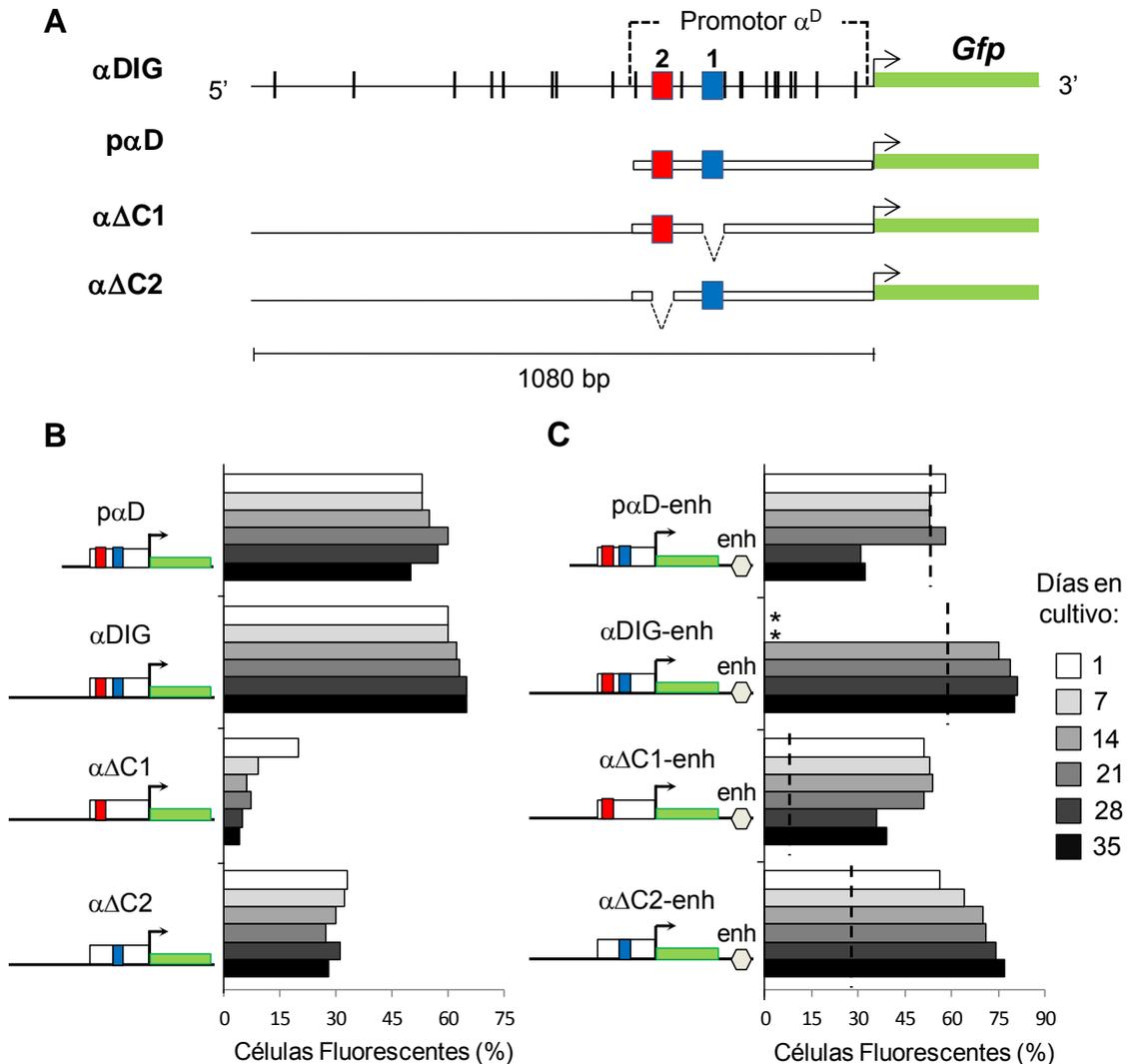


logo empleado fue generado a partir de comparaciones de los más de 15,000 sitios diferentes de unión a CTCF encontrados con en los datos de ChIP-seq en el presente trabajo (Tabla 3)<sup>104</sup>. A partir de la identificación de los probables sitios de unión en la secuencia promotora se elaboraron una serie de construcciones donde fueron eliminados de manera separada los dos motivos de unión a CTCF.

Para la clonación del promotor  $\alpha^D$  y sus versiones donde se eliminan los sitios de unión a CTCF, se realizaron PCRs y clonaciones en el vector de expresión transgénica que contiene al gen que codifica para la GFP, *pEGFP-1* (Figura 41A). Para conocer el nivel de actividad del promotor sin alteraciones se generó la construcción denominada  $p\alpha^D$ . Para analizar el efecto de la eliminación de los sitios de CTCF se diseñaron un par de construcciones donde fue removido de manera independiente ambos sitios de unión presentes en el promotor ( $\alpha\Delta C1$  y  $\alpha\Delta C2$ ; respectivamente). Adicionalmente, se decidió incluir en las construcciones, a la región intergénica río arriba de este donde se presenta la transición de la metilación (Figura 41A). Esto último con la intención de evaluar el papel de CTCF en el promotor  $\alpha^D$  y en presencia de la región intergénica donde ocurre la transición de la metilación. Cabe mencionar que en las construcciones donde fueron eliminados los sitios de unión a CTCF ( $\alpha\Delta C1$  y  $\alpha\Delta C2$ ) ninguno de los motivos para factores de transcripción previamente descritos como reguladores de la actividad del promotor  $\alpha^D$  fueron alterados<sup>159,166</sup>. Estas construcciones fueron transfectadas de manera estable en la línea celular HD3 donde se conoce que el promotor  $\alpha^D$  es activo en este tipo de vectores de expresión<sup>162,163</sup>. Para la selección de las células que integraron establemente los transgenes en su

genoma, se empleó la incubación de las mismas en presencia del antibiótico *geneticina* dado que las construcciones contienen codificado el gen de resistencia a dicho antibiótico.

Para el análisis de la actividad de las construcciones se evaluó el porcentaje de células fluorescentes en un citómetro de flujo o FACS (del inglés, *fluorescence emission by flow cytometry*). En las células que contienen a la construcción del promotor  $\alpha^D$  sin alteraciones ( $p\alpha^D$ ) se observó la expresión de GFP la cual se mantiene a través del tiempo (0-35 días; Figura 41B). El mismo escenario fue registrado para la construcción con el promotor intacto pero que incorpora la región intergénica ( $\alpha^{DIG}$ ). En contraste, cuando el sitio de unión a CTCF-1 fue removido ( $\alpha\Delta C1$ ) el porcentaje de células fluorescentes fue muy bajo y con una rápida extinción a través del tiempo (Figura 41B), lo que pone en manifiesto la participación funcional de CTCF en la actividad transcripcional del promotor  $\alpha^D$ . Para el caso de la eliminación del sitio de CTCF-2 ( $\alpha\Delta C2$ ), se observó una ligera reducción en el porcentaje de células positivas a la GFP, no obstante, el perfil de expresión se mantiene a través del tiempo. Con esta serie de resultados se pudo concluir que el sitio de unión a CTCF-1 (con mayor similitud al *logo* de CTCF en pollo) es el involucrado en la actividad transcripcional del promotor  $\alpha^D$ .



**Figura 41.** El factor multifuncional CTCF es un elemento clave para la actividad del promotor  $\alpha^D$ . **A)** Esquema de las construcciones generadas para el estudio del papel funcional de CTCF. Las cajas azul y roja representan los probables sitios de unión a CTCF identificados (ver Figura 39). En la construcción  $\alpha$ DIG se muestra la región promotora utilizada así como la región intergénica (IG) río arriba donde se ha identificado la transición en el sub-dominio de metilación. Las líneas verticales representan las CpGs. El esquema de las demás construcciones muestra la inclusión o eliminación de los sitios de unión a CTCF 1 (azul) y 2 (roja). **B)** Análisis del porcentaje de células fluorescentes por FACS. **C)** En las construcciones previamente descritas fue incluido por clonación el elemento 3'-Enh del dominio (rombo gris con la leyenda enh). Las líneas discontinuas verticales muestran el porcentaje promedio de células fluorescentes de las construcciones en **B**. Los asteriscos indican experimentos no realizados.

Para corroborar la ausencia del factor CTCF en las construcciones donde se realizó la eliminación de su sitio de unión, se decidió evaluar la presencia *in vivo* de CTCF en las diferentes células transfectadas con los transgenes. Mediante el ChIP de CTCF se observó que en la construcción que presenta actividad robusta del promotor ( $\alpha$ DIG) el factor CTCF se encuentra presente en la región del promotor del transgen (Figura 42B). En cambio, el factor CTCF no presenta unión al promotor en la construcción con la eliminación del sitio de unión 1 ( $\alpha\Delta$ C1; Figura 42B). Para conocer el estado de la cromatina en presencia o ausencia de CTCF, se realizaron ChIPs de las marcas en las histonas H3 y H4ac. Coincidente con la pérdida del factor CTCF en la construcción  $\alpha\Delta$ C1 se observó la ausencia de dichas marcas de cromatina abierta (Figura 42B), lo que sugiere que la presencia de CTCF en el promotor  $\alpha^D$  es necesaria para fomentar una estructura de la cromatina permisible a la transcripción.

Con la intención de distinguir el efecto de la eliminación de CTCF en el promotor  $\alpha^D$  de una posible afectación en la secuencia del promotor que compromete su funcionalidad transcripcional, se decidió incorporar un elemento tipo *enhancer* que puede fomentar un estado cromatínico adecuado para la transcripción. El raciocinio detrás de esta estrategia es que en la ausencia de CTCF el elemento *enhancer* puede ser capaz de compensar la actividad transcripcional del promotor. Para ello en todas las construcciones se incluyó el elemento 3'-Enh del dominio, el cual se sabe es capaz de *trans*-activar a los genes mediante la creación de un ambiente cromatínico óptimo a la transcripción<sup>19,163,166</sup>. Como se muestra en la Figura 41C, todas las

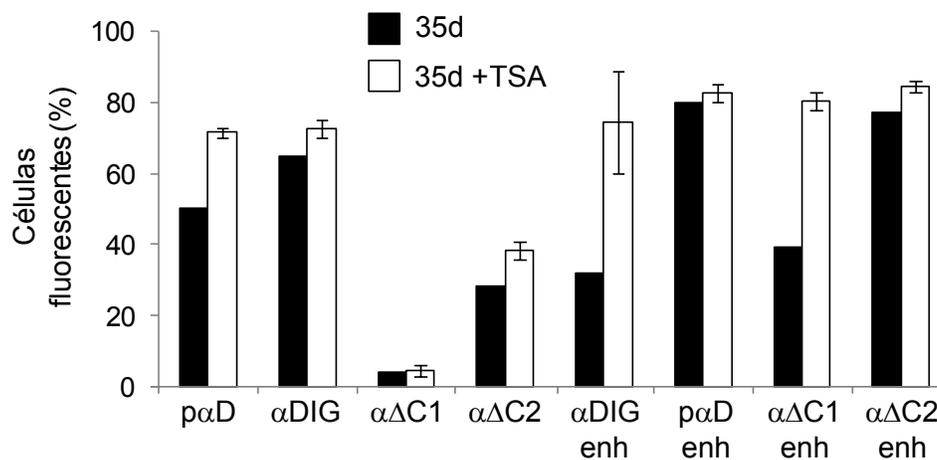


#### **4.1. CTCF establece un estado de la cromatina permisible para la actividad del promotor $\alpha^D$ .**

Aunque se ha descrito que la presencia de CTCF coincide con la incorporación de las marcas de acetilación en las histonas H3 y H4, tanto en las construcciones como en el promotor  $\alpha^D$  endógeno (Figuras 42 y 38, respectivamente). Y que dicha co-existencia se relaciona con la actividad transcripcional del promotor  $\alpha^D$ , fueron necesarios experimentos que pudieran evidenciar el posible papel de CTCF en la formación de una cromatina relajada o permisible a la transcripción (con las marcas de acetilación H3 y H4ac). Para comprobar la capacidad de CTCF de reclutar acetil-transferasas de histonas (HATs) en el promotor  $\alpha^D$ , se realizaron experimentos de inhibición de las desacetilasas de histonas (HDACs) en las células que portan las construcciones. El uso de inhibidores de las HDACs como lo es la Tricostatina-A o TSA, fomenta una hiperacetilación global de la cromatina con lo cual muchos genes se ven favorecidos en su transcripción. En el caso de que las HATs no pueden ser reclutadas por la ausencia de CTCF no se espera detectar un efecto en las células incubadas con dicho inhibidor.

Las células transfectadas con las construcciones con y sin el elemento 3'-Enh, fueron incubadas al día 35 del experimento con el inhibidor TSA. En las células que portan el transgen con la eliminación del sitio de CTCF ( $\alpha\Delta C1$ ) no fue posible inducir su reactivación con el tratamiento de TSA (Figura 43). En contraste, en las células con las construcciones que son capaces de unir a CTCF se observó un incremento en la actividad transcripcional del transgen (Figura 43). Este hecho pone de manifiesto que en la ausencia de CTCF, no hay un

reclutamiento de las HATs que se pudieran ver coadyuvadas o favorecidas en su acción por la presencia del TSA. Cuando los elementos capaces de reclutar a las HATs están presentes, capacidad que presenta el elemento 3'-Enh, el tratamiento con el TSA promueve mayores niveles de actividad transcripcional. Con estos resultados, podemos concluir que el papel de CTCF en el promotor del gen  $\alpha^D$  es el de fomentar un estado de la cromatina que favorece la actividad transcripcional y muy probablemente dicho tipo de cromatina puede contender contra la expansión del sub-dominio de metilación presente el gen embrionario  $\pi$ .



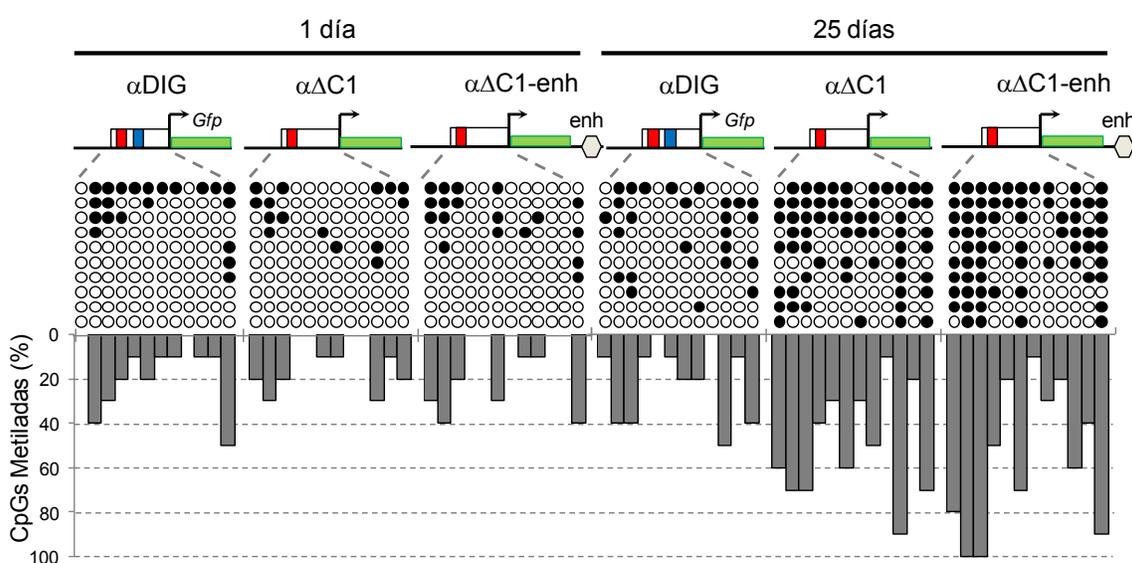
**Figura 43.** CTCF es necesario para la reactivación de los transgenes en presencia del inhibidor TSA. Tratamiento de las células transgénicas al día 35 de cultivo continuo con el inhibidor de las HDACs o TSA. La barra de error proviene de dos experimentos de incubación de las células con el TSA.

## 4.2. Participación de CTCF en la delimitación de la metilación del ADN en el promotor $\alpha^D$ .

Como se ha mencionado anteriormente, una de las funciones descritas al factor CTCF es la de contribuir al establecimiento de una cromatina relajada o abierta mediante el reclutamiento de remodeladores de la cromatina como los son las HATs y las HMTs. Además, se ha observado que el factor CTCF tiene la capacidad de mantener a secuencias promotoras libres de metilación de ADN<sup>57,127,128</sup>. Por tal motivo, en el presente trabajo se decidió analizar el papel de CTCF en la delimitación de la metilación del ADN presente en la región intergénica río arriba al promotor  $\alpha^D$ . Para comprender su participación activa en la delimitación de la metilación del ADN, se evaluaron los niveles de metilación en las células transgénicas con las construcciones del promotor  $\alpha^D$  a través del tiempo en cultivo. Esto fue posible gracias a que en el diseño de las construcciones se incorporó la región intergénica 5' río arriba del promotor  $\alpha^D$  donde se presenta la transición en la metilación del ADN (Figura 42A).

Mediante la técnica de bisulfito de sodio, amplificación por PCR y secuenciación Sanger, fue posible determinar el estado en la metilación en las construcciones. Al primer día de cultivo celular, las construcciones analizadas no presentaron una metilación significativa en el promotor  $\alpha^D$  (Figura 44), inclusive la construcción  $\alpha\Delta C1$  donde los niveles de expresión del transgen son muy bajos y CTCF no se encuentra unido (Figura 41B y 42B; respectivamente). Cuando se analizó la metilación al día 25 de cultivo, únicamente la construcción que presenta

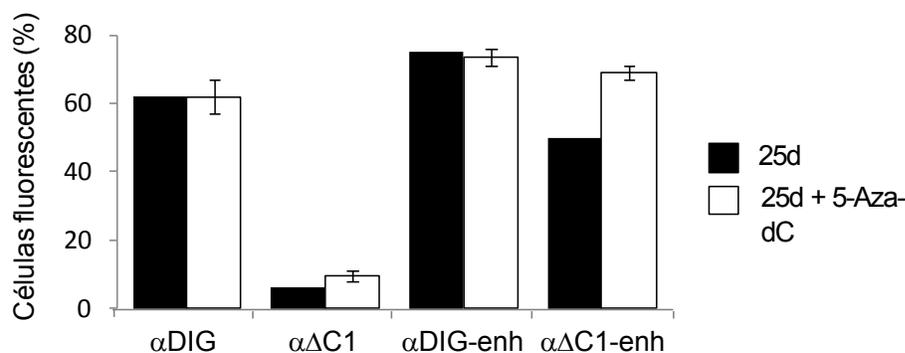
unión a CTCF ( $\alpha$ DIG) se mantuvo libre de metilación. Por el contrario, las construcciones  $\alpha\Delta$ C1 y  $\alpha\Delta$ C1-enh, donde CTCF no es capaz de unirse al promotor, mostraron niveles elevados de metilación del ADN (Figura 44). Por tal motivo, es posible concluir que la presencia de CTCF es necesaria para el mantener al promotor libre de metilación y promover una cromatina permisible para la transcripción.



**Figura 44.** CTCF mantiene libre de metilación al promotor  $\alpha^D$  en las células transgénicas. Análisis de la metilación en la región promotora de las construcciones al día 1 y 25 de cultivo.

Para comprender el impacto que tiene la ganancia de metilación en la actividad transcripcional de las construcciones se decidió evaluar el efecto de la reducción en la metilación adquirida al día 25 de cultivo de las células transgénicas. Para ello, las células fueron tratadas con el fármaco 5-azadC, el cual es un agente inhibidor de la metilación del ADN. Cuando las células que

contienen la construcción  $\alpha\Delta C1$  fueron tratadas con la 5-azadC se observó una ligera reactivación del gen reportero *GFP* (Figura 45), no obstante que en estas células el promotor se encuentra hipermetilado. Sin embargo, cuando las células con la construcción que posee al *enhancer* ( $\alpha\Delta C1$ -enh) fueron tratadas con el inhibidor, se registró una reactivación del transgen. Una posible interpretación de estos resultados es que en ausencia de CTCF, la reducción de la metilación por efectos de la 5-azadC no puede promover una reactivación del transgen debido a la necesidad del reclutamiento de otros remodeladores de la cromatina tal como las HATs, demostrado anteriormente. Por ello, en la construcción que porta al elemento 3'-enh, el *enhancer* presente puede promover un ambiente cromatínico óptimo para la transcripción.



**Figura 45.** Reactivación de los transgenes mediante la inhibición de la metilación. Las células fueron incubadas en presencia del inhibidor de la metilación 5-azadC durante 72 horas. La barra de error fue calculada a partir de dos experimentos.

Con esta serie de resultados podemos concluir que la presencia de CTCF en el promotor del gen  $\alpha^D$  es necesaria para mantener al promotor libre de metilación. No obstante, la metilación del ADN en el promotor  $\alpha^D$  no se considera

como el único mecanismo responsable para la inactivación de los transgenes (y del gen endógeno  $\alpha^D$ ), ya que al reducir sus niveles no es posible reactivar la actividad transcripcional. Esto es un indicativo de que, además de prevenir la metilación del ADN, el factor CTCF es necesario para el establecimiento de una cromatina que permite actividad transcripcional del promotor  $\alpha^D$ .

#### **4.3. Resumen: Distribución de la metilación del ADN, presencia de CTCF y la estructura de la cromatina.**

En el presente trabajo de tesis se ha demostrado que, para en la expresión diferencial de los genes  $\alpha$ -globina en las etapas adultas del desarrollo del pollo, se establece un sub-dominio de metilación del ADN que conlleva al reclutamiento de la proteína MeCP2 y al silenciamiento del gen embrionario  $\pi$ . Dicho sub-dominio de metilación se extiende hasta una región intergénica a 500 pb río arriba del promotor  $\alpha^D$ . Los datos generados a partir del estudio de CTCF y su función en el promotor  $\alpha^D$  nos han llevado a proponer lo siguiente. El papel de CTCF en el promotor del gen  $\alpha^D$  es el de impedir la expansión del sub-dominio de metilación que mantiene silenciado al gen embrionario  $\pi$ . Para llevar a cabo esta función, la presencia de CTCF podría estar vinculada al establecimiento de una cromatina abierta o relajada mediante la hiperacetilación de las histonas y otras marcas.

Es importante mencionar en este punto que estos resultados permiten tener un mejor entendimiento de la manera en que los genes  $\alpha$ -globina son regulados en las etapas adultas del desarrollo. Sin embargo, poco es lo que sabe

de la regulación epigenética que toma lugar en las etapas tempranas del desarrollo, donde el gen embrionario  $\pi$  es activo a la transcripción y los genes adultos se encuentran silenciados. Los resultados hasta aquí descritos indican que la metilación del ADN no se encuentra involucrada en el silenciamiento de los genes adultos en esta etapa temprana (Figura 30). Además, la unión de CTCF al promotor del gen  $\alpha^D$  es constitutiva, es decir, se presenta en ambas etapas del desarrollo del pollo. Por tal motivo, creemos que la regulación de los genes en las etapas tempranas podría involucrar mecanismos distintos a la metilación del ADN, como lo pueden ser la organización tri-dimensional y la intervención de los elementos distales presentes en el dominio  $\alpha$ -globina. Este aspecto se aborda a continuación.

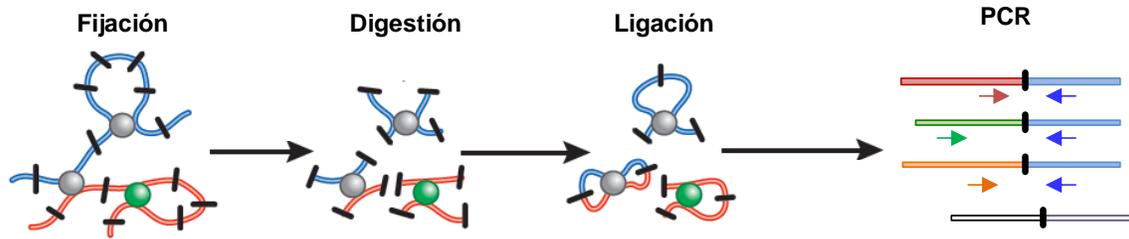
## **5. Organización espacial de los genes $\alpha$ -globina y el elemento 3'-Enh a través del desarrollo del pollo.**

Para la correcta expresión de los genes  $\alpha$ -globina en tiempo y tejido específico se han descrito, en el presente trabajo de tesis, la intervención de diversos procesos como lo son la metilación del ADN, la presencia de factores de transcripción y la estructura de la cromatina. Sin embargo, poco es lo que se sabe del papel que juegan los demás elementos de regulación contenidos en el dominio  $\alpha$ -globina. Entre ellos, el elemento 3'-Enh se ha caracterizado como un *enhancer* capaz de conferir niveles elevados de transcripción a los tres genes globina<sup>163,165,166</sup>. Se conoce que la actividad de este elemento es modulada y regulada por la

estructura de la cromatina presente en él y la acción de factores de transcripción como lo son GATA-1 y YY1<sup>163,165,166</sup>. Sin embargo, se desconoce la manera en que este elemento puede *trans*-activar a los genes globina en las diferentes etapas del desarrollo del pollo. Se sabe que un mecanismo de acción de los *enhancers* es el de establecer contactos distales con los promotores de los genes mediante la formación de *loops* o asas de cromatina<sup>32,90,140</sup>. Estos contactos se cree que sirven para transferir a los promotores los factores necesarios y a la maquinaria de transcripción presentes en los *enhancers*, como por ejemplo el complejo Mediador y la ARN pol II<sup>15,178,179</sup>. Un estudio previo del grupo del Dr. Sergey Razin y colaboradores pudo determinar que este elemento puede establecer contactos físicos con el promotor del gen  $\alpha^D$  en las etapas avanzadas del desarrollo, postulando que esta interacción participa en la regulación transcripcional de dicho gen<sup>171</sup>. Sin embargo, no se ha determinado la manera en que el 3'-Enh puede *trans*-activar al gen embrionario. En particular, se desconoce si el elemento 3'-Enh es capaz de establecer contactos con el gen embrionario  $\pi$  para su *trans*-activación en las etapas tempranas del desarrollo. Para abordar este punto, en el presente trabajo se decidió determinar la distribución espacial de los promotores de los genes  $\pi$  y  $\alpha^D$  y su relación con el elemento 3'-Enh. De esta manera se pretende integrar los datos de la metilación y estructura de la cromatina aquí descritos con la participación de la organización espacial de estos elementos en las diferentes etapas del desarrollo del pollo.

### **5.1. Establecimiento de la técnica experimental: Captura conformacional de cromosomas (3C), para el estudio de la organización espacial del dominio $\alpha$ -globina.**

Con la intención de conocer la organización espacial de los genes globina se empleo la técnica conocida como captura conformacional de cromosomas, conocida como 3C. La técnica de 3C consiste en la fijación *in vivo* de las interacciones presentes en la organización espacial del genoma en el núcleo mediante el uso del formaldehído como un agente químico fijador (Figura 46)<sup>135</sup>. Estas interacciones son aisladas mediante el uso de endonucleasas de restricción quedando en asociación aquellas regiones que estaban en contacto, constituyendo así una macro-molécula con extremos cohesivos de ADN. Posteriormente, estas interacciones son ligadas intra-molecularmente por sus extremos cohesivos constituyendo híbridos entre las regiones del genoma que estuvieron en cercanía<sup>135</sup>. Para evaluar la formación de híbridos entre dos elementos que se encuentran en cercanía al interior del núcleo, se diseñan oligonucleótidos específicos a los híbridos y mediante PCR se puede comprobar su formación <sup>135,136</sup>(Figura 46). De esta manera, sí se presenta una amplificación por PCR, significa que existe una interacción física *in vivo* entre dichos elementos.

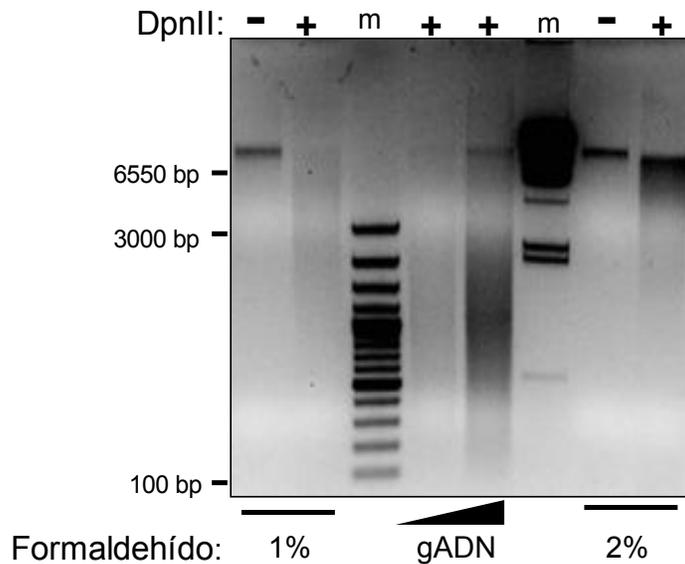


**Figura 46.** Captura conformacional de cromosomas (3C). Con la fijación de las interacciones del genoma al interior del núcleo, su digestión con endonucleasas de restricción y la ligación intra-molecular, se pueden obtener moléculas de ADN híbridas las cuales pueden ser analizadas por PCR. Las flechas representan los oligonucleótidos diseñados para detectar la formación de los híbridos de ADN. Tomada y modificada de<sup>139</sup>.

Para conocer si el elemento 3'-Enh puede establecer contactos distales con los promotores y contribuir a la expresión diferencial de los genes a través del desarrollo, se utilizó el protocolo de 3C en eritrocitos colectados al día 5 y 10 del desarrollo (5dRBCs y 10dRBCs; respectivamente). Para detectar específicamente los posibles contactos entre el elemento 3'-Enh y el promotor de del gen  $\pi$ , así como del 3'-Enh con el promotor  $\alpha^D$ , se realizó un análisis *in silico* para identificar los sitios de corte de diversas enzima endonucleasas de restricción en la región genómica de los genes globina y el 3'-Enh. Para la selección de la endonucleasa a utilizar sus sitios de corte deben estar presentes o en la cercanía a los elementos a ser analizados. Con base en este criterio se seleccionó la endonucleasa DpnII, ya que sus sitios de corte se distribuyen ampliamente a través de la región de estudio (Figura 48A). Además, la enzima DpnII reconoce una secuencia de ADN de cuatro nucleótidos (secuencia 5'-GATC-3'), lo que permite separar de una mejor manera las interacciones fijadas, aumentando

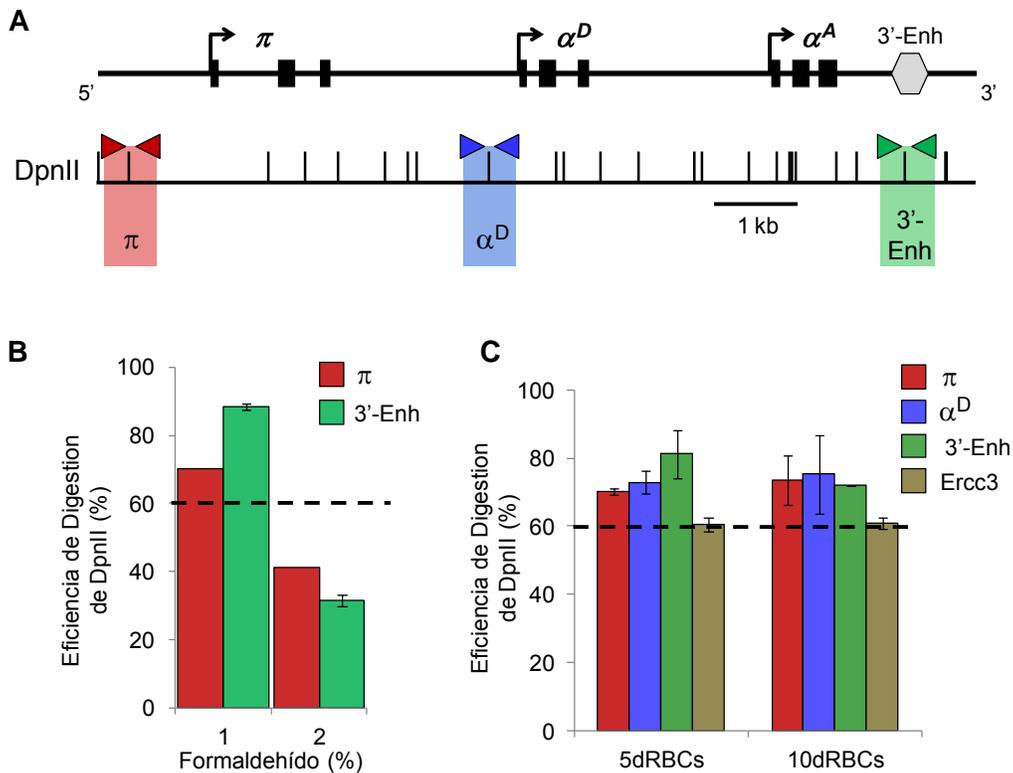
considerablemente la resolución del experimento por encima del uso de endonucleasas convencionales que reconocen secuencias de 6 nucleótidos<sup>175</sup>.

Una vez considerada la enzima de restricción a emplear, las células colectadas fueron fijadas con la adición de formaldehído. Es importante mencionar que la concentración (porcentaje) de formaldehído a utilizarse debe ser evaluada para no interferir con la accesibilidad de la endonucleasa de restricción<sup>174,175,180</sup>. Por tal motivo, se analizó la accesibilidad de la enzima DpnII y su frecuencia de corte en las regiones de interés. Se evaluó la eficiencia de digestión de la enzima en células fijadas con una solución de formaldehído al 1 y 2%. Posterior a la fijación, se aislaron los núcleos de las células y se incubaron con aproximadamente 800 U de la enzima DpnII por toda la noche. El ADN fue purificado como se especifica en la sección de Material y Métodos. En primera instancia, la digestión fue analizada mediante el patrón de migración electroforética del ADN en geles de agarosa (Figura 47). De esta manera, se pudo determinar que el tratamiento de las células con 2% de formaldehído tiene un efecto inhibitorio en la actividad de corte de la enzima DpnII.



**Figura 47.** Control de digestión de la enzima DpnII en células fijadas. Gel de agarosa al 0.6%. Además del ADN digerido con DpnII (+) proveniente de las células fijadas con formaldehído, para fines comparativos se incluyó un control de ADNg digerido donde se presenta el mayor grado de digestión de DpnII ya que la digestión se realizó en ADN purificado (sin cromatina). Tinción con bromuro de etidio.

Para la determinación cuantitativa de la eficiencia de digestión, se realizaron ensayos de PCR en tiempo real empleando oligonucleótidos diseñados en ambos extremos de los sitios corte de DpnII en los elementos de estudio. Como se muestra en la Figura 48, fue posible determinar que el entrecruzamiento con 2% de formaldehído tiene un efecto negativo en la actividad de corte de la enzima DpnII. En contraste, las células fijadas con la solución al 1% de formaldehído presentaron una eficiencia de digestión superior al 60% en las regiones de estudio. Se ha reportado que una eficiencia de digestión menor al 60% puede introducir sesgo en los experimentos de 3C<sup>174,180</sup>, por lo que se decidió utilizar 1% de formaldehído en todos los experimentos realizados.

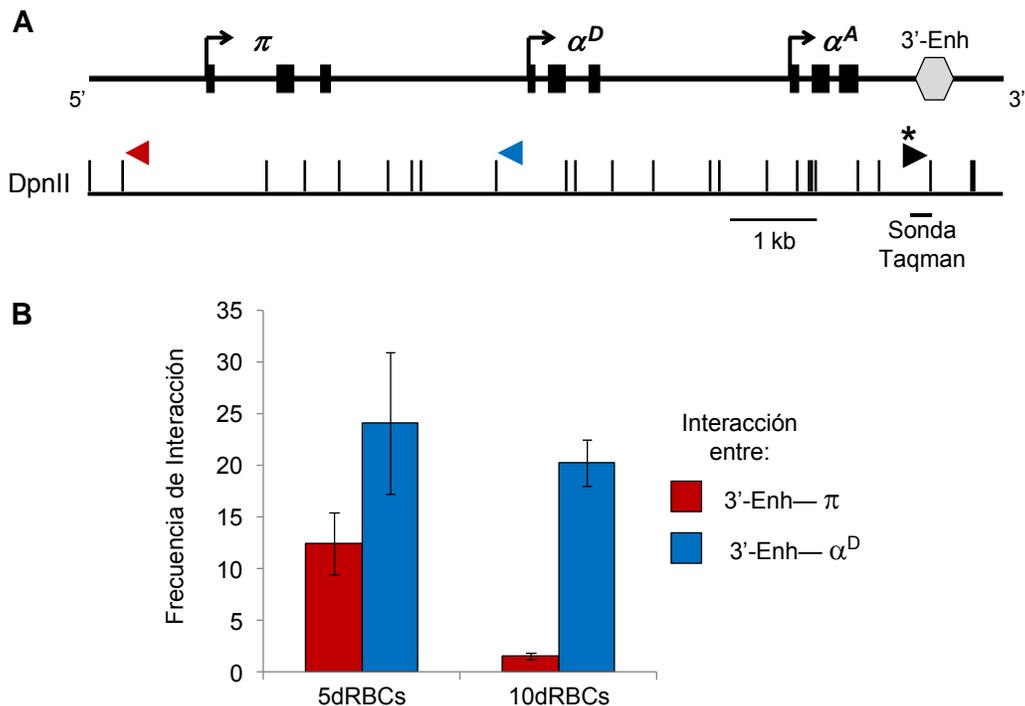


**Figura 48.** Eficiencias de digestión en las células fijadas con formaldehído. **A)** Esquema de los genes  $\alpha$ -globina de pollo y del elemento 3'-Enh donde se muestran los sitios de corte para la endonucleasa DpnII. Las puntas de flecha de colores representan los oligonucleótidos diseñados para los análisis mediante PCR en tiempo real. **B)** Comparación de la eficiencia de digestión de la enzima DpnII en células fijadas con 1 y 2% de formaldehído, respectivamente. **C)** Eficiencia de digestión en los experimentos de 3C en las células 5dRBCs y 10dRBCs fijadas con 1% de formaldehído. Los colores utilizados en B y C corresponden con oligonucleótidos empleados que están indicados en el esquema A. Las líneas horizontales discontinuas indican el porcentaje mínimo para considerarse una digestión eficiente<sup>174</sup>. Las barras de error fueron calculadas a partir de dos o más experimentos independientes.

## 5.2. Interacción física entre el 3'-Enh y los promotores de los genes $\alpha$ -globina mediante *loops* de cromatina.

Con las condiciones experimentales estandarizadas, se realizaron los ensayos de 3C en eritrocitos colectados de embriones de pollo al día 5 y 10 de desarrollo (5dRBCs y 10dRBCs, respectivamente). Una vez digeridos los núcleos de estas células con la enzima DpnII, se evaluaron las eficiencias de digestión en los elementos de estudio para verificar que estas fueran superiores al 60% que es lo recomendado para este tipo de experimentos<sup>174,180</sup> (Figura 48C). La ligación intra-molecular y la purificación del ADN se realizaron de acuerdo con trabajos previamente publicados<sup>136,174,180</sup>. Para evaluar las frecuencias de interacción entre el elemento 3'-Enh y los promotores de los genes  $\pi$  y  $\alpha^D$  mediante PCR en tiempo real, se diseñó un oligonucleótido sentido (*forward*) en la cercanía al sitio de corte DpnII en el elemento *enhancer*, el cual es combinado con su contraparte anti-sentido (*reverse*) en cada uno de los promotores en cuestión (Figura 49A). Por ejemplo, la combinatoria del oligonucleótido *forward* en el 3'-Enh y su contraparte *reverse* en la cercanía al sitio de corte DpnII en el promotor  $\pi$ , permitirá amplificar por PCR el híbrido de ADN generado a partir de la interacción del elemento 3'-Enh y el promotor  $\pi$ . De esta manera se pudo determinar que en las etapas tempranas del desarrollo, en los 5dRBCs se presenta un *loop* de cromatina entre el elemento 3'-Enh y el promotor del gen  $\pi$  (Figura 49B). Este contacto es específico de la etapa temprana del desarrollo ya que en los 10dRBCs la frecuencia de interacción entre el 3'-Enh y el promotor  $\pi$  fue considerablemente menor (Figura 30B). Este resultado es consistente con datos previos donde se ha

descrito que el gen  $\pi$  se encuentra libre de metilación y activo transcripcionalmente (Figura 30B). Por tal motivo, proponemos que el contacto entre el elemento 3'-Enh es importante para la activación transcripcional del gen embrionario, y la ausencia de dicho contacto en las etapas avanzadas contribuye al silenciamiento del gen.



**Figura 49.** Formación de *loops* de cromatina entre el elemento *enhancer* y los promotores de los genes  $\pi$  y  $\alpha^D$ . **A)** Esquema de los genes  $\alpha$ -globina de pollo y del elemento 3'-Enh donde se muestran los sitios de corte para la endonucleasa DpnII. Las puntas de flecha de colores representan los oligonucleótidos diseñados para los análisis mediante PCR en tiempo real. El asterisco señala el oligonucleótido que se emplea de manera constante (*forward*) en las reacciones de PCR en combinación con alguno de los diseñados en los promotores  $\pi$  y  $\alpha^D$  (*reverse*; puntas de flechas roja y azul; respectivamente). **B)** Frecuencias de interacción entre el elemento *enhancer* y los promotores de los genes. En las PCRs en tiempo real se empleó una sonda tipo Taqman como un reportero fluorescente de la amplificación de PCR<sup>174</sup>. Esta fue diseñada en la región de amplificación del 3'-Enh (A). Las barras de error fueron generadas a partir de dos experimentos independientes.

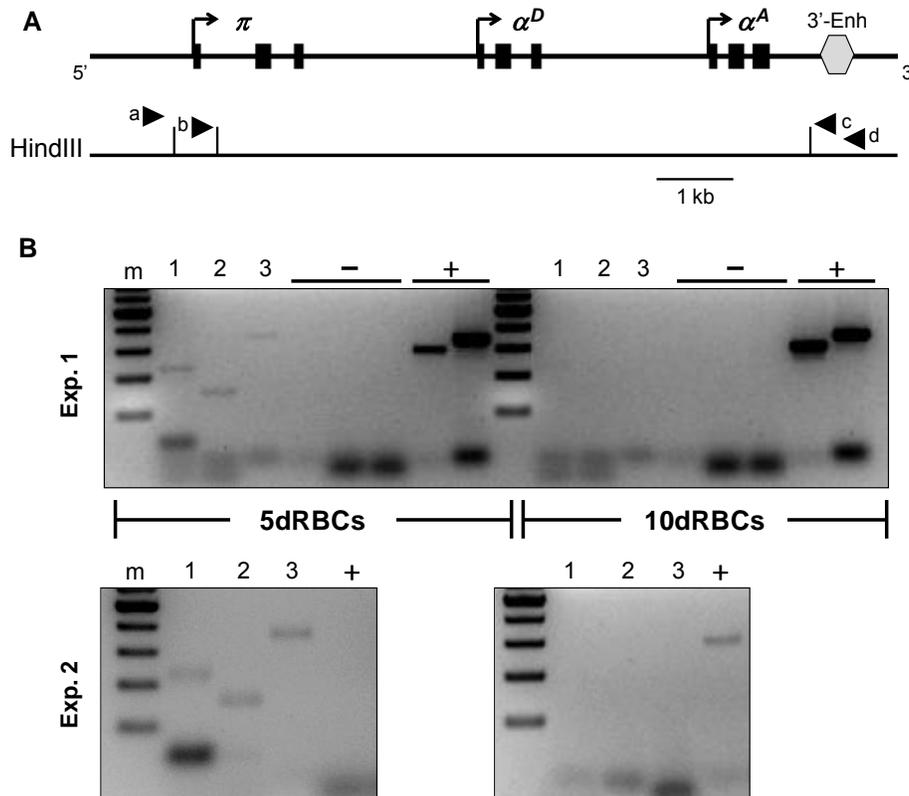
De manera interesante, el elemento 3'-Enh también presentó interacción con el promotor del gen  $\alpha^D$  en los 5dRBCs (Figura 49B). Lo cual sugiere que ambos promotores ( $\pi$  y  $\alpha^D$ ) convergen espacialmente y son contactados por el elemento *enhancer*. Sin embargo, el gen  $\alpha^D$  no presenta actividad transcripcional en esta etapa del desarrollo, por lo que la selectividad en la *trans*-activación del *enhancer* con los promotores de los genes en esta etapa representa un aspecto que debe ser abordado en futuras investigaciones.

En las etapa avanzada del desarrollo fue posible caracterizar el contacto del elemento 3'-Enh exclusivamente con el promotor del gen  $\alpha^D$  (Figura 49B), lo que sugiere que este contacto participa en la activación transcripcional del gen  $\alpha^D$ . En esta etapa, el 3'-Enh no presentó contacto con el gen embrionario, el cual se encuentra hipermetilado y silenciado en su expresión (Figura 30B).

### **5.3. Corroboración del *loop* diferencial entre el elemento 3'-Enh y el gen embrionario $\pi$ .**

Para comprobar que el elemento enhancer establece contactos diferenciales con el gen embrionario a través del desarrollo del pollo, se decidió repetir los ensayos de 3C ocupando una enzima de restricción diferente lo cual ayudaría a descartar un sesgo experimental en los resultados obtenidos. Para ello, se empleó la endonucleasa HindIII la cual presenta sitios de corte en la región del elemento 3'-Enh y en el promotor del gen  $\pi$  (Figura 50A). De manera reproducible y similar a lo observado con la enzima DpnII, se pudo detectar un contacto entre el elemento

3'-Enh y el promotor  $\pi$  en la etapa temprana del desarrollo del pollo (Figura 50B). Dicho contacto se pierde en las células 10dRBCs, indicando que el contacto es específico de la etapa temprana donde el gen embrionario es activo transcripcionalmente.



**Figura 50.** Validación del *loop* de cromatina entre el elemento 3'-Enh y el promotor  $\pi$ . **A)** Esquema de los genes  $\alpha$ -globina de pollo y del elemento 3'-Enh donde se muestran los sitios de corte para la endonucleasa HindIII. Las puntas de flecha representan los oligonucleótidos diseñados. **B)** Experimentos de 3C con la enzima HindIII en 5 y 10dRBCs. Los geles de agarosa muestran los productos de PCR punto final. Empleando las combinatorias de oligonucleótidos siguientes: 1) a y c; 2) b y c; 3) a y d; se pudo detectar una amplificación de los productos de la interacción exclusivamente en los 5dRBCs. Se incluyen controles negativos para las combinatorias 1-3 el cual consta de ADNg (-). El control positivo (+) de amplificación se trata de una región genómica que no es digerida por la enzima HindIII. m= Marcador de peso molecular.

En su conjunto, estos resultados permiten concluir que existen cambios en la organización espacial de los elementos de regulación contenidos en el dominio  $\alpha$ -globina que contribuyen a la expresión diferencial de los genes  $\alpha$ -globina a través de las etapas del desarrollo del pollo.

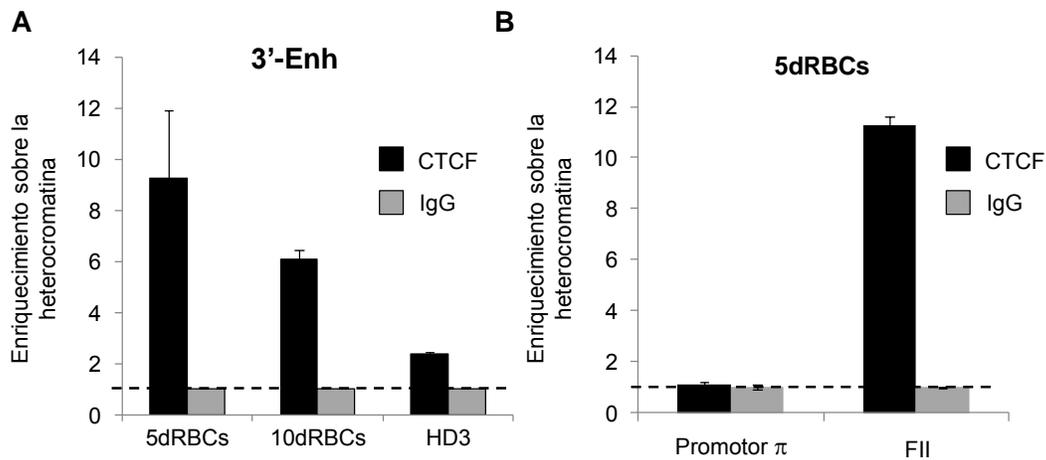
#### **5.4. Presencia de CTCF en las regiones que participan en los *loops* de cromatina.**

En la formación de *loops* o asas de cromatina que permiten los contactos entre elementos distales en el genoma, se ha descrito que existen proteínas y complejos que son capaces de establecer dichos contactos moleculares. Entre esas proteínas se encuentra CTCF y los complejos de cohesinas<sup>15,108,109,179</sup>. Con la intención de tener una idea más detallada de los procesos involucrados en la estructuración espacial del dominio  $\alpha$ -globina, se decidió analizar si CTCF pudiera participar en los contactos detectados entre el elemento 3'-Enh y los promotores de los genes  $\pi$  y  $\alpha^D$ . Para ello se realizaron ChIP de este factor y se evaluó su presencia en las regiones de los promotores  $\pi$  y  $\alpha^D$ , así como en el elemento 3'-Enh. Para el caso del promotor del gen  $\alpha^D$ , como ha sido descrito anteriormente en el presente trabajo, el factor CTCF presenta unión de manera constitutiva (5 y 10dRBCs; Figuras 37 y 38). Este hecho, correlaciona con el tipo de interacciones observadas para este promotor, ya que se encuentra de manera constitutiva en contacto con el elemento 3'-Enh (Figura 49B).

Derivado de los análisis de ChIP-seq de CTCF, se observó que este factor también se encuentra unido en la región del elemento 3'-Enh en los 5dRBCs. Para corroborar este resultado, se realizaron ensayos de ChIPs de CTCF en las células 5 y 10dRBCs y se analizaron mediante PCR en tiempo real. De manera inesperada, se detectó reproduciblemente la unión de CTCF en ambas etapas del desarrollo (Figura 51), cuando en los experimento de ChIP-seq se presentó exclusivamente en los 5dRBCs (Figura 37). Estas diferencias entre los experimentos pueden ser explicadas por aspectos técnicos al análisis del ChIP-seq; cuando se emplea un método de *peak calling* pueden ser asignados o descartados sitios de unión de acuerdo a los parámetros de análisis seleccionados<sup>71,181</sup>. Esta presencia constitutiva del factor CTCF en el elemento 3'-Enh podría estar involucrada en los contactos que presenta el *enhancer* con los promotores en ambas etapas del desarrollo. De hecho, proponemos que la presencia de CTCF en el elemento 3'-Enh en los 5dRBCs puede ser el responsable de mediar el contacto con el promotor de gen  $\pi$  ya que CTCF no es capaz de unirse en dicho promotor (Figura 51).

Colectivamente, estos resultados permiten proponer que el factor multifuncional CTCF podría estar involucrado en la organización espacial del dominio. En etapas tempranas del desarrollo, cuando el gen  $\pi$  es activo a la transcripción, la presencia de CTCF tanto en el promotor  $\alpha^D$  y en el elemento 3'-Enh podría promover los contactos entre estos elementos aunado con la incorporación del promotor del gen  $\pi$  en dicha interacción. En las etapas avanzadas del desarrollo, donde los genes adultos son activos, la unión de CTCF

en el promotor  $\alpha^D$  y en 3'-Enh podría intervenir en la formación de un asa de cromatina entre estos elementos. Adicionalmente, la presencia de CTCF en el promotor de  $\alpha^D$  parece participar en la delimitación del sub-dominio de metilación. Sin duda, son necesarios estudios detallados donde se pueda abordar el papel funcional de CTCF en la constitución de este tipo de interacciones y la delimitación de la estructura de la cromatina.



**Figura 51.** Análisis de la unión de CTCF al elemento 3'-Enh y en el promotor del gen  $\pi$ . A) ChIPs de CTCF y su análisis en el elemento 3'-Enh. B) ChIP de CTCF en 5dRBCs, el enriqueciendo fue evaluado en el promotor del gen  $\pi$ . Como control positivo del experimento se muestra el enriquecimiento de CTCF en el *insulator* 3'-CHS4 (FII, Figura 35). El método de normalización es el mismo descrito para la Figura 38. Las barras de error fueron generadas a partir de dos o más experimentos independientes.

## XI. Discusiones

La información genética necesaria para la ejecución de todos los procesos biológicos está contenida en el genoma de los organismos. En los eucariontes, el genoma consta de varias moléculas de ADN, conocidas como cromosomas, donde se codifican los genes para las proteínas y para los ARNs no-codificantes implicados en la gran variedad de funciones y estructuras que constituyen a los seres vivos. Compartiendo el mismo genoma, se ha descrito una gran diversidad de tejidos y tipos celulares que conforman a los organismos. Esto es posible gracias a que cada tipo celular muestra un programa particular de expresión de los genes el cual es distinto entre los tipos celulares. Por tal motivo, para la comprensión de los procesos biológicos tales como el desarrollo de los organismos, es importante entender los procesos y mecanismos implicados en el establecimiento de dichos programas de expresión diferencial de los genes.

Se ha descrito que tanto los elementos de regulación contenidos en el genoma así como la manera en que el genoma se encuentra compactado en cromatina, participan en el establecimiento de dichos programas de expresión génica<sup>35</sup>. Por tal motivo, resulta importante realizar investigaciones que permitan comprender la manera en que dichos elementos de regulación, el estado de la cromatina y otros factores pueden intervenir en la correcta expresión de los genes en los diferentes tipos celulares que constituyen a los seres vivos. En el presente trabajo de tesis, se empleó al grupo de genes  $\alpha$ -globina como modelo de estudio para comprender el papel de los elementos contenidos en el ADN y la estructura de la cromatina en la regulación de la expresión de estos genes.

## 1. El papel de la metilación del ADN en el silenciamiento del gen embrionario

$\pi$ .

El dominio  $\alpha$ -globina de pollo representa un modelo muy interesante para el estudio de los mecanismos y procesos implicados en la expresión diferencial de los genes globina. En las etapas tempranas del desarrollo del pollo, las células eritroides en circulación presenta la expresión del gen  $\pi$ , mientras los genes adultos  $\alpha^D$  y  $\alpha^A$  se encuentran inactivos. Este patrón de expresión se ve modificado en las etapas avanzadas o tardías del desarrollo, donde el gen  $\pi$  se ve silenciado y los genes adultos presentan actividad transcripcional. En el silenciamiento del gen  $\pi$ , mediante el empleo de técnicas experimentales generales, estudios previos han reportado la presencia de la metilación del ADN en la región promotora del gen  $\pi$ , sin precisar su distribución y papel en dicho silenciamiento<sup>166,182</sup>. En el presente trabajo de tesis, mediante el análisis por secuenciación del ADN modificado con bisulfito de sodio de los eritrocitos en diferentes estadios del desarrollo del pollo, se logró determinar el estado en la metilación del ADN de cada di-nucleótido CpG en una región genómica que va desde el promotor del gen  $\pi$ , el cuerpo del gen, la región intergénica, hasta el promotor del gen adulto  $\alpha^D$  (Figuras 29-34).

Mediante el análisis de ChIP de la proteína de unión al ADN metilado MeCP2, fue posible determinar que esta proteína es reclutada en el promotor del gen  $\pi$  en etapas adultas del desarrollo (Figura 28). Además, la reducción en los niveles de MeCP2 provocó la re-expresión del gen embrionario lo que permitió

vincular directamente a la metilación del ADN en el silenciamiento de este gen (Figura 28). Este resultado está de acuerdo con el papel represor de esta proteína mediante su capacidad para reclutar a los complejos de desacetilación de las histonas y represión génica<sup>39,46,176,183</sup>. No obstante, estudios recientes que abordan la distribución global de MeCP2 en el genoma de humano, han revelado que esta proteína puede localizarse en regiones del genoma que presentan actividad transcripcional<sup>183,184</sup>. Por tal motivo, resulta importante conocer el papel de MeCP2 en la regulación específica del o los genes a los cuales regula. Los hallazgos aquí descritos permiten reconocer el efecto represor de la proteína MeCP2 en el gen embrionario  $\pi$ , contribuyendo a la expresión diferencial de los genes  $\alpha$ -globina de pollo.

Dada la capacidad para reprimir la transcripción por parte de la metilación del ADN en conjunto con la unión de MeCP2, se decidió conocer los procesos involucrados en la regulación y establecimiento de dichos factores en el gen embrionario. El análisis de la distribución de la metilación del ADN evidenció una zona de transición, de un estado hipermetilado a uno desmetilado, a 500 pb río arriba del TSS del gen adulto  $\alpha^D$  (Figuras 29 a 34). Este hallazgo permitió complementar la hipótesis de que existe un silenciamiento autónomo del gen embrionario donde la metilación del ADN participa constituyendo un sub-dominio de represión que abarca la región del gen  $\pi$ . Por tal motivo, esta serie de resultados están de acuerdo con la propuesta de que es necesario un silenciamiento autónomo del gen embrionario para la expresión diferencial de los genes globina<sup>20,155,164,166</sup>. Sin embargo, se requieren más estudios que permitan

conocer los procesos que gobiernan dicho silenciamiento, como pueden ser el reclutamiento de factores nucleares capaces de atraer a las metil-transferasas de ADN y otros complejos vinculados con el establecimiento de zonas de represión transcripcional. Al respecto, diversos estudios han vinculado la acción de factores tales como GATA-1, SOX6, *BCL11A*, entre otros, en la represión de los genes globina fetales y embrionarios en el humano y el ratón, ya sea mediante su interacción con complejos de represión tales como NuRD o reclutando metilasas de ADN<sup>185-189</sup>.

De manera interesante, se ha descrito que el factor nuclear Sox6, además de tener un papel central en la diferenciación condrogénica y osteogénica<sup>190</sup>, es necesario no solo para inducir el silenciamiento de los genes globina, sino además para la consolidación de la diferenciación eritroide. Ratones deficientes en dicho gen presentan anemia embrionaria y ausencia de células eritroides maduras<sup>191,192</sup>. Estudios específicos de su papel en la regulación de los genes globina han demostrado su participación en el silenciamiento de los genes embrionarios y fetales en mamíferos<sup>188,189,193</sup>. Con estos antecedentes, al analizar la secuencia promotora del gen embrionario  $\pi$ , se pudo distinguir, de manera preliminar, que existen secuencias similares al motivo de unión al factor Sox6 (datos no mostrados). Por tal motivo, para la comprensión de los procesos y mecanismos moleculares involucrados en el silenciamiento autónomo del gen  $\pi$ , resulta primordial conocer la posible contribución del factor de transcripción Sox6.

## 2. Desmetilación activa del ADN en el dominio $\alpha$ -globina

Se ha postulado que la modificación en el ADN conocida como 5-hidroximetilcitosina o 5hmC, puede contribuir a la desmetilación activa en los promotores de los genes principalmente en células madres embrionarias de ratón (ESC; del inglés, *Embryonic Stem Cells*)<sup>61,67</sup>. Para el establecimiento de la 5hmC es necesario un estado de metilación del ADN que puede ser reconocido por la familia de proteínas TET, las cuales son las involucradas en realizar la oxidación de la 5mC en 5hmC. Este mismo grupo de proteínas tienen la capacidad de llevar a cabo oxidaciones subsecuentes que llevan a estados de 5-formil-citosina y posteriormente a 5-carboxil-citosina (Figura 7)<sup>63</sup>. Estudios recientes en ESC, han evidenciado la presencia de la 5hmC y estos derivados en promotores de genes tejido-específicos los cuales presentan desmetilación del ADN y actividad transcripcional cuando estas células son inducidas a su diferenciación<sup>194,195</sup>.

En el presente trabajo fue posible detectar la presencia de 5hmC en la región promotora del gen  $\alpha^D$  coincidente con una disminución en el estado de metilación de ADN (Figura 35). La 5hmC se presenta en etapas de la eritropoyesis donde el gen  $\alpha^D$  se encuentra reprimido transcripcionalmente (células 6C2 y HD3). Sin embargo, en las células que son inducidas a la diferenciación, lo que conlleva a la expresión de los genes adultos (dif-HD3), la 5hmC presentó un ligero incremento en comparación con las células 6C2 y HD3, lo cual sugiere su papel en la activación del gen adulto  $\alpha^D$ . Sin embargo, el hecho de detectar a la 5hmC en las etapas previas a la expresión de los genes globina (6C2 y HD3) puede ser un indicativo de que esta región se encuentra marcada para su desmetilación

conforme progresa la diferenciación eritroide y la consecuente activación de los genes  $\alpha$ -globina adultos. Por tal motivo, resulta importante el comprender si la 5hmC se presenta de manera exclusiva en la región del promotor  $\alpha^D$  (marcando para su desmetilación) o ésta se distribuye en la totalidad del dominio de metilación (incluyendo al promotor  $\pi$  y la región intergénica).

En cuanto a la regulación en el establecimiento de la 5hmC en el promotor del gen  $\alpha^D$ , existen reportes de la presencia de CTCF en ciertas regiones del genoma donde la marca de 5hmC se encuentra presente<sup>196</sup>. En las células inducidas a la diferenciación (dif-HD3), fue posible detectar la incorporación tanto de la 5hmC como del factor CTCF en la región promotora del gen  $\alpha^D$ . Esta observación implica que CTCF, y/o las proteínas con las que interactúa, podría estar involucrado en el establecimiento de la 5hmC en el promotor del gen  $\alpha^D$ . Por lo tanto, son necesarias investigaciones futuras que aborden la desmetilación del ADN en los promotores de los genes en respuesta a la diferenciación celular y el papel, si lo presenta, de CTCF en dicho proceso.

### **3. Delimitación de la metilación del ADN mediada por el factor CTCF.**

Como se ha mencionado, en el presente trabajo fue posible determinar que existe un sub-dominio de metilación que mantiene silenciado específicamente al gen embrionario en las etapas adultas del desarrollo del pollo. En el promotor del gen  $\alpha^D$ , se ha descrito la presencia de 5hmC la cual pudiera contribuir a la desmetilación de esta zona y acotando así el sub-dominio de metilación

proveniente del gen embrionario  $\pi$ . Sin embargo, se ha reportado que los promotores de genes que permanecen libres de metilación del ADN presentan la unión de factores nucleares como lo son CTCF, Sp1, VEZF1 entre otros<sup>122,183,197</sup>. Los experimentos de CHIP-seq del factor CTCF realizados en el presente trabajo, permitieron localizar su unión en la región de la transición de dicho sub-dominio de metilación (Figuras 37 y 38). CTCF se une específicamente en las etapas del desarrollo donde esta región se presenta metilación en el ADN (5 y 10dRBCs), lo que pone de manifiesto su probable participación en la delimitación de la metilación.

En coincidencia con la ocupación de CTCF en dicha región, se presenta una estructura de la cromatina enriquecida en marcas en las histonas como lo son la H3K4me2 y la H3 y H4ac, lo cual permitió sugerir que CTCF podría fomentar un estado de la cromatina abierto que contribuye tanto a la delimitación de la metilación del ADN, como a la activación transcripcional del gen  $\alpha$ D. No obstante, se han publicado evidencias de que la proteína CTCF puede presentar modificaciones post-traduccionales como lo es la ADP-ribosilación<sup>198</sup>. La adición de polímeros de ADP-ribosa en la proteína CTCF se encuentra mediada por la enzima PARP-1. Se ha descrito que la interacción de la forma ADP-ribosilada de CTCF y la PARP-1 es capaz de fomentar un estado hipometilado en el ADN debido a que la metil-transferasa de ADN (DNMT1) es reclutada en dicha interacción e inhibida por una ADP-ribosilación<sup>123,124</sup>. Además, el tratamiento de células eritroides de pollo con un inhibidor de la ADP-ribosilación llevo a la reducción en la expresión de los genes globina adultos<sup>162</sup>. Por tal motivo, es

posible sugerir nuevas investigaciones donde se aborde el posible papel de CTCF en el promotor del gen  $\alpha^D$  y la desmetilación activa del ADN por mecanismos relacionados con la ADP-ribosilación.

Para conocer de manera funcional el papel de CTCF en la región del promotor  $\alpha^D$ , se realizaron una serie de construcciones plasmídicas donde fue removido el motivo de unión de CTCF. Esto permitió demostrar que, en ausencia de CTCF, el promotor del gen  $\alpha^D$  carece de actividad transcripcional así como de una pérdida de marcas características de una cromatina abierta (Figuras 40 a 42). De manera interesante, la ausencia de CTCF provocó una ganancia de metilación en las CpGs presentes en el promotor de dichas construcciones, lo que pone de manifiesto la importancia de CTCF en la protección contra la metilación del ADN.

Esta serie de resultados representan quizás la evidencia más importante para proponer que el factor CTCF se encuentra involucrado en la protección en contra de la expansión de la metilación del ADN. Esta actividad “protectora” de CTCF se ha podido demostrar previamente con experimentos similares en el promotor del gen supresor de tumores *Retinoblastoma*<sup>57,127,128</sup>. Por tal motivo, en el presente trabajo de tesis, se ha propuesto que la unión de CTCF en el promotor del gen  $\alpha^D$  es necesaria para la delimitación de la metilación del ADN proveniente del gen embrionario  $\pi$  y el establecimiento de una cromatina permisible a la transcripción. No obstante, no podemos dejar de lado la capacidad de CTCF para organizar la distribución espacial del genoma al interior del núcleo; un aspecto que de igual manera fue abordado en el presente trabajo de tesis y que se discute a continuación.

#### **4. Participación del elemento 3'-Enh en la expresión diferencial de los genes $\alpha$ -globina.**

Contenidos en el genoma de los eucariontes, se han descrito una gran variedad de elementos de ADN que poseen la capacidad de regular a los genes como lo son los *enhancers*. Un *enhancer* puede llevar a la *trans*-activación de uno o más genes blanco ya que en el pueden ser nucleados factores de transcripción y la maquinaria necesaria para la expresión génica como lo son los complejos de acetilación de histonas, el complejo mediador y la ARN polimerasa II, entre otros<sup>15</sup>. Dentro de sus mecanismos de acción se ha descrito la capacidad de contactar físicamente a las regiones promotoras la cual puede ser regulada en las diferentes etapas del desarrollo y entre los tipos celulares<sup>32,108,117</sup>.

Estudios previos han demostrado la importancia de elemento 3'-Enh presente en el dominio  $\alpha$ -globina como un potenciador de la transcripción de los genes<sup>145,163,165,166</sup>. Sin embargo, la manera en que el elemento 3'-Enh puede *trans*-activar selectivamente a cada uno de los tres genes a través del desarrollo del pollo es desconocida.

Trabajos previos han caracterizado la organización espacial del dominio  $\alpha$ -globina en células eritroides del linaje adulto. Mediante el empleo de la técnica conocida como 3C, dicho estudios pudieron determinar que existen interacciones entre los elementos distales del dominio y los promotores de los genes<sup>171,199</sup>. Sin embargo, la estrategia experimental empleada dificulta la interpretación de las posibles interacciones entre el elemento 3'-Enh y los promotores de los genes (se

emplearon endonucleasas de restricción con pocos sitios de corte en el genoma). Esto se debe a que la distancia genómica que existe entre el 3'-Enh y los promotores de los genes  $\pi$  y  $\alpha^D$  es reducida lo que representa una limitante técnica para la realización de este tipo de experimentos. En el presente trabajo de tesis, para evitar dicha contrariedad, los experimentos de 3C fueron realizados ocupando una endonucleasa de restricción que presenta múltiples sitios de corte entre dichos elementos (DpnII), lo cual incrementa considerablemente la resolución del experimento y permite el análisis entre elementos donde se presentan pequeñas distancias genómicas de separación<sup>175</sup>. Cabe mencionar, que como parte alterna al presente proyecto de tesis, se realizó una estancia de investigación en el laboratorio del Dr. Wouter de Laat, en el Instituto Hubrecht, en Utrecht; Holanda, donde fue posible contribuir al establecimiento de este tipo de metodologías y su publicación<sup>175</sup>.

Gracias a la implementación de la técnica de 3C de alta resolución, fue posible determinar que el elemento 3'-Enh es capaz de contactar al gen embrionario  $\pi$  y al gen  $\alpha^D$  en la etapa temprana del desarrollo (5dRBCs). La presencia del promotor  $\alpha^D$  en la interacción con el promotor  $\pi$  y el 3'-Enh fue un resultado inesperado puesto que en esta etapa del desarrollo el gen  $\alpha^D$  se encuentra inactivo transcripcionalmente. Una posible explicación podría estar relacionada con el hecho de que el factor CTCF se une al promotor del gen  $\alpha^D$  en esta etapa del desarrollo (Figuras 37 y 38), por lo que la presencia de CTCF podría estar involucrada en la organización topológica óptima que permite la interacción directa entre el promotor del gen  $\pi$  y del 3'-Enh. La inactividad

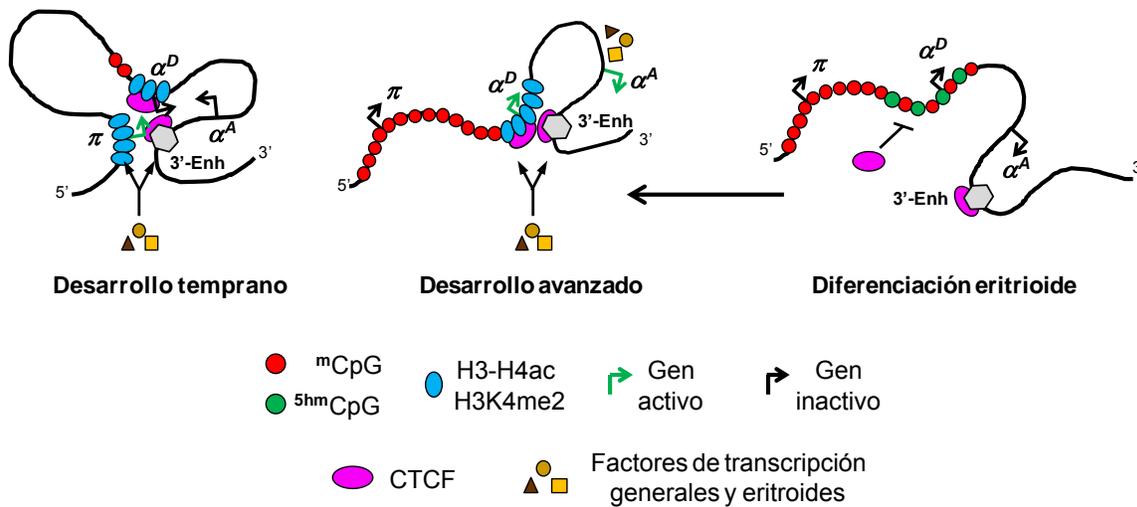
transcripcional del gen  $\alpha^D$ , no obstante su contacto con el *enhancer*, podría ser explicada por la necesidad de factores de transcripción adicionales que no se presenta en esa etapa del desarrollo<sup>166</sup>. A favor de esta propuesta, se sabe que el factor CTCF es capaz de promover la interacción regulada entre regiones del genoma distales mediante la homo-dimerización con otras moléculas de CTCF y su interacción con los complejos de cohesinas<sup>15,108,109,200</sup>. Por tal motivo, son necesarios estudios futuros donde se investigue el probable papel de CTCF y los complejos de cohesinas en la organización topológica del dominio  $\alpha$ -globina de pollo.

Para comprender el papel de CTCF no solo en el establecimiento de una conformación espacial entre los genes y el elemento *enhancer*, sino además en la delimitación de la metilación del ADN, son necesarios estudios puntuales que permitan evidenciar la función de este factor en estas regiones del genoma de pollo. Para ello, se requiere de ensayos de manipulación genética donde se pueda alterar o remover el sitio de unión para este factor de manera endógena. Este tipo de abordajes son ahora posibles gracias al desarrollo de herramientas genéticas para la “edición” del genoma como lo son el uso de las proteínas TALEN y CRISPR-Cas9<sup>201,202</sup>. Dichas proteínas permiten la manipulación del genoma de una manera altamente específica, por lo que se podrían emplear para alterar el motivo de unión a CTCF en la región del gen  $\alpha^D$  y evaluar las consecuencias en términos de la conformación tri-dimensional del dominio y la metilación del ADN; experimentos que se encuentran en desarrollo en nuestro laboratorio.

## 5. Propuesta: Modelo de regulación de los genes $\alpha$ -globina de pollo.

En el presente trabajo de tesis fueron abordados distintos aspectos que tienen que ver con la regulación genética y epigenética del grupo de genes  $\alpha$ -globina de pollo a través de la diferenciación eritroide y el desarrollo del pollo. Los resultados aquí obtenidos y en conjunto con trabajos previos, nos han permitido postular un modelo de la regulación genética y epigenética para este grupo de genes, el cual se describe a continuación (Figura 52).

En las etapas tempranas del desarrollo del pollo, donde el gen embrionario  $\pi$  es activo transcripcionalmente, existe una interacción física entre el promotor de dicho gen, el elemento 3'-Enh y el promotor del gen  $\alpha^D$  (Figura 52). En dicha interacción, el gen embrionario  $\pi$  se ve favorecido en su *trans*-activación por parte del *enhancer*. La presencia de factores de transcripción tanto en el promotor del gen  $\pi$  como en el elemento 3'-Enh podrían promover la selectividad en la *trans*-activación del gen embrionario<sup>163,165,166</sup>. Por su parte, la presencia del factor CTCF tanto en el elemento 3-Enh así como en el promotor  $\alpha^D$ , promueve esta conformación espacial facilitando la interacción entre los promotores de los genes y el elemento *enhancer*, una función ampliamente descrita para este factor de transcripción<sup>31,108,109,120,121</sup>. A pesar de la presencia de CTCF y un estado de la cromatina abierto en la región del promotor  $\alpha^D$ , la inactividad transcripcional sugiere la necesidad de factores de transcripción adicionales a los presentes en esta etapa del desarrollo del pollo<sup>163,165,166</sup>.



**Figura 52.** Modelo de la regulación genética y epigenética del grupo de genes  $\alpha$ -globina de pollo. En este modelo, se ha podido integrar el papel de la metilación ya la 5-hidroximetilación del ADN (mCpG y 5hmCpG; respectivamente), la estructura de la cromatina, el reclutamiento del factor CTCF y la conformación espacial del dominio en la regulación de la expresión diferencial de los genes a través del desarrollo del pollo y la diferenciación eritroide.

Conforme el desarrollo del organismo progresa, el recambio en la expresión de los genes  $\alpha$ -globina tiene lugar; en las células eritroides en circulación el gen  $\pi$  se encuentra silenciado y los genes adultos son activos a la transcripción. En estas células, las CpGs contenidas en el promotor del gen  $\pi$ , el cuerpo del gen y la región intergénica 3' de este, se encuentran metiladas lo cual conlleva al silenciamiento autónomo del gen embrionario (Figura 52). Dicho silenciamiento es dependiente del reclutamiento de proteínas represoras de la transcripción como lo es MeCP2. Participando en la delimitación de este subdominio de metilación en el ADN, la proteína CTCF es reclutada en la región del promotor del gen  $\alpha^D$  impidiendo que este subdominio se extienda e interfiriera con la correcta expresión de los genes adultos. Aunado con su función barrera en

esta región, el factor CTCF podría estar involucrado en el establecimiento de contactos entre el elemento 3'-Enh y el promotor  $\alpha^D$  lo cual promueve la actividad transcripcional de dicho gen. La ausencia del promotor del gen  $\pi$  en esta interacción física entre el enhancer y el gen  $\alpha^D$ , puede ser explicada por su estado hipermetilado, lo cual impide el reclutamiento de factores de transcripción capaces de interactuar con la proteína CTCF presente en el *enhancer* y en el promotor  $\alpha^D$ .

Como se ha descrito anteriormente, la expresión de los genes  $\alpha$ -globina ocurre en los eritrocitos terminalmente diferenciados (RBCs). En las etapas previas de la diferenciación eritroide adulta (donde los genes se encuentran inactivos transcripcionalmente), tanto los promotores de los genes  $\pi$  y  $\alpha^D$ , así como la región intergénica entre estos, se encuentran hiper-metilados interfiriendo con la unión de los factores de transcripción de entre ellos CTCF. No obstante, dicho factor se une al elemento 3'-Enh, un hecho que está de acuerdo con estudios previos que proponen que la estructura de la cromatina y el arreglo nucleosomal en el elemento 3'-Enh puede ser funcional en etapas de la diferenciación eritroide previas a la expresión de los genes genes<sup>145,165,166</sup>. Ante la ausencia del factor CTCF en el promotor  $\alpha^D$ , el elemento 3'-Enh no establece contactos distales con los promotores de los genes. Conforme estas células inician el programa de diferenciación terminal, la marca de 5hmC es incorporada en el promotor  $\alpha^D$  promoviendo una desmetilación activa del ADN. De esta manera el sub-dominio de metilación se ve acotado a la región del gen  $\pi$  y la región intergénica. Coincidente con la desmetilación, el factor CTCF es reclutado en el promotor  $\alpha^D$  fomentando un estado de la cromatina relajado y llevando a la

interacción física con el elemento 3'-Enh, y por consecuencia, la activación del gene adulto  $\alpha^D$ . Futuros trabajos deben ser realizados para comprender la manera el gen adulto  $\alpha^A$  puede ser regulado en su transcripción. En su regulación, no descartamos una posible relación con los procesos de regulación descritos en el presente trabajo de tesis.

## XII. Conclusiones

1. En el silenciamiento autónomo del gen embrionario  $\pi$ , la proteína de unión a ADN metilado MeCP2 es reclutada en el promotor del gen y participa directamente en el silenciamiento en las etapas adultas del desarrollo del pollo y en la diferenciación eritroide adulta.
2. En el silenciamiento autónomo del gen embrionario  $\pi$ , un sub-dominio de metilación es establecido en su región promotora y se extiende hasta el promotor del gen  $\alpha^D$ .
3. El factor de transcripción CTCF es reclutado en el promotor del gen  $\alpha^D$  en las células eritroides terminalmente diferenciadas. La incorporación del factor CTCF en esta región es necesaria para la actividad transcripcional del promotor del gen  $\alpha^D$ .
4. La presencia del factor CTCF en la región promotora del gen  $\alpha^D$  podría estar involucrada con la delimitación del sub-dominio de metilación presente en el gen embrionario  $\pi$ . Además, CTCF podría estar participando en el establecimiento de una cromatina abierta con la capacidad de contener contra la expansión de la metilación del ADN.

5. El elemento 3'-Enh establece contactos físicos con los promotores de los genes  $\pi$  y  $\alpha^D$  a través de la formación de loops de cromatina. Dicha capacidad podría estar modulada por la presencia del factor CTCF en el elemento 3'-Enh.

### XIII. Perspectivas

Derivados de los resultados del presente estudio, es posible sugerir diversos aspectos que deben ser abordados para profundizar en el papel de la regulación epigenética en la expresión diferencial de los genes  $\alpha$ -globina de pollo.

En lo que se refiera a la metilación del ADN que se presenta en la región del gen embrionario  $\pi$  en las etapas adultas del desarrollo, son necesarios futuros estudios que se enfoquen en determinar los mecanismos y/o procesos involucrados en su establecimiento. En el presente trabajo, hemos descrito que el sub-dominio de metilación abarca la región del promotor y el cuerpo del gen  $\pi$ , extendiéndose hasta la región intergénica 3' con el gen  $\alpha^D$ . Sin embargo, se desconoce la distribución del sub-dominio de metilación de ADN en el costado 5' río arriba del promotor del gen  $\pi$ . El conocer la distribución de la metilación en dicha región podría ayudar a identificar otros elementos regulación implicados en el silenciamiento de los genes globina tal como se ha descrito para el humano y el ratón<sup>155,157,185</sup>.

En lo que se refiere a la represión de los genes globina embrionarios y fetales en mamíferos, además de la participación de la metilación del ADN, se ha descrito la acción de diversos factores de transcripción asociados a represión transcripcional<sup>185-189</sup>. Por ejemplo, el factor Sox6 se ha descrito como un represor importante para el silenciamiento de los genes globina en el humano y el ratón, además de un regulador clave en la diferenciación eritroide<sup>188,189,191,193</sup>. Por tal motivo, es necesario conocer la posible presencia y acción de este y otros

factores en los promotores de los genes  $\alpha$ -globina del pollo. Al respecto, en nuestro laboratorio hemos encontrado que el promotor del gen  $\pi$  posee la secuencia ADN para la unión del factor Sox6 (datos no mostrados). Estudios preliminares de sobre-expresión mediante ensayos de transfecciones del ADNc del factor Sox6, han demostrado su capacidad de reprimir la transcripción del gen  $\pi$ . Por consiguiente, es necesario evidenciar su unión *in vivo* en el promotor del gen  $\pi$  y su relación (sí presenta) con la metilación del ADN y la estructura de la cromatina.

En la actualidad, vinculados a la regulación transcripcional y la estructura de la cromatina, se ha descrito la intervención de los ARN no codificantes en dichos procesos. En particular, los ARN no codificantes largos han tomado gran importancia debido a su papel en el establecimiento de estados de la cromatina mediante su interacción con complejos de modificación de las histonas<sup>203–205</sup>. En el laboratorio, como una colaboración al proyecto doctoral de Cristian Arriaga Canon, se ha podido describir un transcrito no codificante largo el cual abarca la región codificante de los genes globina y el elemento 3'-Enh<sup>206</sup>. El transcrito fue denominado lncRNA- $\alpha$ GT (del inglés,  *$\alpha$ -globin transcript long non-coding RNA*). Experimentos de abatimiento en su transcripción (mediante el uso de ARNi) mostraron que el lncRNA- $\alpha$ GT es necesario para la óptima expresión de los genes adultos  $\alpha^D$  y  $\alpha^A$  en el linaje eritroide definitivo. Por tal motivo, resulta interesante el comprender la probable participación del lncRNA- $\alpha$ GT en la delimitación del sub-dominio de metilación presente en el gen  $\pi$ . Además, recientemente en nuestro laboratorio se ha podido determinar que el factor CTCF

es capaz de interactuar con moléculas de ARN en el núcleo<sup>207</sup>. De tal manera que son necesarios estudios donde se aborde el papel del factor CTCF presente en el promotor del gen  $\alpha^D$  y en el elemento 3'-Enh y su relación con el lncRNA- $\alpha$ GT. Experimentos de análisis de la metilación y de 3C deben realizarse en el contexto del abatimiento del lncRNA- $\alpha$ GT como una manera de comprender el papel de este transcrito en dichos procesos de regulación epigenética del dominio.

Uno de los aspectos claves del modelos de regulación aquí propuesto (Figura 52) se centra en la unión del factor CTCF tanto en el promotor del gen  $\alpha^D$  como en el elemento 3'-Enh, por lo tanto los experimentos de remoción del sitio de unión a CTCF mediante las herramientas experimentales de “edición” del genoma, CRISPR-Cas9, que ya se encuentran en curso en nuestro laboratorio, permitirán conocer el papel de dichos elementos en la expresión diferencial de los genes  $\alpha$ -globina.

En lo que se refiere a la organización espacial del dominio, adicionalmente a conocer el papel de CTCF en la formación de los contactos físicos entre los elementos de regulación, son necesarios experimentos que permitan conocer la presencia y distribución de otros “jugadores” en la conformación de la fibra de cromatina. De entre ellos, la unión de las proteínas cohesinas y su relación con la unión del factor CTCF debe ser abordada.

## XIV. Referencias

1. Benjamin, Lewin. *GENES IX*. (2008). Ed. MC GRAW HILL.
2. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
3. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–80 (2001).
4. Lenhard, B., Sandelin, A. & Carninci, P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat. Rev. Genet.* **13**, 233–45 (2012).
5. M. Escamilla-del-Arenal, F. R.-T. Participación de la estructura de la cromatina en la regulación de la expresión génica. *Mensaje Bioquímico* **28**, 173–201 (2004).
6. Maston, G. A., Evans, S. K. & Green, M. R. Transcriptional regulatory elements in the human genome. *Annu. Rev. Genomics Hum. Genet.* **7**, 29–59 (2006).
7. Saxonov, S., Berg, P. & Brutlag, D. L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 1412–7 (2006).
8. Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–22 (2011).
9. Spitz, F. & Furlong, E. E. M. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–26 (2012).
10. Rincón-Arano, H. & Recillas-Targa, F. Sustained heterologous transgene expression in mammalian and avian cell lines. *Methods Mol. Biol.* **267**, 435–50 (2004).
11. Bulger, M. & Groudine, M. Looping versus linking: toward a model for long-distance gene activation. *Genes Dev.* **13**, 2465–77 (1999).
12. Hatzis, P. & Talianidis, I. Dynamics of enhancer-promoter communication during differentiation-induced gene activation. *Mol. Cell* **10**, 1467–77 (2002).
13. Kadauke, S. & Blobel, G. A. Chromatin loops in gene regulation. *Biochim. Biophys. Acta* **1789**, 17–25 (2009).
14. Dean, A. On a chromosome far, far away: LCRs and gene expression. *Trends Genet.* **22**, 38–45 (2006).
15. Ong, C.-T. & Corces, V. G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* **12**, 283–93 (2011).
16. Maston, G. A., Landt, S. G., Snyder, M. & Green, M. R. Characterization of enhancer function from genome-wide analyses. *Annu. Rev. Genomics Hum. Genet.* **13**, 29–57 (2012).
17. Melo, C. A. *et al.* eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol. Cell* **49**, 524–35 (2013).

18. Redmond, A. M. & Carroll, J. S. Enhancer-derived RNAs: “spicing up” transcription programs. *EMBO J.* **32**, 2096–8 (2013).
19. Li, Q., Peterson, K. R., Fang, X. & Stamatoyannopoulos, G. Locus control regions. *Blood* **100**, 3077–86 (2002).
20. Harju, S., McQueen, K. J. & Peterson, K. R. Chromatin structure and control of beta-like globin gene switching. *Exp. Biol. Med. (Maywood)*. **227**, 683–700 (2002).
21. Forsberg, E. C., Downs, K. M. & Bresnick, E. H. Direct interaction of NF-E2 with hypersensitive site 2 of the beta-globin locus control region in living cells. *Blood* **96**, 334–9 (2000).
22. Recillas-Targa, F., Valadez-Graham, V. & Farrell, C. M. Prospects and implications of using chromatin insulators in gene therapy and transgenesis. *Bioessays* **26**, 796–807 (2004).
23. Capelson, M. & Corces, V. G. Boundary elements and nuclear organization. *Biol. Cell* **96**, 617–29 (2004).
24. Schneider, R. & Grosschedl, R. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes Dev.* **21**, 3027–43 (2007).
25. Chung, J. H., Whiteley, M. & Felsenfeld, G. A 5' element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell* **74**, 505–14 (1993).
26. Burgess-Beusse, B. *et al.* The insulation of genes from external enhancers and silencing chromatin. *Proc. Natl. Acad. Sci. U. S. A.* **99 Suppl 4**, 16433–7 (2002).
27. Bell, A. C., West, A. G. & Felsenfeld, G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98**, 387–96 (1999).
28. Recillas-Targa, F., Bell, A. C. & Felsenfeld, G. Positional enhancer-blocking activity of the chicken beta-globin insulator in transiently transfected cells. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 14354–9 (1999).
29. Wallace, J. A. & Felsenfeld, G. We gather together: insulators and genome organization. *Curr. Opin. Genet. Dev.* **17**, 400–7 (2007).
30. Filippova, G. N. Genetics and epigenetics of the multifunctional protein CTCF. *Curr. Top. Dev. Biol.* **80**, 337–60 (2008).
31. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* **20**, 2349–54 (2006).
32. Splinter, E. & de Laat, W. The complex transcription regulatory landscape of our genome: control in three dimensions. *EMBO J.* **30**, 4345–55 (2011).
33. Allis, C. D., Jenuwein, T., Reinberg, D. & Caparros, M.-L. *Epigenetics*. (Cold Spring Harbor Laboratory Press). at <<http://www.amazon.com/Epigenetics-Danny-Reinberg/dp/0879698756>>

34. Cremer, T., Küpper, K., Dietzel, S. & Fakan, S. Higher order chromatin architecture in the cell nucleus: on the way from structure to function. *Biol. Cell* **96**, 555–67 (2004).
35. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448–53 (2003).
36. Bell, A. C., West, A. G. & Felsenfeld, G. Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* **291**, 447–50 (2001).
37. Beisel, C. & Paro, R. Silencing chromatin: comparing modes and mechanisms. *Nat. Rev. Genet.* **12**, 123–35 (2011).
38. Chen, Z. & Riggs, A. D. DNA methylation and demethylation in mammals. *J. Biol. Chem.* **286**, 18347–53 (2011).
39. Klose, R. J. & Bird, A. P. Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **31**, 89–97 (2006).
40. Lee, T.-F., Zhai, J. & Meyers, B. C. Conservation and divergence in eukaryotic DNA methylation. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 9027–8 (2010).
41. Singal, R. & Ginder, G. D. DNA methylation. *Blood* **93**, 4059–70 (1999).
42. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
43. Escamilla-Del-Arenal, M., da Rocha, S. T. & Heard, E. Evolutionary diversity and developmental regulation of X-chromosome inactivation. *Hum. Genet.* **130**, 307–27 (2011).
44. Buck-Koehntop, B. A. & Defossez, P.-A. On how mammalian transcription factors recognize methylated DNA. *Epigenetics* **8**, 131–7 (2013).
45. Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. & Bird, A. P. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **58**, 499–507 (1989).
46. Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **18**, 6538–47 (1998).
47. Ng, H. H. & Bird, A. DNA methylation and chromatin modification. *Curr. Opin. Genet. Dev.* **9**, 158–63 (1999).
48. Martín Caballero, I., Hansen, J., Leaford, D., Pollard, S. & Hendrich, B. D. The methyl-CpG binding proteins Mecp2, Mbd2 and Kaiso are dispensable for mouse embryogenesis, but play a redundant function in neural differentiation. *PLoS One* **4**, e4315 (2009).
49. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457–66 (2007).
50. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–22 (2009).
51. Ball, M. P. *et al.* Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* **27**, 361–8 (2009).

52. Schübeler, D. Molecular biology. Epigenetic islands in a genetic ocean. *Science* **338**, 756–7 (2012).
53. Baylin, S. B. *et al.* Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum. Mol. Genet.* **10**, 687–92 (2001).
54. Brown, R. & Strathdee, G. Epigenomics and epigenetic therapy of cancer. *Trends Mol. Med.* **8**, S43–8 (2002).
55. Toraño, E. G., Petrus, S., Fernandez, A. F. & Fraga, M. F. Global DNA hypomethylation in cancer: review of validated methods and clinical significance. *Clin. Chem. Lab. Med.* **50**, 1733–42 (2012).
56. You, J. S. & Jones, P. A. Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell* **22**, 9–20 (2012).
57. Dávalos-Salas, M. *et al.* Gain of DNA methylation is enhanced in the absence of CTCF at the human retinoblastoma gene promoter. *BMC Cancer* **11**, 232 (2011).
58. Feng, S., Jacobsen, S. E. & Reik, W. Epigenetic reprogramming in plant and animal development. *Science* **330**, 622–7 (2010).
59. Ooi, S. K. T. & Bestor, T. H. The colorful history of active DNA demethylation. *Cell* **133**, 1145–8 (2008).
60. Lee, J. *et al.* Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development* **129**, 1807–17 (2002).
61. Pastor, W. A. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–7 (2011).
62. Guo, J. U., Su, Y., Zhong, C., Ming, G. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–34 (2011).
63. Williams, K., Christensen, J. & Helin, K. DNA methylation: TET proteins-guardians of CpG islands? *EMBO Rep.* **13**, 28–35 (2012).
64. Branco, M. R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* **13**, 7–13 (2012).
65. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–5 (2009).
66. Frauer, C. *et al.* Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS One* **6**, e21306 (2011).
67. Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398–402 (2011).
68. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–8 (2011).

69. He, Y.-F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–7 (2011).
70. Murr, R. Interplay between different epigenetic modifications and mechanisms. *Adv. Genet.* **70**, 101–41 (2010).
71. Furey, T. S. ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat. Rev. Genet.* **13**, 840–52 (2012).
72. Carrozza, M. J., Utley, R. T., Workman, J. L. & Côté, J. The diverse functions of histone acetyltransferase complexes. *Trends Genet.* **19**, 321–9 (2003).
73. Yang, X.-J. The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. *Nucleic Acids Res.* **32**, 959–76 (2004).
74. De la Cruz, X., Lois, S., Sánchez-Molina, S. & Martínez-Balbás, M. A. Do protein motifs read the histone code? *Bioessays* **27**, 164–75 (2005).
75. Kanno, T. *et al.* Selective recognition of acetylated histones by bromodomain proteins visualized in living cells. *Mol. Cell* **13**, 33–43 (2004).
76. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–37 (2007).
77. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903 (2008).
78. Wang, Z. *et al.* Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138**, 1019–31 (2009).
79. Cui, K. *et al.* Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* **4**, 80–93 (2009).
80. Wang, Z., Schones, D. E. & Zhao, K. Characterization of human epigenomes. *Curr. Opin. Genet. Dev.* **19**, 127–34 (2009).
81. Natoli, G. & Andrau, J.-C. Noncoding transcription at enhancers: general principles and functional models. *Annu. Rev. Genet.* **46**, 1–19 (2012).
82. De Ruijter, A. J. M., van Gennip, A. H., Caron, H. N., Kemp, S. & van Kuilenburg, A. B. P. Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem. J.* **370**, 737–49 (2003).
83. Bogdanović, O. & Veenstra, G. J. C. DNA methylation and methyl-CpG binding proteins: developmental requirements and function. *Chromosoma* **118**, 549–65 (2009).
84. Reynolds, N., O'Shaughnessy, A. & Hendrich, B. Transcriptional repressors: multifaceted regulators of gene expression. *Development* **140**, 505–12 (2013).
85. Black, J. C., Van Rechem, C. & Whetstone, J. R. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol. Cell* **48**, 491–507 (2012).

86. Greer, E. L. & Shi, Y. Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* **13**, 343–57 (2012).
87. Shi, Y. *et al.* Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941–53 (2004).
88. Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811–6 (2006).
89. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279–83 (2011).
90. Ong, C.-T. & Corces, V. G. Enhancers: emerging roles in cell fate specification. *EMBO Rep.* **13**, 423–30 (2012).
91. Loh, Y.-H., Zhang, W., Chen, X., George, J. & Ng, H.-H. Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev.* **21**, 2545–57 (2007).
92. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–43 (2002).
93. Czermin, B. *et al.* Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185–96 (2002).
94. Hansen, K. H. *et al.* A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* **10**, 1291–300 (2008).
95. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349–53 (2006).
96. Lee, T. I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–13 (2006).
97. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–26 (2006).
98. Vastenhouw, N. L. & Schier, A. F. Bivalent histone modifications in early embryogenesis. *Curr. Opin. Cell Biol.* **24**, 374–86 (2012).
99. Lobanenko, V. V. *et al.* A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene* **5**, 1743–53 (1990).
100. Vostrov, A. A. & Quitschke, W. W. The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation. *J. Biol. Chem.* **272**, 33353–9 (1997).
101. Ohlsson, R., Renkawitz, R. & Lobanenko, V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* **17**, 520–7 (2001).

102. Heath, H. *et al.* CTCF regulates cell cycle progression of alphabeta T cells in the thymus. *EMBO J.* **27**, 2839–50 (2008).
103. Torrano, V. *et al.* CTCF regulates growth and erythroid differentiation of human myeloid leukemia cells. *J. Biol. Chem.* **280**, 28152–61 (2005).
104. Martin, D. *et al.* Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. *Nat. Struct. Mol. Biol.* **18**, 708–14 (2011).
105. Kim, T. H. *et al.* Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**, 1231–45 (2007).
106. Jothi, R., Cuddapah, S., Barski, A., Cui, K. & Zhao, K. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res.* **36**, 5221–31 (2008).
107. Renda, M. *et al.* Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *J. Biol. Chem.* **282**, 33336–45 (2007).
108. Holwerda, S. J. B. & de Laat, W. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368**, 20120369 (2013).
109. Merckenschlager, M. & Odom, D. T. CTCF and cohesin: linking gene regulatory elements with their targets. *Cell* **152**, 1285–97 (2013).
110. Cuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19**, 24–32 (2009).
111. West, A. G., Huang, S., Gaszner, M., Litt, M. D. & Felsenfeld, G. Recruitment of histone modifications by USF proteins at a vertebrate barrier element. *Mol. Cell* **16**, 453–63 (2004).
112. Saitoh, N. *et al.* Structural and functional conservation at the boundaries of the chicken beta-globin domain. *EMBO J.* **19**, 2315–22 (2000).
113. Recillas-Targa, F. *et al.* Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 6883–8 (2002).
114. Farrell, C. M., West, A. G. & Felsenfeld, G. Conserved CTCF insulator elements flank the mouse and human beta-globin loci. *Mol. Cell. Biol.* **22**, 3820–31 (2002).
115. Bulger, M. *et al.* A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse beta-globin locus. *Mol. Cell. Biol.* **23**, 5234–44 (2003).
116. Bender, M. A. *et al.* Flanking HS-62.5 and 3' HS1, and regions upstream of the LCR, are not required for beta-globin transcription. *Blood* **108**, 1395–401 (2006).
117. Palstra, R.-J. *et al.* The beta-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* **35**, 190–4 (2003).

118. Recillas-Targa, F. DNA methylation, chromatin boundaries, and mechanisms of genomic imprinting. *Arch. Med. Res.* **33**, 428–38
119. Ferguson-Smith, A. C., Sasaki, H., Cattanach, B. M. & Surani, M. A. Parental-origin-specific epigenetic modification of the mouse H19 gene. *Nature* **362**, 751–5 (1993).
120. Kurukuti, S. *et al.* CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10684–9 (2006).
121. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–211 (2009).
122. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–5 (2011).
123. Guastafierro, T. *et al.* CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery. *J. Biol. Chem.* **283**, 21873–80 (2008).
124. Zampieri, M. *et al.* ADP-ribose polymers localized on Ctfp-Parp1-Dnmt1 complex prevent methylation of Ctfp target sites. *Biochem. J.* **441**, 645–52 (2012).
125. Wang, H. *et al.* Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22**, 1680–8 (2012).
126. Recillas-Targa, F., De La Rosa-Velázquez, I. A., Soto-Reyes, E. & Benítez-Bribiesca, L. Epigenetic boundaries of tumour suppressor gene promoters: the CTCF connection and its role in carcinogenesis. *J. Cell. Mol. Med.* **10**, 554–68
127. Recillas-Targa, F., de la Rosa-Velázquez, I. A. & Soto-Reyes, E. Insulation of tumor suppressor genes by the nuclear factor CTCF. *Biochem. Cell Biol.* **89**, 479–88 (2011).
128. De La Rosa-Velázquez, I. A., Rincón-Arano, H., Benítez-Bribiesca, L. & Recillas-Targa, F. Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Res.* **67**, 2577–85 (2007).
129. Butcher, D. T. & Rodenhiser, D. I. Epigenetic inactivation of BRCA1 is associated with aberrant expression of CTCF and DNA methyltransferase (DNMT3B) in some sporadic breast tumours. *Eur. J. Cancer* **43**, 210–9 (2007).
130. Francastel, C., Schübeler, D., Martin, D. I. & Groudine, M. Nuclear compartmentalization and gene activity. *Nat. Rev. Mol. Cell Biol.* **1**, 137–43 (2000).
131. Cremer, M. *et al.* Inheritance of gene density-related higher order chromatin arrangements in normal and tumor cell nuclei. *J. Cell Biol.* **162**, 809–20 (2003).
132. Bolzer, A. *et al.* Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* **3**, e157 (2005).
133. Rajapakse, I. & Groudine, M. On emerging nuclear order. *J. Cell Biol.* **192**, 711–21 (2011).
134. Cremer, T. & Cremer, M. Chromosome territories. *Cold Spring Harb. Perspect. Biol.* **2**, a003889 (2010).

135. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–11 (2002).
136. Simonis, M., Kooren, J. & de Laat, W. An evaluation of 3C-based methods to capture DNA interactions. *Nat. Methods* **4**, 895–901 (2007).
137. De Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* **26**, 11–24 (2012).
138. De Laat, W. & Dekker, J. 3C-based technologies to study the shape of the genome. *Methods* **58**, 189–91 (2012).
139. Van Steensel, B. & Dekker, J. Genomics tools for unraveling chromosome architecture. *Nat. Biotechnol.* **28**, 1089–1095 (2010).
140. Ghirlando, R. *et al.* Chromatin domains, insulators, and the regulation of gene expression. *Biochim. Biophys. Acta* **1819**, 644–51 (2012).
141. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–80 (2012).
142. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–93 (2009).
143. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–51 (2008).
144. Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38**, 603–13 (2010).
145. Guerrero, G. *et al.* Globin genes transcriptional switching, chromatin structure and linked lessons to epigenetics in cancer: a comparative overview. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **147**, 750–60 (2007).
146. Storz, J. F., Opazo, J. C. & Hoffmann, F. G. Phylogenetic diversification of the globin gene superfamily in chordates. *IUBMB Life* **63**, 313–22 (2011).
147. Hughes, J. R. *et al.* Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 9830–5 (2005).
148. Razin, S. V, Farrell, C. M. & Recillas-Targa, F. Genomic domains and regulatory elements operating at the domain level. *Int. Rev. Cytol.* **226**, 63–125 (2003).
149. Orkin, S. H. Diversification of haematopoietic stem cells to specific lineages. *Nat. Rev. Genet.* **1**, 57–64 (2000).
150. Keller, G., Lacaud, G. & Robertson, S. Development of the hematopoietic system in the mouse. *Exp. Hematol.* **27**, 777–87 (1999).
151. Bresnick, E. H., Lee, H.-Y., Fujiwara, T., Johnson, K. D. & Keles, S. GATA switches as developmental drivers. *J. Biol. Chem.* **285**, 31087–93 (2010).

152. Kaneko, H., Shimizu, R. & Yamamoto, M. GATA factor switching during erythroid differentiation. *Curr. Opin. Hematol.* **17**, 163–8 (2010).
153. Deng, W. *et al.* Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**, 1233–44 (2012).
154. Song, S.-H., Hou, C. & Dean, A. A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. *Mol. Cell* **28**, 810–22 (2007).
155. Harju-Baker, S., Costa, F. C., Fedosyuk, H., Neades, R. & Peterson, K. R. Silencing of Agamma-globin gene expression during adult definitive erythropoiesis mediated by GATA-1-FOG-1-Mi2 complex binding at the -566 GATA site. *Mol. Cell. Biol.* **28**, 3101–13 (2008).
156. Rupon, J. W., Wang, S. Z., Gnanapragasam, M., Labropoulos, S. & Ginder, G. D. MBD2 contributes to developmental silencing of the human  $\epsilon$ -globin gene. *Blood Cells. Mol. Dis.* **46**, 212–9 (2011).
157. Sargent, T. G., Buller, A. M., Teachey, D. T., McCanna, K. S. & Lloyd, J. A. The gamma-globin promoter has a major role in competitive inhibition of beta-globin gene expression in early erythroid development. *DNA Cell Biol.* **18**, 293–303 (1999).
158. Recillas-Targa, F. & Razin, S. V. Chromatin domains and regulation of gene expression: familiar and enigmatic clusters of chicken globin genes. *Crit. Rev. Eukaryot. Gene Expr.* **11**, 227–42 (2001).
159. Knezetic, J. A. & Felsenfeld, G. Mechanism of developmental regulation of alpha pi, the chicken embryonic alpha-globin gene. *Mol. Cell. Biol.* **13**, 4632–9 (1993).
160. Razin, S. V, Shen, K., loudinkova, E. & Scherrer, K. Functional analysis of DNA sequences located within a cluster of DNase I hypersensitive sites colocalizing with a MAR element at the upstream border of the chicken alpha-globin gene domain. *J. Cell. Biochem.* **74**, 38–49 (1999).
161. Valadez-Graham, V., Razin, S. V & Recillas-Targa, F. CTCF-dependent enhancer blockers at the upstream region of the chicken alpha-globin gene domain. *Nucleic Acids Res.* **32**, 1354–62 (2004).
162. Furlan-Magaril, M. *et al.* An insulator embedded in the chicken  $\alpha$ -globin locus regulates chromatin domain configuration and differential gene expression. *Nucleic Acids Res.* **39**, 89–103 (2011).
163. Escamilla-Del-Arenal, M. & Recillas-Targa, F. GATA-1 modulates the chromatin structure and activity of the chicken alpha-globin 3' enhancer. *Mol. Cell. Biol.* **28**, 575–86 (2008).
164. Targa, F. R., de Moura Gallo, C. V, Huesca, M., Scherrer, K. & Marcaud, L. Silencer and enhancer elements located at the 3'-side of the chicken and duck alpha-globin-encoding gene domains. *Gene* **129**, 229–37 (1993).
165. Rincón-Arano, H., Valadez-Graham, V., Guerrero, G., Escamilla-Del-Arenal, M. & Recillas-Targa, F. YY1 and GATA-1 interaction modulate the chicken 3'-side alpha-globin enhancer activity. *J. Mol. Biol.* **349**, 961–75 (2005).

166. Rincón-Arano, H., Guerrero, G., Valdes-Quezada, C. & Recillas-Targa, F. Chicken alpha-globin switching depends on autonomous silencing of the embryonic pi globin gene by epigenetics mechanisms. *J. Cell. Biochem.* **108**, 675–87 (2009).
167. Bruns, G. A. & Ingram, V. M. The erythroid cells and haemoglobins of the chick embryo. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **266**, 225–305 (1973).
168. Hayman, M. J. *et al.* Cell transformation by avian defective leukaemia viruses. *Proc. R. Soc. Lond. B. Biol. Sci.* **210**, 397–409 (1980).
169. Graf, T. & Beug, H. Role of the v-erbA and v-erbB oncogenes of avian erythroblastosis virus in erythroid cell transformation. *Cell* **34**, 7–9 (1983).
170. Beug, H., Doederlein, G., Freudenstein, C. & Graf, T. 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 (1982).
171. Gavrillov, A. A. & Razin, S. V. Spatial configuration of the chicken alpha-globin gene domain: immature and active chromatin hubs. *Nucleic Acids Res.* **36**, 4629–40 (2008).
172. Weitzel, J. M., Buhmester, H. & Strätling, W. H. Chicken MAR-binding protein ARBP is homologous to rat methyl-CpG-binding protein MeCP2. *Mol. Cell. Biol.* **17**, 5656–66 (1997).
173. Rincón-Arano, H., Furlan-Magaril, M. & Recillas-Targa, F. Protection against telomeric position effects by the chicken cHS4 beta-globin insulator. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 14044–9 (2007).
174. Hagège, H. *et al.* Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* **2**, 1722–33 (2007).
175. Van de Werken, H. J. G. *et al.* Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* **9**, 969–72 (2012).
176. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–9 (1998).
177. Clark, S. J., Harrison, J., Paul, C. L. & Frommer, M. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* **22**, 2990–7 (1994).
178. Carlsten, J. O. P., Zhu, X. & Gustafsson, C. M. The multitasking Mediator complex. *Trends Biochem. Sci.* **38**, 531–7 (2013).
179. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–5 (2010).
180. Splinter, E., Grosveld, F. & de Laat, W. 3C technology: analyzing the spatial organization of genomic loci in vivo. *Methods Enzymol.* **375**, 493–507 (2004).
181. Hawkins, R. D., Hon, G. C. & Ren, B. Next-generation genomics: an integrative approach. *Nat. Rev. Genet.* **11**, 476–86 (2010).
182. Singal, R., vanWert, J. M. & Ferdinand, L. Methylation of alpha-type embryonic globin gene alpha pi represses transcription in primary erythroid cells. *Blood* **100**, 4217–22 (2002).

183. Baubec, T., Ivánek, R., Lienert, F. & Schübeler, D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153**, 480–92 (2013).
184. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**, 1417–30 (2012).
185. Costa, F. C., Fedosyuk, H., Chazelle, A. M., Neades, R. Y. & Peterson, K. R. Mi2 $\beta$  is required for  $\gamma$ -globin gene silencing: temporal assembly of a GATA-1-FOG-1-Mi2 repressor complex in  $\beta$ -YAC transgenic mice. *PLoS Genet.* **8**, e1003155 (2012).
186. Xu, Z. *et al.* The role of WDR5 in silencing human fetal globin gene expression. *Haematologica* **97**, 1632–40 (2012).
187. Vassen, L. *et al.* Growth factor independence 1b (gfi1b) is important for the maturation of erythroid cells and the regulation of embryonic globin expression. *PLoS One* **9**, e96636 (2014).
188. Yi, Z. *et al.* Sox6 directly silences epsilon globin expression in definitive erythropoiesis. *PLoS Genet.* **2**, e14 (2006).
189. Xu, J. *et al.* Transcriptional silencing of  $\gamma$ -globin by BCL11A involves long-range interactions and cooperation with SOX6. *Genes Dev.* **24**, 783–98 (2010).
190. Lefebvre, V. The SoxD transcription factors--Sox5, Sox6, and Sox13--are key cell fate modulators. *Int. J. Biochem. Cell Biol.* **42**, 429–32 (2010).
191. Dumitriu, B. *et al.* Sox6 cell-autonomously stimulates erythroid cell survival, proliferation, and terminal maturation and is thereby an important enhancer of definitive erythropoiesis during mouse development. *Blood* **108**, 1198–207 (2006).
192. Cohen-Barak, O. *et al.* Stem cell transplantation demonstrates that Sox6 represses epsilon  $\gamma$  globin expression in definitive erythropoiesis of adult mice. *Exp. Hematol.* **35**, 358–67 (2007).
193. Sankaran, V. G., Menne, J. & Heller, R. Heterozygous disruption of human SOX6 is insufficient to impair erythropoiesis or silencing of fetal hemoglobin. *Blood* **117**, 4396–7 (2011).
194. Dawlaty, M. M. *et al.* Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev. Cell* **29**, 102–11 (2014).
195. Wheldon, L. M. *et al.* Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. *Cell Rep.* **7**, 1353–61 (2014).
196. Sun, Z. *et al.* High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. *Cell Rep.* **3**, 567–76 (2013).
197. Dickson, J. *et al.* VEZF1 elements mediate protection from DNA methylation. *PLoS Genet.* **6**, e1000804 (2010).

198. Caiafa, P. & Zlatanova, J. CCCTC-binding factor meets poly(ADP-ribose) polymerase-1. *J. Cell. Physiol.* **219**, 265–70 (2009).
199. Loudinkova, E. S. *et al.* The inactivation of the  $\pi$  gene in chicken erythroblasts of adult lineage is not mediated by packaging of the embryonic part of the  $\alpha$ -globin gene domain into a repressive heterochromatin-like structure. *Epigenetics* **6**, 1481–8 (2011).
200. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 996–1001 (2014).
201. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat. Methods* **10**, 957–63 (2013).
202. Gaj, T., Gersbach, C. A. & Barbas, C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**, 397–405 (2013).
203. Bergmann, J. H. & Spector, D. L. Long non-coding RNAs: modulators of nuclear structure and function. *Curr. Opin. Cell Biol.* **26**, 10–8 (2014).
204. Fatica, A. & Bozzoni, I. Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* **15**, 7–21 (2014).
205. Ulitsky, I. & Bartel, D. P. lincRNAs: genomics, evolution, and mechanisms. *Cell* **154**, 26–46 (2013).
206. Arriaga-Canon, C. *et al.* A long non-coding RNA promotes full activation of adult gene expression in the chicken  $\alpha$ -globin domain. *Epigenetics* **9**, 173–81 (2014).
207. Saldaña-Meyer, R. *et al.* CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.* **28**, 723–34 (2014).

## XV. Publicaciones relacionadas al proyecto principal de tesis

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

- Publicación principal derivada del presente trabajo de tesis.

2. Arriaga-Canon C, Fonseca-Guzmán Y, **Valdes-Quezada C**, Arzate-Mejía R, Guerrero G, Recillas-Targa F. 2014. A long non-coding RNA promotes full activation of adult gene expression in the chicken  $\alpha$ -globin domain. *Epigenetics*. 9(1):173-81

- En esta publicación se contribuyó a la caracterización de la estructura de la cromatina en el promotor del gen  $\alpha^D$  y su relación con un transcrito largo no codificante involucrado en la expresión de dicho gen.

3. van de Werken HJ, Landan G, Holwerda SJ, Hoichman M, Klous P, Chachik R, Splinter E, **Valdes-Quezada C**, Oz Y, Bouwman BA, Verstegen MJ, de Wit E, Tanay A, de Laat W. 2012. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods*. 9(10):969-72.

-Se realizaron experimentos de 3C en círculo acoplados a secuenciación genómica (4C-seq) donde se logró contribuir con replicas experimentales que sustentan los resultados principales de esta publicación. La implementación del 3C de alta resolución descrito en esta publicación representó un aspecto experimental clave para los resultados del presente trabajo de tesis.

4. 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, Furlan-Magaril M, Alcina A, Cañón S, Fedetz M, Blasco MA, Pereira PS, Ovcharenko I, Recillas-Targa F, Montoliu L, Manzanares M, Guigó R, Serrano M, Casares F, Gómez-Skarmeta JL. 2011. Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. *Nat Struct Mol Biol*. 18(6):708-14.

-Los experimentos de ChIP-seq del factor CTCF realizados en el presente trabajo de tesis fueron utilizados en esta publicación para los estudios comparativos la unión genómica de este factor entre los vertebrados.

5. 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(3):675-87.

-En esta publicación se incluyen los resultados obtenidos del estudio de la unión de MeCP2 en el promotor  $\pi$  y el efecto de su eliminación mediante ARNi.

6. Guerrero G, Delgado-Olguin P, Escamilla-Del-Arenal M, Furlan-Magaril M, Rebollar E, De La Rosa-Velazquez IA, Soto-Reyes E, Rincon-Arano H, **Valdes-Quezada C**, Valadez-Graham V, Recillas-Targa F. 2007. Globin genes transcriptional switching, chromatin structure and linked lessons to epigenetics in cancer: A comparative overview. Comp Biochem Physiol A Mol Integr Physiol. 147(3):750-60.

-Esta publicación se trata de una revisión donde se plantea la importancia del estudio de los dominios de genes  $\alpha$  y  $\beta$ -globin de pollo en la comprensión de los procesos involucrados en la regulación epigenética.

7. Guerrero G, Rincón-Arano H, Casasola A, **Valdes-Quezada C** and Recillas-Targa F. Targeted deletion of the chicken 3'-side  $\alpha$ -globin Enhancer reveals the presence of a novel regulatory element. En preparación.

- Para esta publicación, la cual se encuentra en preparación, se ha colaborado con la realización de experimentos de 3C de alta resolución y análisis de expresión en tiempo real que permitieron comprender la participación *in vivo* del elemento 3'-Enh en la estructura del dominio  $\alpha$ -globina de pollo.

**XVI. Artículo principal derivado del presente trabajo de investigación**

**XVII. Carátulas de los artículos publicados relacionados con el presente trabajo de investigación**

# CTCF demarcates chicken embryonic $\alpha$ -globin gene autonomous silencing and contributes to adult stage-specific gene expression

Christian Valdes-Quezada, Cristian Arriaga-Canon, Yael Fonseca-Guzmán, Georgina Guerrero and Félix Recillas-Targa\*

Instituto de Fisiología Celular; Departamento de Genética Molecular; Universidad Nacional Autónoma de México; México D.F., México

**Keywords:** DNA methylation, 5-hydroxymethylcytosine, chromosome conformation capture, epigenetic silencing, CTCF, enhancer-promoter interaction

Genomic loci composed of more than one gene are frequently subjected to differential gene expression, with the chicken  $\alpha$ -globin domain being a clear example. In the present study we aim to understand the globin switching mechanisms responsible for the epigenetic silencing of the embryonic  $\pi$  gene and the transcriptional activation of the adult  $\alpha^D$  and  $\alpha^A$  genes at the genomic domain level. In early stages, we describe a physical contact between the embryonic  $\pi$  gene and the distal 3' enhancer that is lost later during development. We show that such a level of regulation is achieved through the establishment of a DNA hypermethylation sub-domain that includes the embryonic gene and the adjacent genomic sequences. The multifunctional CCCTC-binding factor (CTCF), which is located upstream of the  $\alpha^D$  gene promoter, delimits this sub-domain and creates a transition between the inactive sub-domain and the active sub-domain, which includes the adult  $\alpha^D$  gene. In avian-transformed erythroblast HD3 cells that are induced to differentiate, we found active DNA demethylation of the adult  $\alpha^D$  promoter, coincident with the incorporation of 5-hydroxymethylcytosine (5hmC) and concomitant with adult gene transcriptional activation. These results suggest that autonomous silencing of the embryonic  $\pi$  gene is needed to facilitate an optimal topological conformation of the domain. This model proposes that CTCF is contributing to a specific chromatin configuration that is necessary for differential  $\alpha$ -globin gene expression during development.

## Introduction

Compartmentalization of the eukaryotic genome is one of the initial regulatory steps for gene expression. Subsequently, interdependent chromatin remodeling is needed to define sub-domains to allow highly specific regulation of gene expression at a local level.<sup>1,2</sup> To address these aspects of gene regulation in more detail, we adopted the chicken  $\alpha$ -globin gene domain as a paradigm to understand the influence of chromatin domain structure on gene expression (Fig. 1A).<sup>3</sup> For years we have been interested in understanding genomic domain-level mechanisms of differential  $\alpha$ -globin gene expression, not only during organism development but also during erythroid differentiation.

Hemoglobin switching represents the transition from embryonic to fetal and adult globin peptide synthesis in order to adjust oxygen requirements in the organism during development. This is accomplished through coordinated and differentially regulated  $\alpha$ - and  $\beta$ -globin gene expression (Fig. 1A). This process depends on a complex interplay among gene promoters, distal regulatory elements, nuclear dynamics, and the epigenetic milieu.<sup>4–6</sup>

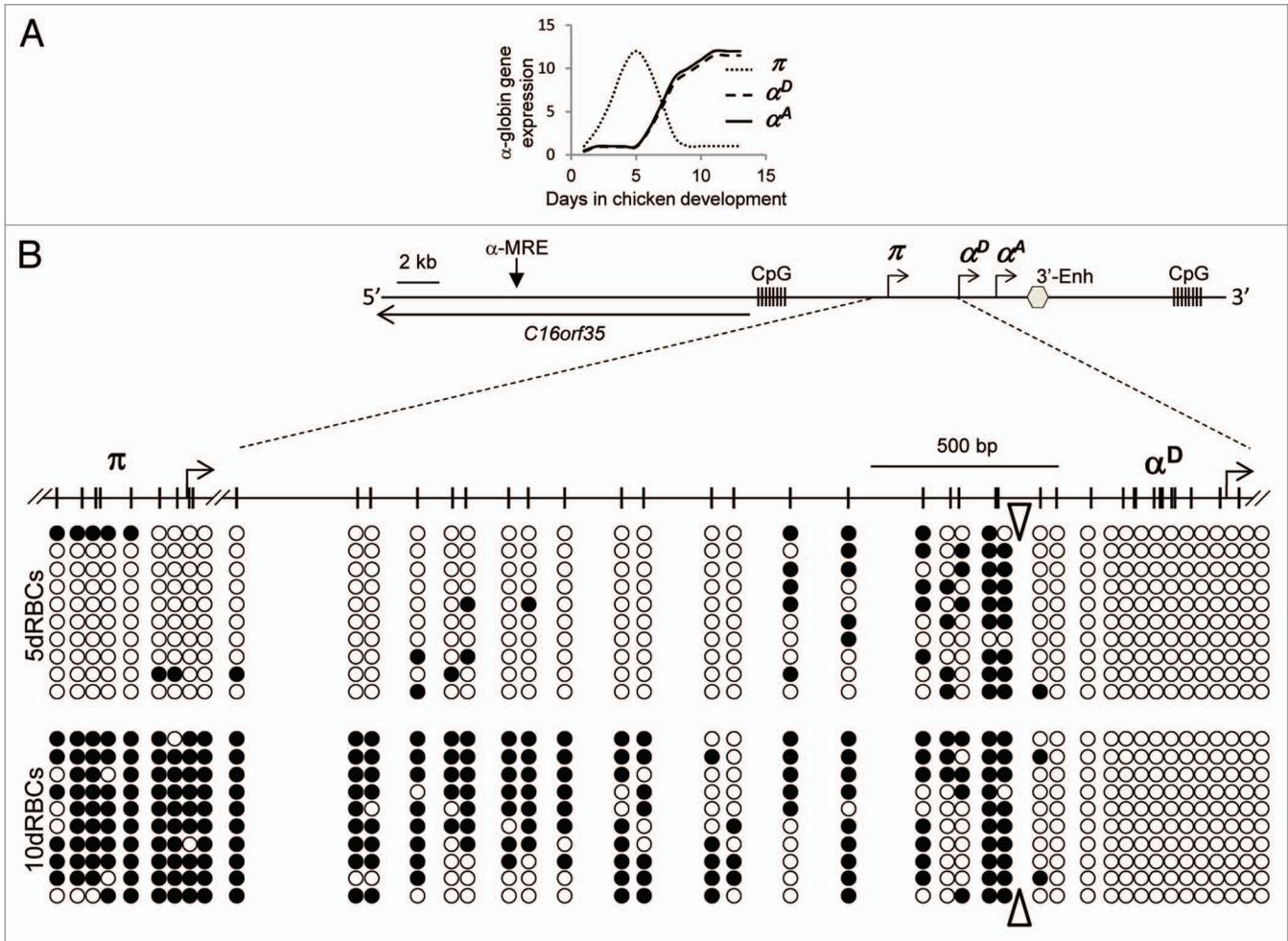
Unlike the chicken  $\beta$ -globin gene domain, the  $\alpha$ -globin domain lies in a constitutive open chromatin context on

minichromosome 14 and contains the embryonic  $\pi$  gene and adult  $\alpha^D$  and  $\alpha^A$  genes (Fig. 1B). The  $\alpha$ -globin locus covers a genomic region of around 50 kilobases (kb). On its 5' non-coding region there are several DNase I hypersensitive sites, including the locus control region known as the  $\alpha$ -major regulatory element ( $\alpha$ -MRE), followed by the  $\alpha$ EHS-1.4 element, which is a CTCF-dependent insulator.<sup>7</sup> This insulator is located within an intron of the *C16orf35* gene antisense transcript,<sup>8,9</sup> between the  $\alpha$ -MRE element and the  $\alpha$ -globin genes (Fig. 1B). Further downstream, 4 kb upstream of the  $\pi$  gene, there is a CpG island that corresponds to the promoter element of the *C16orf35* antisense transcript (Fig. 1B).

Besides three  $\alpha$ -globin genes ( $\pi$ ,  $\alpha^D$ , and  $\alpha^A$ ) and their promoters, the domain contains several regulatory elements located on its 3' non-coding region. In particular, there is an erythroid-specific 3' enhancer located at a distance of 0.9 kb after the  $\alpha^A$  gene.<sup>10</sup> As the promoters of the 3 genes, the 3' enhancer activity depends on the binding of the erythroid-specific transcription factor GATA-1, other transcription factors and its chromatin structure.<sup>10</sup>

In the field, it has been suggested that hemoglobin switching is a process that involves, in part, autonomous silencing

\*Correspondence to: Félix Recillas-Targa; Email: frecilla@ifc.unam.mx  
Submitted: 04/11/13; Revised: 06/17/13; Accepted: 06/20/13  
<http://dx.doi.org/10.4161/epi.25472>



**Figure 1.** Different DNA methylation patterns of the chicken  $\alpha$ -globin locus during development. **(A)** Differential  $\alpha$ -globin gene expression profiles during chicken development.<sup>3</sup> **(B)** Scheme of the  $\alpha$ -globin domain showing the location of intergenic CpG-islands, the  $\alpha$ -MRE locus control region and the anti-sense *cgthba/C16orf35* transcript. Dotted lines define the genomic region analyzed by bisulfite transformation. In the lower panel we show the DNA methylation distribution of the 5- and 10-days RBCs (5dRBCs and 10RBCs, respectively). Open circles correspond to unmethylated CpGs and black circles represent methylated CpGs. Only the 3' end of the embryonic  $\pi$  body gene was analyzed. The empty triangles indicate the methylation transition zone.

of the embryonic gene promoter and competition for proximal and distal regulatory elements.<sup>3,4,11</sup> Unfortunately, there is scarce experimental evidence describing the mechanisms for such autonomous silencing. Previous work from our research group suggests there is autonomous epigenetic repression of the embryonic  $\pi$  gene.<sup>3</sup> We have shown that the embryonic  $\pi$  gene promoter is DNA methylated and bound by MeCP2, and the histones are mainly deacetylated in order to induce embryonic gene silencing, which then allows regulated transcriptional activation of the adult chicken  $\alpha$ -globin genes late in development.<sup>3</sup> In our previous studies, we have described the regulatory action of an enhancer, located in the 3' non-coding region of the locus, over the three chicken  $\alpha$ -globin genes.<sup>3,10,12</sup> One conclusion of those studies is that the autonomous silencing of the embryonic  $\pi$  gene also contributes to the selectivity of the 3' enhancer during development. Here, we decided to address the mechanisms responsible for the establishment of the autonomous silencing.

In the present study, we systematically analyzed the DNA methylation status of the embryonic  $\pi$  gene, its promoter and the 3' intergenic sequences up to the adult  $\alpha^D$  promoter region in different developmental stages and during erythroid cell differentiation. Based on our previously published genome-wide distribution of the multifunctional nuclear factor CTCF, we found a recognition site in proximity of the adult  $\alpha^D$  promoter. A series of mutant constructs stably integrated in an erythroid cell line seem to contribute to the formation of a stage-specific DNA methylation sub-domain. These observations further support our model in which delimited chromatin structural changes are responsible for directing the autonomous silencing of the embryonic  $\pi$  gene, which subsequently allows activation of the adult  $\alpha^D$  and  $\alpha^A$  genes late in development and during definitive erythropoiesis. A model is discussed based on the possible role of CTCF in defining the sub-domain chromatin transition, in addition to its possible role in long distance interactions between the regulatory elements of the locus.

## Results

### Differential DNA methylation of the embryonic $\pi$ globin gene and surrounding genomic sequences during development.

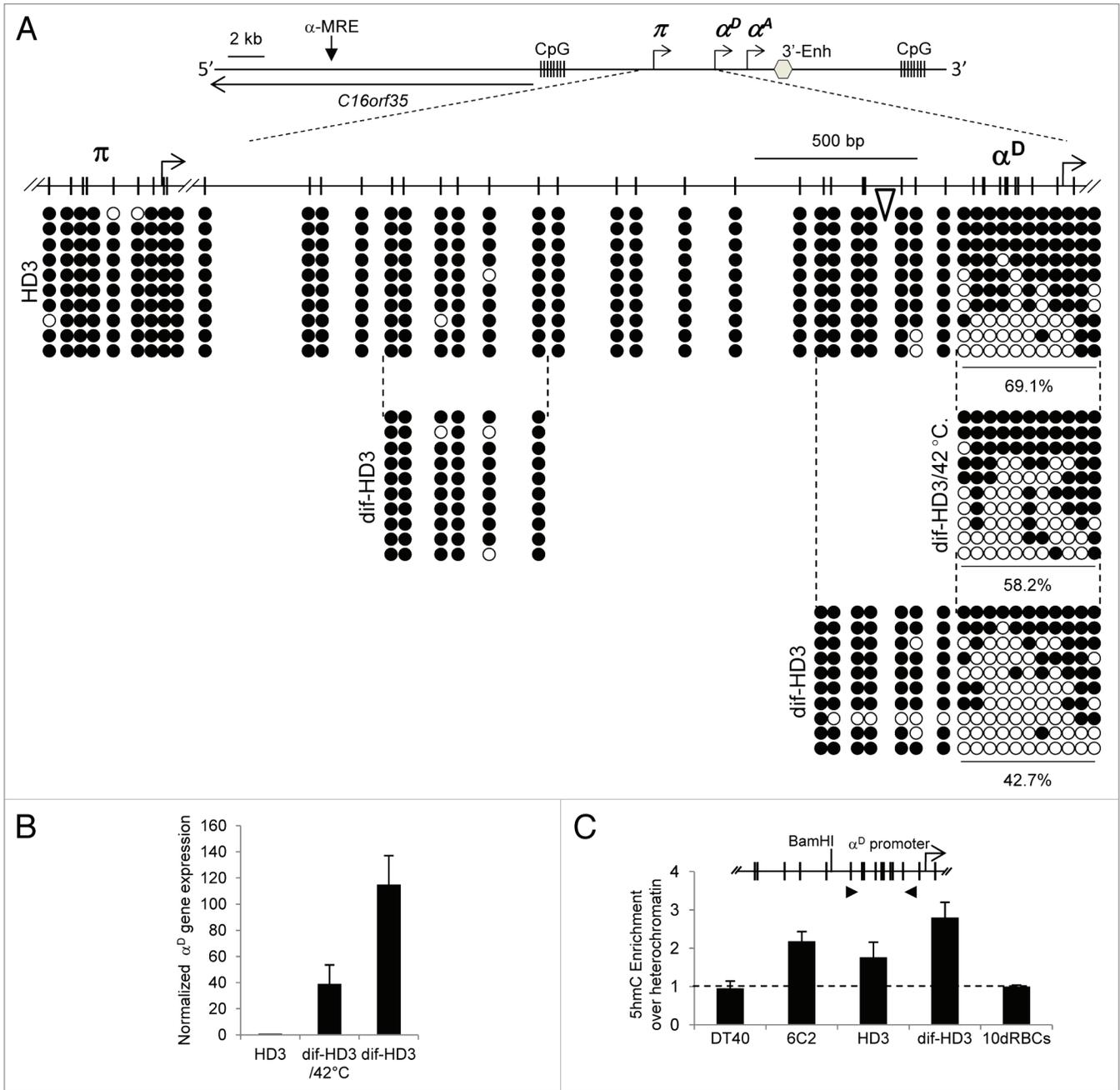
Based on our previous observations that the embryonic  $\pi$  gene promoter is not DNA methylated in 5-days-old chicken embryos (5dRBCs) but is methylated in 8dRBCs and 10dRBCs,<sup>3</sup> we asked whether such methylation is restricted to the promoter or if it can be expanded, in particular, toward the adult  $\alpha^D$  gene (Fig. 1B). This aim is further supported by the fact that DNA methylation can be propagated along the genome.<sup>13-15</sup> We performed sodium bisulfite transformation of genomic DNA and sequencing from 5dRBCs (where only the embryonic  $\pi$  gene is expressed) in comparison to 10dRBCs (where the adult  $\alpha^D$  and  $\alpha^A$  genes are expressed) (Fig. 1A and B). We found that the promoter regions of the  $\pi$  and  $\alpha^D$  genes and the intergenic sequences in 5dRBCs are unmethylated, with the exception of two characteristic CpGs that are located 540 bp from the transcription start site (TSS) of the  $\alpha^D$  gene (Fig. 1B). In contrast, in 10dRBCs, the  $\pi$  promoter and the intergenic region between the  $\pi$  and  $\alpha^D$  genes are heavily methylated (Fig. 1B). However, the most appealing observation is the presence of a notorious transition between a DNA methylated and non-methylated state delimited by 2 constitutively methylated CpG-dinucleotides ( $C^m pG$ ). This result supports a well-defined DNA methylation sub-domain (Fig. 1B, see empty triangles) coincident with the silencing of the embryonic  $\pi$  gene and the transcription of the adult  $\alpha^D$  and  $\alpha^A$  genes.

**Partial DNA demethylation during erythroid differentiation.** In order to further understand the role of DNA methylation during erythroid differentiation we took advantage of HD3 cells, which are AEV-transformed chicken pro-erythroblast cells corresponding to chicken adult erythroid cells arrested at early stages of differentiation.<sup>16</sup> Importantly, these cells do not express globin genes, even though they possess a permissive globin domain chromatin structure.<sup>10,17</sup> Interestingly, these cells can specifically transcribe the adult  $\alpha^D$  and  $\alpha^A$  genes after induction of terminal erythroid differentiation (Fig. S1A). With this inducible system we were able to examine the DNA methylation distribution during the transition from the inactive state to transcriptional activation. Thus, we compared the DNA methylation patterns of genomic DNA from HD3 cells and differentiated HD3 cells (Fig. 2A). To induce HD3 differentiation (dif-HD3) we cultured the cells in the presence of the Iso-H-7 inducer at 42°C for 48 h or exclusively at 42°C (dif-HD3/42°C) (Fig. S1A).<sup>17</sup> Total RNA and genomic DNA were then isolated to verify the induction efficiency through the transcriptional activation of the adult  $\alpha^D$  and  $\alpha^A$  genes (Fig. 2B).<sup>3,17</sup> We found that the entire region is DNA hypermethylated in undifferentiated conditions (Fig. 2A). However, once the HD3 cells were induced to terminally differentiate and  $\alpha^D$  and  $\alpha^A$  gene expression was initiated, we observed a localized DNA demethylation of the  $\alpha^D$  promoter region (69.1% vs. 42.7%) (Fig. 2A). This is confirmed by a control genomic region located in the upstream intergenic region that does not show any reduction in DNA methylation (Fig. 2A). Supporting these data, intermediate levels of induction reached only by a shift in temperature to 42°C (Fig. 2B,

dif-HD3/42°C; Fig. S1A) clearly show intermediate levels of  $\alpha^D$  gene expression and promoter DNA demethylation (58.2%) in comparison to the levels seen upon robust induction conditions (dif-HD3). Interestingly, we notice that the position of individual CpGs in the  $\alpha^D$  promoter that suffer a pronounced DNA demethylation under differentiation conditions are located in proximity to the binding sites for GATA-1 and Sp1 transcription factors (Fig. 3).<sup>3</sup>

Since there is growing evidence supporting active enzymatic reactions capable of erasing or modifying pre-existing DNA methylation patterns,<sup>18</sup> we first investigated the possibility of active DNA demethylation over the  $\alpha^D$  promoter region. We analyzed the presence and abundance of the 5-hydroxymethylcytosine (5hmC) modification in RBCs and chicken cell lines (Fig. S1C). Using a specific antibody against 5hmC, we detected a modest recognition in chicken erythroid and non-erythroid cells compared with mouse embryonic stem cell genomic DNA, which was used as positive control (Fig. S1C).<sup>19</sup> Then, we performed a quantitative 5hmC-DNA immunoprecipitation (hMe-DIP) analysis (Fig. S1D). We found that 5hmC is already present in early erythroid differentiation stages as seen in the pre-erythroblast 6C2 cells and the undifferentiated HD3 cells (Fig. 2C). These results are further supported by no enrichment of 5hmC and DNA hypermethylation in the non-erythroid chicken lymphoid DT40 cell line (Fig. 2C; Fig. S1B). Therefore, we observe an erythroid-specific gain in 5hmC that is coincident with the activation of the  $\alpha^D$  gene (Fig. 2C). We do not believe that DNA demethylation of the  $\alpha^D$  gene promoter has a direct effect in its robust transcription activation. Different levels of regulation are probably involved including the activation of specific-signal transduction cascades, the synthesis of stage-specific transcription factors, the recruitment of chromatin remodelers and the silencing of the embryonic  $\pi$  gene. This finding suggests an indirect correlation between DNA demethylation and transcriptional activation of the adult  $\alpha^D$  gene once HD3 cells are induced to differentiate. Taking these results together, we conclude that DNA demethylation is, to some extent, coincident with adult gene transcriptional activation. Further investigations are needed to determine a direct link between DNA demethylation and transcriptional activation of  $\alpha$ -globin adult genes.

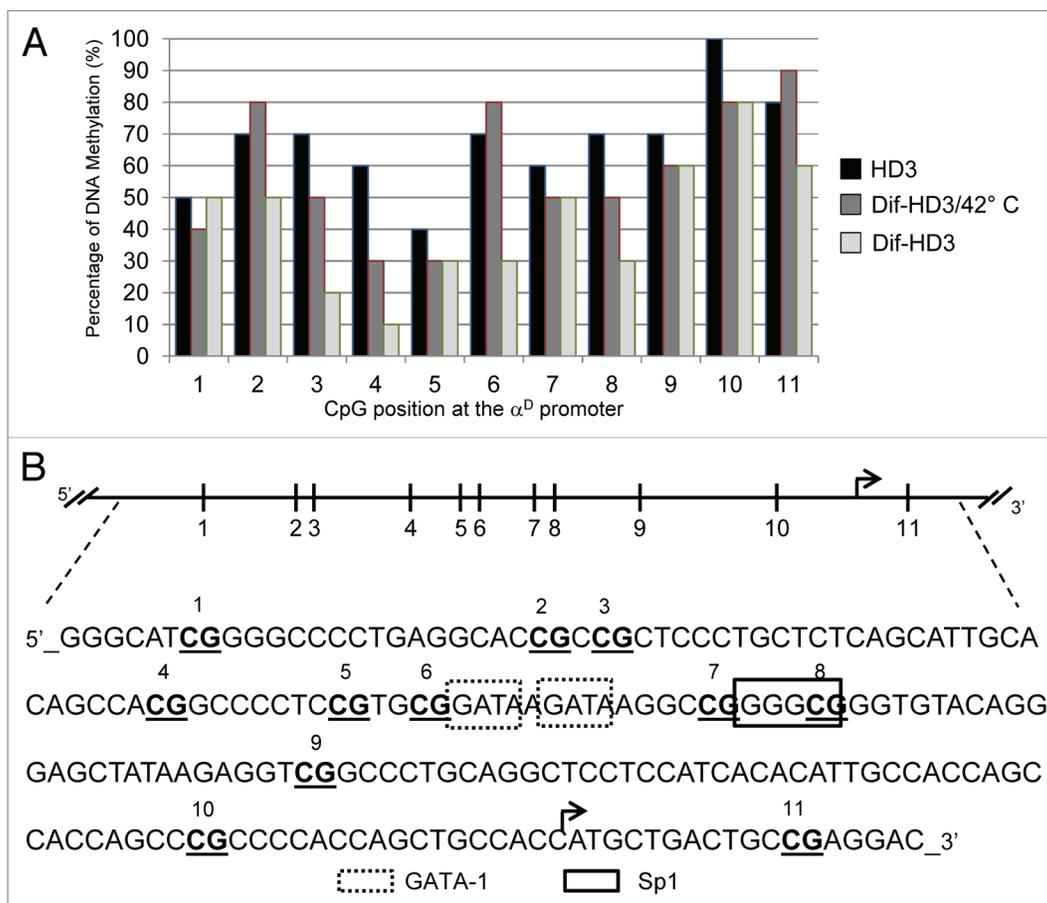
**Active histone marks and CTCF enrichment at the DNA methylation transition zone.** Next, we focused our study on understanding the chromatin features of the DNA methylation transition zone (Fig. 1B, see empty triangle). We assessed the enrichment of a series of histone marks, the presence of the H2A.Z histone variant, and the nuclear factor CTCF. For CTCF, we used an antibody that we generated to recognize the N-terminal of chicken CTCF, and we took advantage of our previously published ChIP-seq series of experiments using immunoprecipitated chromatin from RBCs isolated from chicken embryos at 5 and 10 days of development.<sup>20</sup> In this report, we found more than 5,000 CTCF binding sites were identified in 5dRBCs, and more than 21,000 in 10dRBCs,<sup>20</sup> among them a novel CTCF-binding site that was found 190 bp upstream of the  $\alpha^D$  gene TSS (Fig. 4A; Fig. 5A). Based on this novel CTCF location, in the present study we decided to validate its presence by semiquantitative duplex



**Figure 2.** DNA demethylation and gain of the 5hmC modification correlates with adult  $\alpha^D$  gene transcriptional activation. **(A)** Analysis of the DNA methylation status in the transformed avian HD3 cell line in non-induced (HD3) and differentiation-induced (dif-HD3; Iso-H-7 and 42°C) conditions. For comparative purposes an intermediate degree of induction is incorporated (see dif-HD3/42°C; induction only with 42°C). The percentage of the CpG methylation in the promoter region of the  $\alpha^D$  gene is shown. The empty triangle indicates the methylation transition zone as in (Fig. 1B). **(B)** Quantitative real-time PCR showing the differentiation efficiency through the adult  $\alpha^D$  gene transcriptional activation. **(C)** Quantitative 5-hydroxymethylcytosine-DNA immunoprecipitation was assessed using an antibody against 5hmC and primers located in the  $\alpha^D$  promoter. As controls we used bisulfite transformation to test the chicken B-lymphocyte DT40 cell line, where the  $\alpha^D$  promoter is DNA hypermethylated, and also 6C2 pre-erythroblast cells (Fig. S1B). Values higher than 1 (discontinuous line) are considered real enrichments. All the error bars were calculated from two independent experiments.

PCR using primers extending from the CTCF site and associated histone marks (Fig. 4B; Fig. S2). The results show that CTCF is bound to the upstream limit of the chicken  $\alpha^D$  promoter in 5dRBCs and 10dRBCs, but absent in HD3 cells, consistent with the DNA hypermethylation seen in this cell line. Interestingly,

there is no enrichment of negative histone marks in any condition (H3K9me3 and H3K27me3; Fig. S3), and this observation is in agreement with the data published by others.<sup>21</sup> Instead, we observed a constant incorporation of active histone marks like H3ac, H4ac, and H3K4me2 (Fig. 4B).

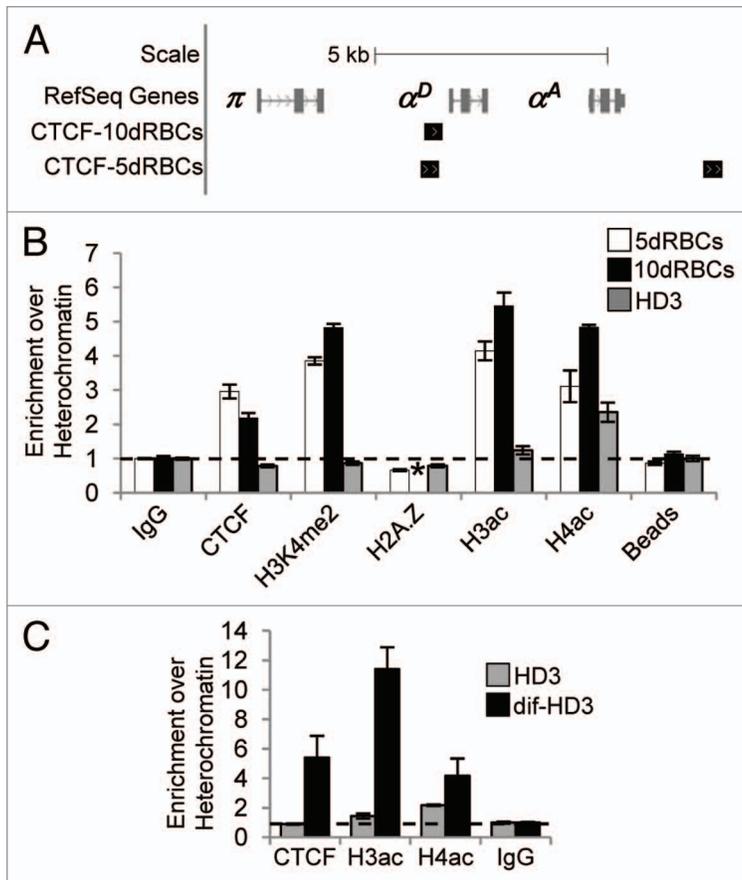


**Figure 3.** DNA methylation overview at individual CpG-dinucleotides in the  $\alpha^D$  promoter. **(A)** A comparative percentage of DNA methylation is shown in the graph. The data are generated from the results in Figure 2A. HD3 cells represent undifferentiated erythrocytes, dif-HD3/42°C which corresponds to the intermediate condition of differentiation induction, and dif-HD3 represents the highest differentiation induction condition. At the bottom, the distribution and position of the CpGs in the  $\alpha^D$  promoter is indicated. Interestingly, CpGs located in positions 3, 4, 6, and 8 are subjected to most significant levels of DNA demethylation. **(B)** DNA nucleotide sequence of the  $\alpha^D$  promoter region analyzed by sodium bisulfite conversion and sequencing. The CpGs that are preferentially demethylated (positions 3, 4, 6, and 8) coincide with a DNA segment where a tandem GATA-1 binding-site and one Sp1 binding-site are found.<sup>3</sup>

In order to verify the chromatin status of the region in induced HD3 cells, we performed a comparative ChIP from HD3 cells and induced HD3 cells (Fig. 4C). In undifferentiated HD3 cells there is no enrichment of H3ac and a modest increment of H4ac. Furthermore, there is no binding of CTCF. These observations are in accordance with the DNA hypermethylation status of the region and the lack of  $\alpha^D$  gene expression. In contrast, under HD3 cells differentiation conditions (dif-HD3) there is a gain of open histone chromatin marks and binding of CTCF, favoring the idea of barrier element formation. Such a configuration may delimit the spreading of DNA methylation and it defines a chromatin sub-domain necessary for the autonomous silencing of the embryonic  $\pi$  gene and the expression of the adult genes in terminally differentiated erythroid cells.

**The absence of the CTCF site in the transition zone induces rapid epigenetic silencing in stably transformed cells.** Based on its location and the results obtained up to this point, we decided to demonstrate the barrier function of this CTCF motif located upstream of the  $\alpha^D$  gene promoter. Reducing the relative

abundance of CTCF in HD3 cells was an alternative that we discarded based on the fact that CTCF is bound in vivo at least to four different sites along the locus, rendering the interpretation of the results very difficult.<sup>7,20,22</sup> In addition, RNA interference experiments are inefficient in the non-dividing 5dRBCs and 10dRBCs. Thus, to address this aim we generated a series of constructs that incorporate or remove the different CTCF-binding sites (Fig. 5). On the basis of the canonical CTCF motif that emerged from the published ChIP-seq data and bioinformatics prediction,<sup>20</sup> we analyzed the genomic sequence in more detail and found two potential binding motifs about 190 bp upstream of the  $\alpha^D$  gene TSS (Fig. 5A).<sup>20</sup> Derived from this prediction and the size of CTCF-binding sites, we designed two different deletions that excluded each predicted motif (C1: proximal site and C2: distal site) (Fig. S4). It is worth mentioning that none of the previously characterized binding sites for transcription factors required for  $\alpha^D$  promoter activity were removed.<sup>3,12</sup> These constructs were stably transfected in HD3 cells and pools were selected and maintained in continuous cell culture. *GFP*



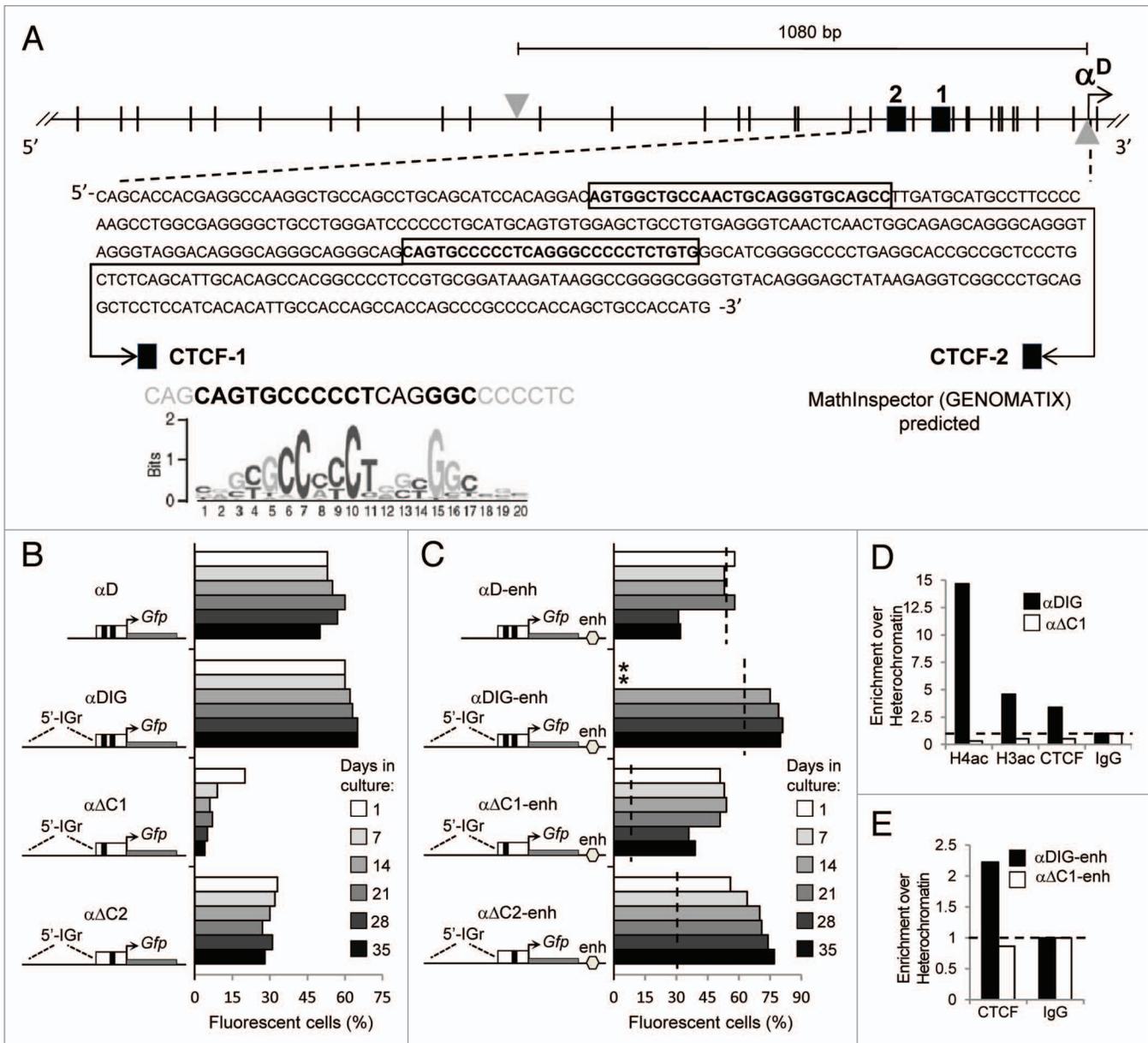
**Figure 4.** In vivo association of the CTCF nuclear factor and histone marks over the adult  $\alpha^D$  gene promoter. (A) Location of the new CTCF binding site determined in a previous publication by chromatin immunoprecipitation and massive sequencing (ChIP-seq);<sup>20</sup> black squares represent CTCF in vivo occupancy (May 2006, WUGSC 2.1/galGal3 Assembly; UCSC, Genome Browser). (B) Semi-quantitative and comparative ChIP assays validating the in vivo binding of CTCF and the incorporation of histone post-translational modifications at the  $\alpha^D$  gene promoter. (C) ChIP experiments comparing non-induced HD3 cells (HD3) and differentiation-induced cells (dif-HD3). The same primers are used in (B) and (C). Asterisks indicate experiments not done. Values higher than 1 represent real enrichments (discontinuous line). The error bars were calculated from two or more independent experiments.

expression was evaluated by GFP fluorescence emission by flow cytometry (FACS), revealing that the loss of the proximal CTCF site ( $\alpha\Delta C1$ ) is associated with a rapid and robust decay of transgene expression (Fig. 5B). Next, we performed a ChIP assay comparing the intact transgene ( $\alpha DIG$ ) and the transgene lacking the proximal CTCF site ( $\alpha\Delta C1$ ), which resulted in loss of CTCF binding, a drastic decrease in histone H3 and H4 acetylation and a decrease in *GFP* expression after 35 days of continuous cell culture (Fig. 5D). From these data we predict that the CTCF-1 motif is the preferential site responsible for such activity. These results support our model in which CTCF is creating a barrier against silencing chromatin signals. Nevertheless, based on the rapid loss of transgene activity in the  $\alpha\Delta C1$  construct, we cannot rule out the possibility that CTCF is also having a transcriptional activation function on the  $\alpha^D$  gene promoter (Fig. 5B; Day 1).

With the intention to differentiate CTCF activity from classical transcription factor and barrier element activity, we repeated the same experiment but this time incorporating the 3'  $\alpha$ -globin enhancer in our transgenes.<sup>3,10</sup> The rationale behind this experiment is that in the absence of CTCF, the enhancer may compensate by *trans*-activating the transgene through the formation of an optimal chromatin configuration. This prediction is further supported by published data demonstrating that enhancers can positively influence the chromatin structure of a transgene and favor their *trans*-activation potential.<sup>23,24</sup> However, with the incorporation of the 3' enhancer there was between 40–60% transgene expression reduction in the  $\alpha\Delta C1$ -enh transgene, but the overall expression was maintained in the other constructs and over time (Fig. 5C). The absence of CTCF binding was corroborated by ChIP analysis (Fig. 5E). In agreement with our prediction the absence of the CTCF site did not affect promoter functionality and the enhancer retained its *trans*-activation potential. Of note, and based on the ChIP-seq data,<sup>20</sup> we validated another CTCF site in the 3' enhancer (Fig. 7 and B). Thus, the presence of CTCF at both the  $\alpha^D$  promoter and the 3' enhancer favors a robust and sustained transgene expression by the establishment of a permissive chromatin configuration. Therefore, these results also suggest that CTCF is acting as a chromatin remodeler but at this point we are not able to discard a role of CTCF as a classical transcriptional factor. In any case, both activities cannot be excluded.

**Gain of DNA methylation in the  $\alpha^D$  gene promoter upon CTCF binding site removal.** To further assess the role of CTCF and on the basis of its capacity to nucleate histone acetylases and methylases, we incubated the cells expressing the different transgenes with the histone deacetylase inhibitor, trichostatin-A (TSA), in the absence or the presence of the 3' enhancer (Fig. 6A). Interestingly, we found that in the absence of the enhancer there was no transgene reactivation in the  $\alpha\Delta C1$  constructs. In contrast, when the enhancer was incorporated in the transgene, we observed a more robust transgene expression in response to TSA (Fig. 6A). Thus, the presence of CTCF is probably necessary to recruit histone modifiers such as HATs and HMTs to the  $\alpha^D$  promoter region.

To further explore the contribution of CTCF in demarcating the DNA methylation sub-domain, we surveyed the DNA methylation of the  $\alpha^D$  gene promoter in the transgenes both with and without the CTCF-binding sequence ( $\alpha DIG$  vs.  $\alpha\Delta C1$  and  $\alpha\Delta C1$ -enh). We found a gain in DNA methylation after 25 days of continuous cell culture in the absence of the CTCF-binding sequence (Fig. 6B; Fig. S5). It is worth mentioning that at Day 1 no DNA methylation was found in any of the three transgenes analyzed (Fig. 6B). This is consistent with the low expression levels of the  $\alpha\Delta C1$  transgene (Fig. 5B), and is in agreement with previous published data demonstrating that transgene silencing begins with histone deacetylation followed by gain in DNA methylation.<sup>25</sup> In addition, even in the presence of the 3'

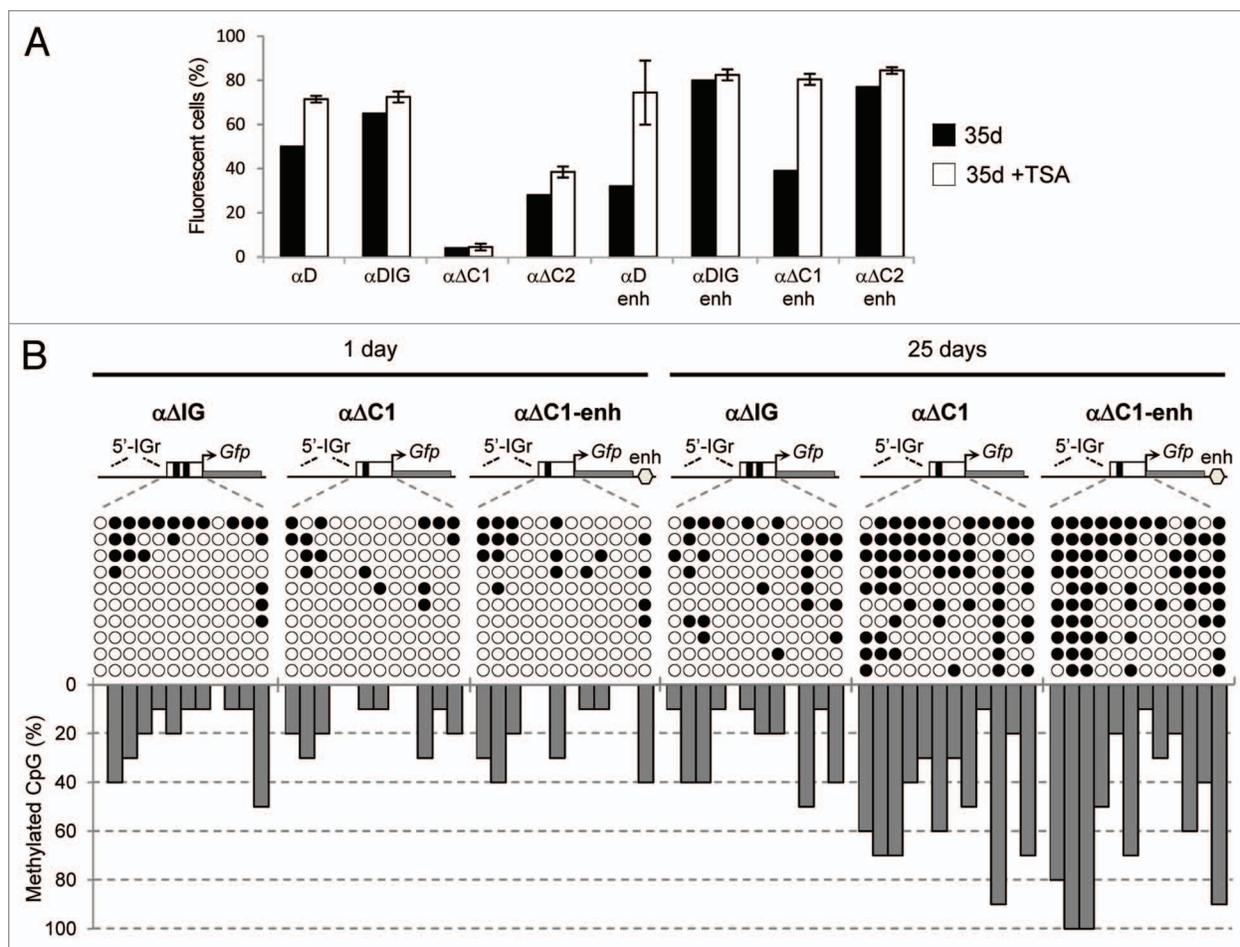


**Figure 5.** CTCF is a key component for  $\alpha^D$  gene promoter function. **(A)** Schematic representation and nucleotide sequence of the putative CTFC binding motifs in the  $\pi$  -  $\alpha^D$  intergenic region. Using the Logo generated from the chicken CTFC ChIP-seq experiments we identified a potential binding motif located 190 bp upstream of the  $\alpha^D$ -TSS (CTCF-1, black square 1).<sup>20</sup> Additionally, we looked for other potential binding sites using bioinformatics tools (MathInspector) and we found another CTFC-binding motif (CTCF-2, black square 2). Both binding sites are shown and for the CTCF-1 site the published Logo is included for comparison purposes. The vertical lines show the distribution of CpG-dinucleotides along the studied region. The gray triangles indicate the sequences cloned for the transgene experiments (see below). We incorporated the core  $\alpha^D$  promoter (white box in schemes) containing the putative CTFC-binding sites (black boxes inside;  $\alpha^D$  and  $\alpha^{DIG}$ ) driving the expression of the *GFP* gene (for details see Fig. S4). **(B)** We performed deletions of the proximal ( $\alpha\Delta C1$ ) and distal ( $\alpha\Delta C2$ ) CTFC-binding sites. 5'-IGr corresponds to a DNA segment from the upstream intergenic region (938 bp). **(C)** We introduced the 3' enhancer (enh) downstream of the *GFP* gene in each construct described in **(B)**. The discontinuous line corresponds to the average (%) cell fluorescence in **(B)**. **(D)** ChIP over transgene constructs at day 35 of culture using primers over the  $\alpha^D$  gene promoter (forward) and the *GFP* gene (reverse). **(E)** ChIP for transgene as in **(D)** in the constructs that harbor the 3' enhancer.

enhancer ( $\alpha\Delta C1$ -enh), we observed an increment in DNA methylation concomitant with a partial reduction of transgene expression (Fig. 5C; compare  $\alpha^{DIG}$ -enh with  $\alpha\Delta C1$ -enh). This result was confirmed when transgenic cells were treated with the DNA methylation inhibitor 5-aza-2'-deoxycytidine (5-azadC), which induced reactivation of transgene expression in the presence of

the 3' enhancer and absence of the CTFC-1 binding site (Fig. S6;  $\alpha\Delta C1$ -enh).

Altogether, this series of experiments is in agreement with a model whereby the CTFC site located 190 bp from the  $\alpha^D$  gene TSS represents a barrier element that counteracts the propagation of silencing chromatin, in particular, DNA methylation. In



**Figure 6.** Transgene reactivation assays and DNA methylation analysis. (A) TSA-treatment at day 35 of culture. (B) DNA methylation analysis over the transgene. The percentage at individual CpG-dinucleotides of the transgene comparing day 1 and day 25 of continuous cell culture is shown. Interestingly, CpGs located in the vicinity of the deleted CTCF binding-site gain preferentially DNA methylation. The error bars were calculated from two independent experiments.

conjunction, all these factors facilitate the establishment of a chromatin configuration that allows the specific and regulated transcriptional activation of the adult  $\alpha^D$  gene.

**Stage-specific chromatin loop formation between the embryonic  $\pi$  gene promoter region and the 3' enhancer.** Up to this point, we cannot exclude stage-specific three-dimensional contacts between the gene promoters, the 3' enhancer and other regulatory components of the locus.<sup>26</sup> The spatial organization of the chicken  $\alpha$ -globin locus has already been established during erythroid differentiation (HD3 cells) and only during late development, revealing the differential formation of chromatin hubs.<sup>17</sup>

With all this in mind we asked whether there is a physical interaction between the 3' enhancer and the embryonic  $\pi$  promoter region, an aspect that has not been previously explored (Fig. 7A). To address this question, we performed a chromosome conformation capture (3C) assay in 5dRBCs vs. 10dRBCs (Fig. 7; Fig. S7).<sup>27,28</sup> In agreement with the data presented here, we found a physical association between the embryonic promoter region and the 3' enhancer only in 5dRBCs (Fig. 7C). No significant association was seen in 10dRBCs, when the promoter and surrounding

regions of the  $\pi$  gene are DNA hypermethylated and the  $\alpha^D$  gene is transcriptionally active. Of note, such interactions do not seem entirely dependent on CTCF, as suggested for other loci,<sup>29</sup> since we have not been able to demonstrate CTCF binding around the  $\pi$  gene (Fig. S8). In contrast, we found CTCF binding at the 3' enhancer in 5dRBCs and 10dRBCs (Fig. 7A and B). Of note, we found that the 3' enhancer is constitutively contacting the  $\alpha^D$  promoter region even though there is no *trans*-activation (like in the 5dRBCs) (Fig. 7). This is in agreement with the previously proposed role of the  $\alpha^D$  promoter region as a constitutive anchor point for the  $\alpha$ -globin domain chromatin hub.<sup>17,21</sup> In conclusion, even though the locus is in a general open chromatin conformation there is a preference for the 3' enhancer to interact with the  $\pi$  promoter region in early stages of development.

## Discussion

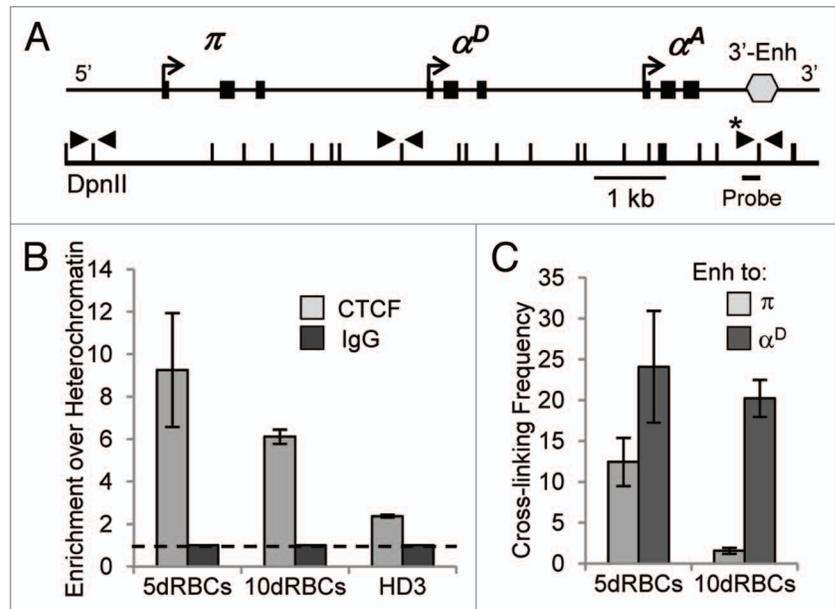
The genomic distribution and its topological orientation within the nucleus hierarchically represent one of the first levels of gene regulation. On the local scale the genome requires another degree

of sophistication with the definition of transcriptionally active domains and sub-domains that are shielded from neighboring signals or even genes. During development we observed the establishment of a DNA methylation sub-domain, which is demarcated by the binding of CTCF and other transcription factors.<sup>3</sup>

Concerning the protection against DNA methylation expansion, our data are consistent with the observation that a sub-set of CTCF-binding sites found in the vicinity of tumor suppressor genes and microRNAs prevent epigenetic silencing.<sup>14,30,31</sup> Furthermore, the presence of CTCF and other nuclear factors upstream of the  $\alpha^D$  gene TSS establishes that DNA-binding factors are involved in creating an unmethylated state even in the absence of transcription, as is the case in 5dRBCs (Fig. 3B; Fig. 4B).<sup>3,15,32,33</sup> Similarly, this type of barrier function that prevents DNA methylation was described in association with the activity of the VEZF1 nuclear factor in the chicken  $\beta$ -globin *chs4* insulator elements.<sup>33</sup> Interestingly, mutation of the VEZF1 binding site in stably transfected constructs demonstrated the spreading of DNA methylation over the  $\beta^A$  promoter and transgene silencing.<sup>33,34</sup> The stable lines that we generated with the deletion of the CTCF site support an equivalent role for CTCF, even though the CTCF partners and/or post-translational modifications associated with such a particular role remain to be determined.<sup>7</sup>

The enrichment of the histone mark H3K4me2 in 5dRBCs and 10dRBCs is in accordance with the published data demonstrating that hypomethylated CpG-islands show elevated levels of H3K4me2, even in the absence of transcription. In fact, it has been demonstrated that the H3K4me2 modification occurs uniformly on all types of CpG-island promoters, supporting the concept that this is an inherent feature of CpG-islands.<sup>35</sup> Furthermore, it has also been shown that broad histone H3 hyperacetylation occurs in CpG-islands.<sup>35</sup> Together, we consider CTCF and its associated co-factors and chromatin remodelers to be responsible for the formation of an active chromatin state which precludes the propagation of DNA methylation. A remarkable aspect of the genomic region under study is its capacity to be DNA demethylated through the gain of 5hmC in response to cellular differentiation signals (dif-HD3 cells). Interestingly, a regular distribution of 5hmC around CTCF binding sites has recently been reported in mouse embryonic stem cells.<sup>36</sup> This observation implies that CTCF and/or its associated partners are involved in the establishment of a 5hmC profile in a regulated manner to achieve regular and optimal nucleosome positioning.<sup>36</sup> Thus, future investigations should address the DNA demethylation of gene promoters in response to cellular differentiation and the role, if any, of CTCF in this process.

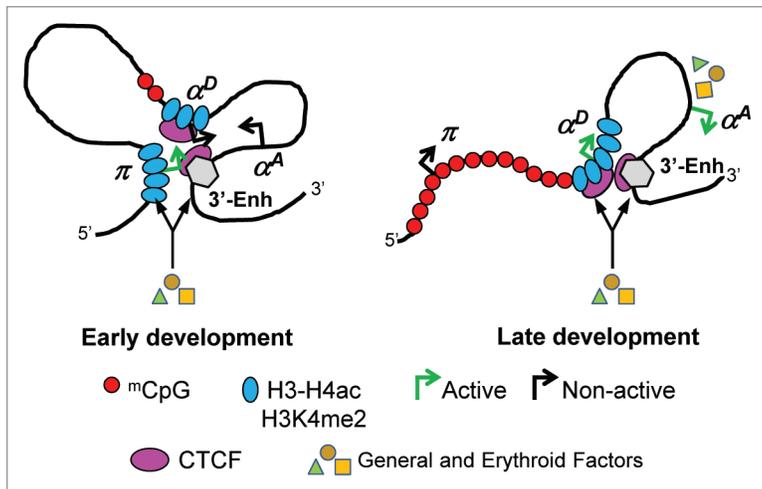
An alternative mechanism for CTCF could be related to the work of Caiafa and collaborators showing that the



**Figure 7.** Physical contacts between the 3' enhancer and the  $\pi$  and  $\alpha^D$  genes. (A) The scheme shows the distribution of the 4-base cutter DpnII restriction enzyme over the  $\alpha$ -globin genes and the 3' enhancer (3'-Enh; black boxes represent gene exons). The arrowheads correspond to the test-primers and the asterisk marks the anchor-primer. (B) CTCF occupancy over the 3'-Enh region analyzed by ChIP. (C) 3C experiments performed in 5dRBCs and 10dRBCs. The cross-linking frequency was calculated from two independent experiments (Table S2) and associated controls are shown in detail; see Figure S7. The error bars were calculated from two or more independent experiments.

poly(ADP-ribose) polymerase 1 (PARP-1) is self-poly(ADP-ribosylated) in the presence of CTCF.<sup>37</sup> Such an interaction generates a CTCF-PARP-1-Dnmt1 complex that interferes with the activity of the Dnmt1 DNA methyltransferase. Interestingly, CTCF dissociation abolishes such interactions and favors DNA hypermethylation in the vicinity of the CTCF sites.<sup>38</sup> Studies from our group in the chicken  $\alpha$ -globin locus demonstrated that CTCF activity can be modulated by poly(ADP-ribosylation).<sup>7</sup> Together, we conclude that the activity of CTCF located in proximity to the  $\alpha^D$  gene promoter can be, in part, regulated by poly(ADP-ribosylation) and/or by blocking DNA methylation through its interaction with PARP-1.

A complementary aspect of the regulation and structure of the chicken  $\alpha$ -globin domain is its three-dimensional organization. One of the most relevant aspects of such a spatial organization is that the promoter region of the adult  $\alpha^D$  gene is apparently a key contact point, which is coincident with the newly identified CTCF-binding site.<sup>17,20</sup> In such a scenario, one possible model is that the embryonic  $\pi$  region can be excluded from the hub, in part, by its autonomous silencing in later stages of development. This view is consistent with the one described for the mouse  $\beta$ -globin locus.<sup>39</sup> Furthermore, and as suggested previously,<sup>7</sup> the combination of molecular allies and post-translational modifications may also induce a differential selectivity in the function of the CTCF sites located along the domain. At this point, we cannot discard the complementary intervention of non-coding RNAs in such multi-step regulation.



**Figure 8.** Autonomous silencing of the embryonic  $\pi$  gene and the derived topological conformation of the  $\alpha$ -globin domain. In early development (5dRBCs), the model shows the formation of a hub composed by interactions between the  $\pi$  gene, the  $\alpha^D$  promoter region and the 3' enhancer. In late development (10dRBCs), the embryonic  $\pi$  gene and surrounding genomic regions are DNA methylated with its concomitant dissociation of the hub. At this stage, the physical contact between the  $\alpha^D$  gene promoter and the 3' enhancer is maintained among other previously described interactions.<sup>17</sup>

In summary, the results presented here allow the proposal of an integrative model of regulation that incorporates diverse components and the coordination of multiple genetic and epigenetic processes (Fig. 8). Historically, it has been assumed that the chicken  $\alpha$ -globin locus is found in a constitutively open chromatin conformation.<sup>40</sup> Tissue- and stage-specific transcription factors are then needed for the activation of the embryonic  $\pi$  gene in a context in which the rest of the domain remains in an open chromatin structure.<sup>3</sup> In such an embryonic context no stage-specific factors are present for the activation of the adult genes. Once switching between embryonic and adult gene expression occurs, CTCF and its co-factors act as a barrier and a nucleation center to shield the silencing of the embryonic  $\pi$  gene and surrounding genomic regions through DNA methylation in an autonomous way (Fig. 8). As a consequence, the embryonic  $\pi$  gene region is excluded from the active chromatin hub. Based on our model, such a well-delimited silencing is needed to allow the transcriptional activation of the adult  $\alpha^D$  and  $\alpha^A$  genes in later stages of development. A controversy has been raised arguing that there is no silencing in any part of the chicken  $\alpha$ -globin locus.<sup>21</sup> This is based only in the study of active and repressive histone marks but no survey on the DNA methylation status of the regions was assessed.<sup>21</sup> Our results support combinatorial regulatory events which could be similar to those observed for RAR $\beta$ 2 gene expression, where the XPG endonuclease promotes DNA breaks and DNA demethylation at promoters, including CTCF recruitment and looping, which are events that can also be mediated by specific TAFs.<sup>41,42</sup> This attempt to integrate genetic and epigenetic levels of regulation represents a starting point to address more detailed mechanistic questions.

## Materials and Methods

**Cell culture.** RBC primary cultures were obtained from chicken embryos at 5 and 10 days of development (5dRBCs and 10dRBCs, respectively). The cell culture conditions for RBCs and the chicken cell lines (DT40, 6C2, and HD3) were as previously described.<sup>3</sup> To induce erythroid differentiation, HD3 cell line were treated with Iso-H-7 (20  $\mu$ M, Sigma) in HD3 cell medium containing 8% of FBS (Multicell), 2% chicken serum (Gibco), 1% penicillin/streptomycin (Gibco), and 10 mM HEPES pH 8.0 in a 1% CO<sub>2</sub> atmosphere at 42°C for 48 h. HD3 cell transfections were performed using Lipofectamine plus 2000 (Invitrogen) under antibiotic resistant selection in the presence of 0.8 mg/ml of Geneticin (Calbiochem). Transgene expression was evaluated by flow cytometry as described.<sup>3</sup> For drug treatments, Trichostatin-A (TSA, Sigma) and 5-aza-2'-deoxycytidine (5-azadC, Sigma) were used at 2.5 ng/ml and 3  $\mu$ M, respectively.

**Plasmids.** The pG $\alpha^D$ , pG $\alpha^D$ DIG, pG $\Delta$ C1, and pG $\Delta$ C2 plasmids and their associated versions containing the 3'-enh were generated by PCR amplification of the  $\alpha^D$  promoter (Fig. S4), the intergenic region between the  $\pi$  and  $\alpha^D$  genes and the 3'-enh using the specific primers listed in Table S1. The PCR DNA fragments were sub-cloned and restriction enzyme-digested for introduction into the pEGFP-1 (Clontech) vector reporter system.

**Sodium bisulfite DNA conversion and sequencing.** Three  $\mu$ g of genomic DNA from DT40, 6C2, HD3, dif-HD3 cells, 5dRBC, and 10dRBCs were denatured at 95°C for 5 min, chilled on ice, and incubated with 0.3 M NaOH at 37°C for 5 min. Freshly-prepared solutions of sodium bisulfite (Sigma), adjusted to pH 5.0 with NaOH, and hydroquinone (Sigma) were added at final concentrations of 1.7 M and 0.5 mM, respectively. DNA solutions were mixed and incubated at 55°C for 13 h in the dark. Non-reacting bisulfite was removed by column purification (Wizard DNA Clean-Up System; Promega). Purified DNA samples were de-sulfonated with NaOH at a final concentration of 0.3 M, incubated at 37°C for 15 min followed by ethanol precipitation. DNA fragments of interest were PCR-amplified using the modified primers listed in Table S1, then cloned into the pGEM-T Easy system (Promega), and sequenced using SP6 or T7 sequence primers.

**Quantitative RT-PCR.** RNA isolation was performed as described previously.<sup>3</sup> For quantitative RT-PCR the primers used for amplification are listed in Table S1. All primers were intron-spanning. For quantitative real time PCR (qPCR), Platinum Taq DNA polymerase (Invitrogen) was used mixed with SYBR-Green I (Sigma) dissolved in DMSO. The qPCR reactions were performed in a StepOne™ detection system (Applied Biosystems). The relative expression levels of the  $\alpha^D$  gene were determined by using the ribosomal 28S mRNA levels as an endogenous reference and the  $\Delta\Delta$ CT method was used for normalization.

**5-hydroxymethylcytosine DNA immunoprecipitation (hMeDIP).** In order to detect only 5hmC present in the  $\alpha^D$  gene

promoter and not in the surrounding sequences, the genomic DNAs were first BamHI-digested and then used for hMeDIP experiments. The 5-hydroxymethylcytosine DNA immunoprecipitation (hMeDIP) technique was performed using an anti-5hmC antibody (Active Motif; 39769) as described.<sup>43</sup> The technique was validated by Dot-Blot analysis (Fig. S1D). qPCR was performed using the primers listed in Table S1. The 5hmC enrichment was calculated, first by normalizing  $\alpha^D$  promoter amplification values over a 5hmC-depleted region (HI, repetitive region in a heterochromatic context),<sup>3,44</sup> and then over the IgG background value. The calculations were performed as follows:  $Enrichment = \frac{antibody (C_{T_r} - C_{C_c})}{IgG (C_{T_r} - C_{C_c})}$ ; where  $C_{T_r}$  = amplification of the test region, and  $C_{C_c}$  = amplification of the negative control region (HI, Fig. S2A).

**Identification of in vivo CTCF interacting regions in the chicken genome.** The chicken CTCF binding maps previously published were incorporated into the chicken assembly Genome Browser (May 2006, WUGSC 2.1/galGal3 Assembly; UCSC) and the in vivo binding of CTCF was determined at the  $\alpha$ -globin locus.<sup>20</sup> The accession codes for the ChIP-seq data previously published are the following: GEO: GSM691016 for 5dRBCs and GSM691017 for 10dRBCs.<sup>20</sup> The identified CTCF binding sites were validated by quantitative PCR (see below).

**Chromatin immunoprecipitation analysis.** Chromatin was prepared for immunoprecipitation as previously described.<sup>3</sup> The chromatin was immunoprecipitated with anti-CTCF antibody,<sup>45</sup> anti-Acetyl histone H3 antibody (06-599, Millipore), anti-Acetyl histone H4 antibody (06-866, Millipore), anti-H3K4me2 antibody (07-030, Millipore), anti-H2A.Z antibody (ab18263, Abcam), anti-H3K27me3, and anti-H3K9me3 antibodies (kindly supplied by Dr Thomas Jenuwein, IMP-Freiburg), and anti-Normal rabbit IgG (12-370, Millipore). The purified and recovered DNA was assayed by qPCR using the primers listed in Table S1. For the CTCF and open chromatin marks, the enrichment was calculated over a repetitive sequence region in a heterochromatic context (HI in Fig. S2A).<sup>3,46</sup> In the inverse sense, for close chromatin marks, the enrichment was calculated over a constitutive open region (FII in Fig. S2A).<sup>3,46</sup> The enrichment calculations were performed as follows:  $Enrichment = \frac{antibody (C_{T_r} - C_{C_c})}{IgG (C_{T_r} - C_{C_c})}$ ; where  $C_{T_r}$  = amplification of the test region, and  $C_{C_c}$  = amplification of the negative control region (HI or FII; Fig. S2A). Alternatively, we used semi-quantitative radioactive duplex PCR using the same formula described for the enrichment calculations.

**Chromosome conformation capture assay (3C).** The 3C experiments were performed as described with minor modifications.<sup>27</sup> In brief,  $4 \times 10^7$  RBCs were cross-linked with either

1% or 2% formaldehyde for 10 min at room temperature (Fig. S7). The reaction was stopped by adding glycine (125 mM final concentration). The cells were washed with cold-PBS and the nuclei were isolated with incubation in ice-cooled lysis buffer (10 mM Tris (pH 8.0), 10 mM NaCl, 0.2% NP-40, and protease inhibitor mix) for 90 min. The nuclei were dispersed into  $1 \times 10^6$  aliquots and frozen in liquid nitrogen. The 1% formaldehyde cross-linked nuclei were used for the following steps of the protocol (Fig. S7). In order to determine the frequency of interacting genomic fragments at high resolution, a  $1 \times 10^6$  aliquot was digested with 1,000 units of the four-cutter DpnII restriction enzyme (New England Biolabs) overnight. Inactivation of the enzyme was performed by incubation at 65°C for 20 min. The samples were incubated in 7 ml of ligase Buffer (NEB) and 400 units of T4 DNA ligase (NEB) was added at 16°C for at least 4 h. The cross-linking was reversed by incubation at 65°C overnight in the presence of Proteinase K and RNase A (Sigma). The DNA was purified with phenol:chloroform:isoamyl alcohol (Sigma), and precipitated with ethanol and glycogen (Roche). To specifically detect the interaction hybrids, the following Taqman probe was designed over the “anchor” site: 5'-6-FAM/ACTCAGCAC/ZEN/CTGGCAGGTTTAC-3'-IABkFQ (IDT Technologies). The calculations and normalization data are shown in Table S2.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

#### Acknowledgments

We thank Catherine Farrell and Paul Delgado-Olguín for reagents and critical reading of the manuscript. We acknowledge the technical assistance of Fernando Suaste-Olmos and the bioinformatic advice of Rodrigo Arzate-Mejía. This work was supported by the Dirección General de Asuntos del Personal Académico-Universidad Nacional Autónoma de México (IN209403 and IN203811), Consejo Nacional de Ciencia y Tecnología, México (CONACyT; 42653-Q and 128464), and a PhD fellowship from CONACyT and Dirección General de estudios de Posgrado-Universidad Nacional Autónoma de México (DGEP) (C V-Q-207086, C A-C-207081, and YF-G-492191). Additional support was provided by the PhD Graduate Program “Doctorado en Ciencias Biomédicas.”

#### Supplemental Materials

Supplemental materials may be found here: <http://www.landesbioscience.com/journals/epigenetics/article/25472>

#### References

- Ghirlando R, Giles K, Gowher H, Xiao T, Xu Z, Yao H, et al. Chromatin domains, insulators, and the regulation of gene expression. *Biochim Biophys Acta* 2012; 1819:644-51; PMID:22326678; <http://dx.doi.org/10.1016/j.bbagg.2012.01.016>
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012; 485:376-80; PMID:22495300; <http://dx.doi.org/10.1038/nature11082>
- Rincón-Arango H, Guerrero G, Valdes-Quezada C, Recillas-Targa F. Chicken  $\alpha$ -globin switching depends on autonomous silencing of the embryonic  $\pi$  globin gene by epigenetics mechanisms. *J Cell Biochem* 2009; 108:675-87; PMID:19693775; <http://dx.doi.org/10.1002/jcb.22304>
- Stamatoyannopoulos G. Control of globin gene expression during development and erythroid differentiation. *Exp Hematol* 2005; 33:259-71; PMID:15730849; <http://dx.doi.org/10.1016/j.exphem.2004.11.007>
- West AG, Fraser P. Remote control of gene transcription. *Hum Mol Genet* 2005; 14(Spec No 1):R101-11; PMID:15809261; <http://dx.doi.org/10.1093/hmg/ddi104>
- Ragoczy T, Bender MA, Telling A, Byron R, Groudine M. The locus control region is required for association of the murine  $\beta$ -globin locus with engaged transcription factories during erythroid maturation. *Genes Dev* 2006; 20:1447-57; PMID:16705039; <http://dx.doi.org/10.1101/gad.1419506>

7. Furlan-Magaril M, Rebollar E, Guerrero G, Fernández A, Moltó E, González-Buendía E, et al. An insulator embedded in the chicken  $\alpha$ -globin locus regulates chromatin domain configuration and differential gene expression. *Nucleic Acids Res* 2011; 39:89-103; PMID:20813760; <http://dx.doi.org/10.1093/nar/gkq740>
8. Vyas P, Vickers MA, Picketts DJ, Higgs DR. Conservation of position and sequence of a novel, widely expressed gene containing the major human  $\alpha$ -globin regulatory element. *Genomics* 1995; 29:679-89; PMID:8575760; <http://dx.doi.org/10.1006/geno.1995.9951>
9. Flint J, Tufarelli C, Peden J, Clark K, Daniels RJ, Hardison R, et al. Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the  $\alpha$  globin cluster. *Hum Mol Genet* 2001; 10:371-82; PMID:11157800; <http://dx.doi.org/10.1093/hmg/10.4.371>
10. Escamilla-Del-Arenal M, Recillas-Targa F. GATA-1 modulates the chromatin structure and activity of the chicken  $\alpha$ -globin 3' enhancer. *Mol Cell Biol* 2008; 28:575-86; PMID:17984219; <http://dx.doi.org/10.1128/MCB.00943-07>
11. Harju-Baker S, Costa FC, Fedosyuk H, Neades R, Peterson KR. Silencing of Agamma-globin gene expression during adult definitive erythropoiesis mediated by GATA-1-FOG-1-Mi2 complex binding at the -566 GATA site. *Mol Cell Biol* 2008; 28:3101-13; PMID:18347053; <http://dx.doi.org/10.1128/MCB.01858-07>
12. Knezetic JA, Felsenfeld G. Mechanism of developmental regulation of  $\alpha$ ,  $\pi$ , the chicken embryonic  $\alpha$ -globin gene. *Mol Cell Biol* 1993; 13:4632-9; PMID:8336706
13. Zhang Y, Shu J, Si J, Shen L, Estecio MR, Issa JP. Repetitive elements and enforced transcriptional repression co-operate to enhance DNA methylation spreading into a promoter CpG-island. *Nucleic Acids Res* 2012; 40:7257-68; PMID:22600741; <http://dx.doi.org/10.1093/nar/gks429>
14. Recillas-Targa F, de la Rosa-Velázquez IA, Soto-Reyes E. Insulation of tumor suppressor genes by the nuclear factor CTCF. *Biochem Cell Biol* 2011; 89:479-88; PMID:21846316; <http://dx.doi.org/10.1139/o11-031>
15. Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schübeler D. Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet* 2011; 43:1091-7; PMID:21964573; <http://dx.doi.org/10.1038/ng.946>
16. Beug H, von Kirchbach A, Döderlein G, Conscience JF, Graf T. Chicken hematopoietic cells transformed by seven strains of defective avian leukemia viruses display three distinct phenotypes of differentiation. *Cell* 1979; 18:375-90; PMID:227607; [http://dx.doi.org/10.1016/0092-8674\(79\)90057-6](http://dx.doi.org/10.1016/0092-8674(79)90057-6)
17. Gavrillov AA, Razin SV. Spatial configuration of the chicken  $\alpha$ -globin gene domain: immature and active chromatin hubs. *Nucleic Acids Res* 2008; 36:4629-40; PMID:18621783; <http://dx.doi.org/10.1093/nar/gkn429>
18. Wu H, Zhang Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev* 2011; 25:2436-52; PMID:22156206; <http://dx.doi.org/10.1101/gad.179184.111>
19. Ficiz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* 2011; 473:398-402; PMID:21460836; <http://dx.doi.org/10.1038/nature10008>
20. Martín D, Pantoja C, Fernández Miñán A, Valdes-Quezada C, Moltó E, Matesanz F, et al. Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. *Nat Struct Mol Biol* 2011; 18:708-14; PMID:21602820; <http://dx.doi.org/10.1038/nsmb.2059>
21. Ioudinkova ES, Ulianov SV, Bunina D, Iarovaia OV, Gavrillov AA, Razin SV. The inactivation of the  $\pi$  gene in chicken erythroblasts of adult lineage is not mediated by packaging of the embryonic part of the  $\alpha$ -globin gene domain into a repressive heterochromatin-like structure. *Epigenetics* 2011; 6:1481-8; PMID:22139578; <http://dx.doi.org/10.4161/epi.6.12.18215>
22. Klochkov D, Rincón-Arango H, Ioudinkova ES, Valadez-Graham V, Gavrillov A, Recillas-Targa F, et al. A CTCF-dependent silencer located in the differentially methylated area may regulate expression of a housekeeping gene overlapping a tissue-specific gene domain. *Mol Cell Biol* 2006; 26:1589-97; PMID:16478981; <http://dx.doi.org/10.1128/MCB.26.5.1589-1597.2006>
23. Kim A, Dean A. A human globin enhancer causes both discrete and widespread alterations in chromatin structure. *Mol Cell Biol* 2003; 23:8099-109; PMID:14585970; <http://dx.doi.org/10.1128/MCB.23.22.8099-8109.2003>
24. Zhao H, Dean A. An insulator blocks spreading of histone acetylation and interferes with RNA polymerase II transfer between an enhancer and gene. *Nucleic Acids Res* 2004; 32:4903-19; PMID:15371553; <http://dx.doi.org/10.1093/nar/gkh832>
25. Mutskov V, Felsenfeld G. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J* 2004; 23:138-49; PMID:14685282; <http://dx.doi.org/10.1038/sj.emboj.7600013>
26. Holwerda S, de Laat W. Chromatin loops, gene positioning, and gene expression. *Front Genet* 2012; 3:217; PMID:23087710; <http://dx.doi.org/10.3389/fgene.2012.00217>
27. Hagège H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, et al. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2007; 2:1722-33; PMID:17641637; <http://dx.doi.org/10.1038/nprot.2007.243>
28. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science* 2002; 295:1306-11; PMID:11847345; <http://dx.doi.org/10.1126/science.1067799>
29. Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, et al. CTCF mediates long-range chromatin looping and local histone modification in the  $\beta$ -globin locus. *Genes Dev* 2006; 20:2349-54; PMID:16951251; <http://dx.doi.org/10.1101/gad.399506>
30. Soto-Reyes E, González-Barrios R, Cisneros-Soberanis F, Herrera-Goepfert R, Pérez V, Cantú D, et al. Disruption of CTCF at the miR-125b1 locus in gynecological cancers. *BMC Cancer* 2012; 12:40; PMID:22277129; <http://dx.doi.org/10.1186/1471-2407-12-40>
31. Witcher M, Emerson BM. Epigenetic silencing of the p16(INK4a) tumor suppressor is associated with loss of CTCF binding and a chromatin boundary. *Mol Cell* 2009; 34:271-84; PMID:19450526; <http://dx.doi.org/10.1016/j.molcel.2009.04.001>
32. Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Nemes A, et al. Sp1 elements protect a CpG island from *de novo* methylation. *Nature* 1994; 371:435-8; PMID:8090226; <http://dx.doi.org/10.1038/371435a0>
33. Dickson J, Gowher H, Strogantsev R, Gaszner M, Hair A, Felsenfeld G, et al. VEZF1 elements mediate protection from DNA methylation. *PLoS Genet* 2010; 6:e1000804; PMID:20062523; <http://dx.doi.org/10.1371/journal.pgen.1000804>
34. Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, Bell AC, Litt MD, West AG, et al. Position-effect protection and enhancer blocking by the chicken  $\beta$ -globin insulator are separable activities. *Proc Natl Acad Sci U S A* 2002; 99:6883-8; PMID:12011446; <http://dx.doi.org/10.1073/pnas.102179399>
35. Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 2007; 39:457-66; PMID:17334365; <http://dx.doi.org/10.1038/ng1990>
36. Sun Z, Terragni J, Borgaro JG, Liu Y, Yu L, Guan S, et al. High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. *Cell Rep* 2013; 3:567-76; PMID:23352666; <http://dx.doi.org/10.1016/j.celrep.2013.01.001>
37. Guastafierro T, Cecchinelli B, Zampieri M, Reale A, Riggio G, Sthandier O, et al. CCCTC-binding factor activates PARP-1 affecting DNA methylation machinery. *J Biol Chem* 2008; 283:21873-80; PMID:18539602; <http://dx.doi.org/10.1074/jbc.M801170200>
38. Zampieri M, Guastafierro T, Calabrese R, Ciccarone F, Bacalini MG, Reale A, et al. ADP-ribose polymers localized on Ctfc-Parp1-Dnm1 complex prevent methylation of Ctfc target sites. *Biochem J* 2012; 441:645-52; PMID:21985173; <http://dx.doi.org/10.1042/BJ20111417>
39. Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W. The  $\beta$ -globin nuclear compartment in development and erythroid differentiation. *Nat Genet* 2003; 35:190-4; PMID:14517543; <http://dx.doi.org/10.1038/ng1244>
40. Razin SV, Farrell CM, Recillas-Targa F. Genomic domains and regulatory elements operating at the domain level. *Int Rev Cytol* 2003; 226:63-125; PMID:12921236; [http://dx.doi.org/10.1016/S0074-7696\(03\)01002-7](http://dx.doi.org/10.1016/S0074-7696(03)01002-7)
41. Liu Z, Scannell DR, Eisen MB, Tjian R. Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell* 2011; 146:720-31; PMID:21884934; <http://dx.doi.org/10.1016/j.cell.2011.08.005>
42. Le May N, Fradin D, Iltis I, Bounghès P, Egly JM. XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes. *Mol Cell* 2012; 47:622-32; PMID:22771116; <http://dx.doi.org/10.1016/j.molcel.2012.05.050>
43. Mohn F, Weber M, Schübeler D, RoloffTC. Methylated DNA immunoprecipitation (MeDIP). *Methods Mol Biol* 2009; 507:55-64; PMID:18987806; [http://dx.doi.org/10.1007/978-1-59745-522-0\\_5](http://dx.doi.org/10.1007/978-1-59745-522-0_5)
44. Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappaport J, et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 2011; 473:343-8; PMID:21490601; <http://dx.doi.org/10.1038/nature10066>
45. Valadez-Graham V, Razin SV, Recillas-Targa F. CTCF-dependent enhancer blockers at the upstream region of the chicken  $\alpha$ -globin gene domain. *Nucleic Acids Res* 2004; 32:1354-62; PMID:14981153; <http://dx.doi.org/10.1093/nar/gkh301>
46. Litt MD, Simpson M, Recillas-Targa F, Prioleau MN, Felsenfeld G. Transitions in histone acetylation reveal boundaries of three separately regulated neighboring loci. *EMBO J* 2001; 20:2224-35; PMID:11331588; <http://dx.doi.org/10.1093/emboj/20.9.2224>