



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
Maestría y Doctorado en Ciencias Bioquímicas

“Efecto de un RNA largo no-codificante en la expresión de los genes α -globina de pollo”

T E S I S
QUE PARA OPTAR POR EL GRADO DE:
Maestra en Ciencias

PRESENTA:
LCG. Ana Martha Yael Fonseca Guzmán

Dr. Félix Recillas Targa
[Instituto de Fisiología Celular](#)
MIEMBROS DEL COMITÉ TUTOR
Dr. Jesús Chimal Monroy
[Instituto de Investigaciones Biomédicas](#)
Dr. José Luis Reyes Taboada
[Instituto de Biotecnología](#)

MÉXICO, D. F., Agosto de 2014



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.

Reconocimientos académicos

El presente trabajo fue elaborado 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 con los apoyos de DGAP-PAPIIT, UNAM (IN203811 e IN201114) y CONACyT (128464 y 220503).

Los estudios de maestría de la LCG. Ana Martha Yael Fonseca Guzmán se realizaron con el apoyo de la beca de maestría del CONACyT.

Se agradece y reconoce al comité tutor que asesoró esta tesis, el cual estuvo constituido por:

Dr. Félix Recillas Targa	Instituto de Fisiología Celular, UNAM
Dr. Jesús Chimal Monroy	Instituto de Investigaciones Biomédicas, UNAM
Dr. José Luis Reyes Taboada	Instituto de Biotecnología, UNAM

Se agradece el apoyo y la excelente asesoría técnica de la Biol. Georgina Guerrero Avendaño y el Dr. Fernando Suaste Olmos. Así mismo, se reconoce el trabajo de la laboratorista Gianelli Cortés González.

Adicionalmente se reconoce la asesoría académica de todos y cada uno de los miembros del laboratorio del Dr. Félix Recillas Targa, en particular del Biol. Christian Valdes Quezada y el Biol. Cristian Gabriel Oliverio Arriaga Canon.

Se reconoce también el apoyo de la Unidad de Biología Molecular del Instituto de Fisiología Celular en las secuenciaciones de plásmidos realizadas en este proyecto, y también se agradece a la asistente de procesos del posgrado Leticia García Gutierrez por su disposición y ayuda en los trámites administrativos.

Finalmente se agradece a los miembros del Jurado que revisaron y enriquecieron con sus observaciones este trabajo:

Dr. Ignacio Camacho Arroyo (Presidente)	Facultad de Química
Dra. Rosa Estela Navarro González	Instituto de Fisiología Celular
Dr. Enrique Merino Pérez	Instituto de Biotecnología
Dra. Marina Macías Silva	Instituto de Fisiología Celular
Dr. Juan Miranda Ríos (Secretario)	Instituto de Investigaciones Biomédicas

Agradecimientos

Muchas gracias al Dr. Félix Recillas, antes que nada por haberme aceptado en su laboratorio y haberme dado la oportunidad de formar parte de su maravilloso equipo de trabajo. Gracias por los consejos, e incluso los jalones de oreja. Su compromiso con la ciencia es innegable, y es un ejemplo de perseverancia, disciplina y esfuerzo para todos nosotros.

Muchas gracias a Geito, por ser “la mamá de los pollitos” del laboratorio. Porque no importa lo mal que nos vaya, o los problemas que tengamos, siempre estás ahí para escucharnos y brindarnos palabras de aliento.

Muchas gracias Giane, porque el labo es un lugar mejor gracias a ti. ¡Eres una persona muy cálida y atenta! Muchas gracias Fer Suaste, por las discusiones, consejos, pláticas e incluso por las partiditas de básquet... pero esto no se queda así, algún día pediré mi revancha y ahora si espero ganarte.

A todos mis compañeros del 122 Nte. porque honestamente encontré un amigo en todos y cada uno de ustedes, gracias por hacer la ciencia más llevadera y divertida. De todos aprendí muchas cosas (y no sólo de ciencia). A los que me enseñaron, que en realidad fueron todos, pero en particular a Cristian y a Christiansito, por ofrecerme su tiempo y sus conocimientos, pero sobre todo su amistad, ¡mil gracias chicos! No tengo cómo agradecerles todo lo que hicieron por mí. Mil gracias Meche por tu amistad que fue una de las primeras, y por los buenos consejos. La verdad es que ser el nuevo siempre es difícil, pero me diste ánimos para seguir adelante. A Charis, Erandí, Edgar, Paco, Rodrigo, Martín, chicos ustedes son el alma de la fiesta completamente!!! Gracias por los buenos momentos y las risas, porque tener compañeros como ustedes no tiene precio, y definitivamente hace de cualquier sitio un lugar mucho más agradable. ¡Mil gracias Jesús! Porque tú eres parte también de la familia Recillas, y eres uno de los compañeros más solidarios que uno puede tener, ¡gracias por salvar mi vida en más de una ocasión!

Muchas gracias Estela, David, Carla y Mariana, por todos los buenos momentos que compartimos y por su sincera amistad. Espero que a dónde quiera que vayamos no nos perdamos la pista, yo sé que les irá super bien hagan lo que hagan, ¡porque son unas excelentes personas! Nunca se olviden de mí, porque al menos yo siempre los recordaré.

A todos los nuevos: Karin, Christina, Hober, Josué, Shana, ¡les deseo todo el éxito del mundo! Nunca se desanimen por más difíciles que estén las cosas, ¡si se puede! Están en un súper grupo y ustedes son unas súper personas, ¡así que todo estará bien!

¡Mil gracias a todos y cada uno de ustedes! ¡Los quiero mucho!

¡Gracias a mi familia porque son el motor que me hace no rendirme!

Gracias mamá porque eres el ejemplo de lucha más grande que tengo, porque no importa lo que yo necesite, tú siempre has estado ahí para ayudarme y para impulsarme a alcanzar mis sueños.

Gracias papá por escucharme y entenderme, porque siempre has tenido fe en mí, y eso me ha hecho tener más fe en mí misma. Gracias por reconfortarme y por estar al pendiente.

Gracias a mis hermanas, Claudia e Itzel, porque a pesar de ser la mayor ustedes han cuidado mucho de mí. A los peques: Juanito y Melanie, gracias por las risas y los juegos, porque no importa lo mal que me sienta, después de verlos sonreír todo se aligera.

A mis abuelitas, Reina y Coco, porque nada me hubiera dado más gusto que compartir este momento con ustedes. Gracias, porque invariablemente en los mejores recuerdos aparecen ustedes. ¡Las llevaré por siempre en mi corazón!

Gracias papá bito por todo el apoyo, por ser el pilar de esta familia tan grande y tan bonita, ¡te quiero mucho! Gracias Nina, Tía Ady y Paty, Tía Adriana: por todas las risas y los buenos momentos, por echarme tantas porras y creer en mí.

A Zaira, Monse, Jorge, Katy...los amigos son los hermanos que elegimos y yo no pude haber elegido mejor!! Los quiero!!

¡¡¡¡Gracias Jonas!!!! Porque tú sabes que este logro en gran parte te lo debo a ti. Gracias por la comprensión y el apoyo. Por creer siempre en mí, por protegerme siempre. Por inspirarme a superarme y a ser mejor persona. Porque no tengo palabras para expresarte mi gratitud, y todo lo que siento por ti... ¡te amo!

Contenido

Reconocimientos académicos	2
Agradecimientos.....	3
RESUMEN	7
ABREVIATURAS	8
1. INTRODUCCIÓN	9
Clasificación e inferencia de función de RNAs largos no-codificantes	14
Mecanismos de acción propuestos para los lncRNAs	18
1.1. ANTECEDENTES GENERALES	42
1.2. ANTECEDENTES PARTICULARES	52
Caracterización de la transcripción a lo largo del dominio α -globina de pollo	53
Determinación del potencial codificante y de un promotor específico del linaje eritroide ..	56
2. PLANTEAMIENTO DEL PROBLEMA	61
3. HIPÓTESIS.....	62
4. OBJETIVO GENERAL	62
OBJETIVOS PARTICULARES	62
5. MATERIALES Y MÉTODOS.....	63
Cultivo celular	63
Plásmidos	63
Inmunoprecipitación de la cromatina (ChIP).....	64
Extracción de RNA y RT-PCR cuantitativa (qRT-PCR).....	65
Transfecciones en HD3	69
Diferenciación de células HD3.....	69
Tabla de secuencias de oligonucleótidos.....	70
6. RESULTADOS	73
6.1. Expresión de lncRNA- α GT y su correlación con la estructura local de la cromatina durante el desarrollo del pollo.....	73
6.2. Determinación de la función del lncRNA- α GT.....	80
6.2.1. Ensayo de pérdida de función del lncRNA- α GT en células de pollo HD3 diferenciadas	80

6.2.2. Ensayo de ganancia de función del lncRNA- α GT en células HD3.....	86
7. DISCUSIÓN	94
8. CONCLUSIÓN.....	102
9. PERSPECTIVAS.....	102
10.REFERENCIAS	104
11.ANEXOS.....	121

RESUMEN

Actualmente se sabe que la mayor parte del genoma se transcribe, dando origen a un gran número de transcritos no-codificantes. Dentro de estos transcritos sobresalen aquellos genéricamente conocidos como RNAs largos no-codificantes o lncRNAs (por sus siglas en inglés: *long noncoding RNAs*). En los últimos años, la caracterización funcional de este tipo de transcritos ha demostrado que son importantes en la fisiología celular, en particular al nivel de la regulación de la expresión génica. En nuestro grupo de investigación se ha caracterizado previamente un lncRNA, al cual se llamó lncRNA- α GT (*long noncoding RNA α -globin transcript*), que se extiende a lo largo de 23 kilobases sobre el dominio α -globina de pollo. El objetivo de este proyecto de maestría fue determinar si esta molécula es funcional dentro del contexto del dominio α -globina de pollo. Se observó mediante qRT-PCR que la expresión del lncRNA- α GT es mayor en los eritrocitos de pollo de 10 días post-fecundación con respecto a etapas del desarrollo anteriores, y que estos niveles de expresión correlacionan positivamente con el enriquecimiento en la marca de histonas H3K4me3 observado en el promotor del lncRNA- α GT en diferentes etapas del desarrollo. Además, mediante un experimento de pérdida de función en la línea celular eritroide HD3 se demostró que el lncRNA- α GT contribuye a la activación transcripcional del gen de expresión adulta α^D durante la diferenciación eritroide. Estos resultados sugieren que lncRNA- α GT es un elemento importante en la regulación diferencial de los genes α -globina de pollo, en particular a nivel de la transición de expresión embrionaria a adulta.

ABREVIATURAS

cDNA- DNA complementario

lncRNA- RNA largo no-codificante

mRNA- RNA mensajero

shRNA- RNA pequeño con estructura tallo asa

siRNA- RNA pequeño de interferencia

H3K4me3- Trimetilación de la lisina 4 de la histona H3

H3K27me3- Trimetilación de la lisina 27 de la histona H3

H3K36me3- Trimetilación de la lisina 36 de la histona H3

ChIP- Inmunoprecipitación de cromatina

qPCR- PCR cuantitativa

RT- Transcripción reversa

qRT-PCR- Transcripción reversa analizada mediante PCR cuantitativa

RNA-seq- Secuenciación de RNA

1. INTRODUCCIÓN

Durante mucho tiempo, el atributo principal conferido al núcleo celular fue el contener a un universo de genes que constituyen la base de la herencia y diversos procesos bioquímicos que perpetúan la vida. Por lo tanto, conocer la cantidad de genes que conforman el genoma, sus productos potenciales, así como sus mecanismos de regulación, se volvió una prioridad. Con la secuenciación del genoma humano (Consortio Internacional de Secuenciación del Genoma Humano 2004) se hizo evidente que la complejidad de los organismos vivos no podía explicarse simplemente a través del número de genes contenidos en cada especie (Wilusz et al. 2009). El dogma central de la biología molecular estableció que un gen se transcribe produciendo un mRNA (por sus siglas en inglés: “messenger RNA”), y este a su vez es traducido en una proteína (Crick, 1970). Sin embargo, actualmente, se sabe que la porción de genes codificantes ocupa únicamente 1.5% del genoma humano (Wang et al. 2011), lo que llevó a pensar durante mucho tiempo que el resto de las secuencias no-codificantes carecían de función, denominando a estas secuencias “DNA basura” (Willingham y Gingeras, 2006).

Sin embargo, estudios globales del estado de transcripción celular sugieren que la mayor parte del genoma eucariota se transcribe (Bertone et al. 2004; Carninci et al. 2005). Las primeras observaciones de transcripción ubicua en el genoma, provenientes de análisis de microarreglos (técnica utilizada para analizar la expresión de genes), sugirieron que el número de genes de RNA no-codificantes era similar al número de genes codificantes (mRNA/proteínas) (Kapranov et al. 2002; Rinn et al. 2003). Estos hallazgos preliminares se tomaron con mucho escepticismo por parte de la comunidad científica

debido a los inconvenientes inherentes a la metodología de microarreglo, y durante mucho tiempo hubo debates controversiales sobre si se trataba de ruido transcripcional o de transcritos con funciones específicas (Ponjavic et al. 2007). En este caso el ruido transcripcional es toda aquella transcripción que surge como un producto inespecífico de la RNA polimerasa a partir de regiones del genoma que no contienen elementos funcionales (genes), y por ello se asume que esta transcripción no es funcional. El ruido transcripcional tiene por tanto un componente estocástico, y la transcripción que se observa en este fenómeno es en general varios niveles de magnitud inferior a la transcripción asociada con promotores o inicios de la transcripción fidedignos.

Innovaciones recientes en las técnicas de secuenciación masiva (RNA-seq) y el mejoramiento de las herramientas computacionales han permitido la realización de análisis transcriptómicos más finos que corroboraron las observaciones iniciales de transcripción ubicua (Djebali et al. 2012), concluyendo que el genoma está constituido por una proporción de RNAs no-codificantes mucho mayor de lo antes pensado.

Dado que la mayor parte del genoma se transcribe, dando origen a una gran cantidad de RNAs no-codificantes, se han impulsado muchos estudios con la finalidad de entender el significado de este fenómeno transcripcional. Aunque no se espera que todas las moléculas de RNA producidas a partir del genoma sean funcionales (Guttman y Rinn, 2012), en la literatura existen estudios que demuestran que varios de ellos son de relevancia a nivel celular (Penny et al. 1996; Martianov et al. 2007; Rinn et al. 2007; Pandey et al. 2008; Bartel, 2009; Wang et al. 2011; Rinn y Chang, 2012).

En un intento por clasificar aquellos nuevos transcritos no-codificantes provenientes de estudios globales del estado de transcripción celular, se hizo una distinción inicial entre RNAs pequeños y RNAs largos. Dentro de la clasificación de RNAs pequeños encontramos a los miRNAs (*microRNAs*), siRNAs (*small interfering RNAs*) y piRNAs (*PIWI-interacting RNAs*), todos ellos moléculas involucradas en la regulación génica tanto a nivel transcripcional como post-transcripcional (Bartel, 2009; Moazed, 2009; Siomi et al. 2011). Por otra parte, los RNAs largos, genéricamente llamados lncRNAs (por sus siglas en inglés: *long noncoding RNAs*), se han definido convencionalmente como transcritos de un tamaño mayor a 200 nucleótidos (nt) que ejercen diversas funciones reguladoras dentro de la célula (Ponting et al. 2009; Mercer y Mattick, 2013), siendo los más largos que se han descrito hasta el momento de hasta 100 kb (Seidl et al. 2006; Pandey et al. 2008).

Los lncRNAs, como su nombre lo indica, no poseen la capacidad de codificar para proteínas, sin embargo, comparten muchas características bioquímicas y estructurales con los mRNAs codificantes: son transcritos por la RNA polimerasa II (RNA pol II), un gran porcentaje de ellos presenta 5'-cap y se poliadenilan en su extremo 3', tienen una estructura exón-intrón, es decir, son procesados post-transcripcionalmente por la maquinaria de *splicing* (Guttman et al. 2009; Cabili et al. 2011; Rinn y Chang, 2012, Derrien et al. 2012). Estudios funcionales recientes han permitido concluir que los lncRNAs participan en la regulación de diversos procesos biológicos importantes como: el mantenimiento de la pluripotencia de células troncales (Loewer et al. 2010; Guttman et al. 2011), regulación del desarrollo a través de la diferenciación celular (Dinger et al. 2008;

Bertani et al. 2011), regulación del metabolismo (Mercer et al. 2009), así como del ciclo celular (Huarte et al. 2010; Hung, et al. 2011), compensación de dosis en organismos con determinación sexual cromosomal (Nagano et al. 2008; Wutz, 2013), y hasta enfermedades como por ejemplo, cáncer (Gupta et al. 2010). Todos estos estudios, y otros que se mencionarán en detalle a continuación, muestran la importancia de los lncRNAs en la fisiología celular, en particular, al nivel de la regulación génica global.

Dada la relevancia de esta nueva clase de genes no-codificantes, se iniciaron esfuerzos en la anotación de lncRNAs dentro del genoma con la finalidad de facilitar su futuro estudio y caracterización funcional. Una de las primeras herramientas utilizadas en la anotación de lncRNAs hizo uso del análisis de la estructura de la cromatina, en particular de las modificaciones post-traduccionales de histonas, para la predicción de unidades transcripcionales (**Figura 1**). Se sabe que los genes activamente transcritos por la RNA Pol II presentan un enriquecimiento en la trimetilación de la lisina 4 en la histona H3 (H3K4me3) en su promotor, así como en la trimetilación de la lisina 36 en la histona H3 (H3K36me3) a lo largo del cuerpo del gen (Mikkelsen et al. 2007). Esta característica de la estructura de la cromatina se conoce como “la firma K4-K36”. Mediante *ChIP-Seq* (*Chromatin Immunoprecipitation*, inmunoprecipitación de la cromatina acoplada con secuenciación masiva) se buscó la presencia de la firma K4-K36 en los genomas de ratón y humano fuera de las regiones previamente anotadas como genes codificantes (**Figura 1**), y mediante microarreglos de DNA se corroboró cuál de estas regiones se transcribía (Guttman et al. 2009). Como la naturaleza de los transcritos descritos en este estudio es intergénica se les llamó lincRNAs (*long intergenic noncoding RNAs*). Este trabajo fue uno

de los primeros en realizar una anotación genómica de transcritos no-codificantes, aunque cabe resaltar que su enfoque estuvo sesgado a las zonas intergénicas. Estudios posteriores, utilizando RNA-seq, ampliaron el catálogo de lncRNAs conocidos, demostrando que la complejidad del transcriptoma es mucho mayor de lo antes propuesto (**Figura 1**) (Guttman et al. 2010; Garber et al. 2011; Djebali et al. 2012).

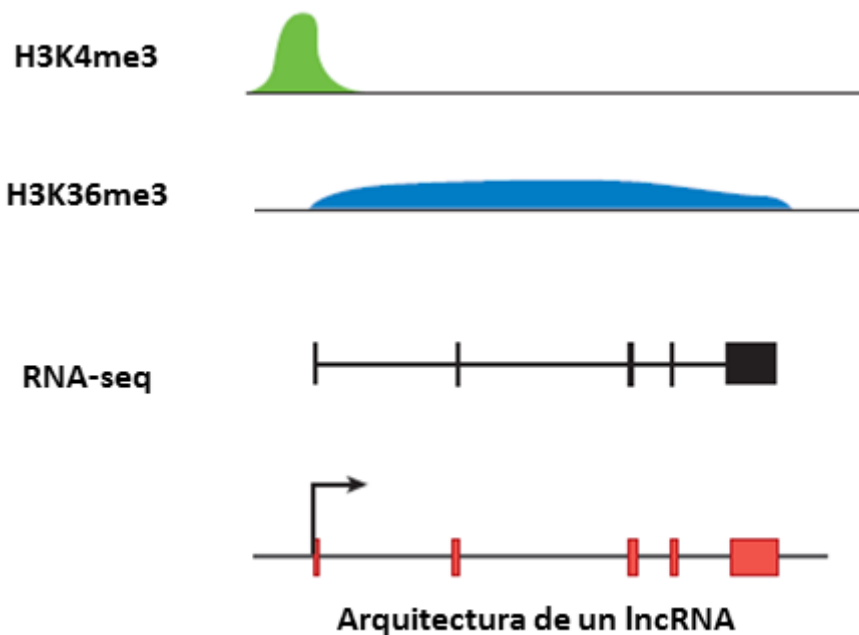


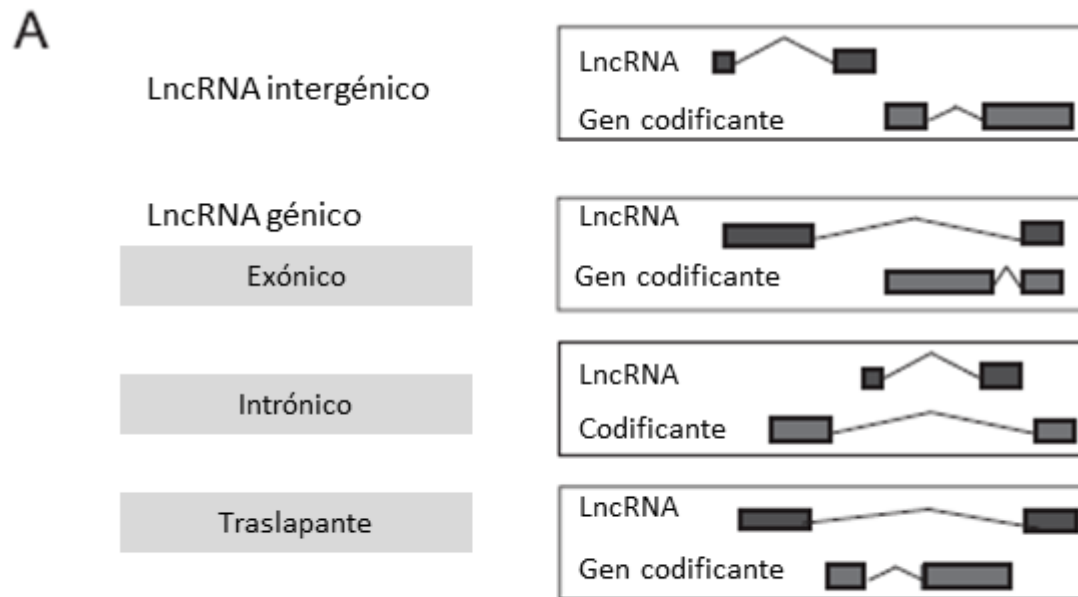
Figura 1.- Descubrimiento y anotación global de lncRNAs en el genoma. En esta figura se dan ejemplos de las técnicas empleadas en la anotación genómica de los lncRNAs, tales como el mapeo de las marcas de cromatina H3K4me3 (inicio de la transcripción) y H3K36me3 (elongación transcripcional), así como secuenciación masiva del transcriptoma (*RNA-seq*, por sus siglas en inglés: *RNA sequencing*). El RNA-seq ayuda de manera directa a definir la estructura madura o procesada (estructura exónica), e incluso las diferentes isoformas, de un lncRNA (Tomada y modificada a partir de Rinn y Chang 2012)

Otra característica de los lncRNAs es que se expresan en muy bajos niveles con respecto a los genes codificantes (Ravasi et al. 2006; Dinger et al 2009; Ponting et al. 2009, Cabili et al. 2011, Derrien et al. 2012), lo cual fue tomado para muchos como un argumento en favor de la teoría del ruido transcripcional. Además, la secuencia nucleotídica de los

lncRNAs parece estar poco conservada (Wang et al. 2004; Pang et al. 2006), lo que apoya la teoría de que se trata de elementos que evolucionan de manera rápida, y por lo tanto podrían tener una influencia en las diferencias biológicas observadas entre especies (Mattick, 2007; Kutter et al. 2012). Estas características (bajos niveles de expresión y secuencias poco conservadas) propias de los lncRNAs dificultaron su descubrimiento y caracterización inicial, ocasionando que la transcripción no-codificante fuera un fenómeno poco tomado en consideración de manera previa al desarrollo del RNA-seq. Los métodos descritos anteriormente se utilizan de manera conjunta en la anotación de lncRNAs (**Figura 1**). Por un lado, la búsqueda de la firma K4-K36 en el genoma permite identificar loci transcripcionalmente activos; por otro lado, el RNA-seq facilita una detección más fina y sensible, no solamente de diversas isoformas provenientes de un mismo locus, sino también de aquellos lncRNAs poco abundantes (Rinn y Chang, 2012).

Clasificación e inferencia de función de RNAs largos no-codificantes

Gracias a todos los esfuerzos realizados en la caracterización y anotación global de los lncRNAs, se ha propuesto una clasificación de los mismos basada en su localización relativa con respecto al gen codificante más cercano (Derrien et al. 2012). En primer lugar, un lncRNA puede ser génico (es decir, que traslapa con otros genes) o intergénico con respecto a los genes codificantes previamente anotados (**Figura 2A**). Dependiendo de la ubicación del inicio de su transcripción y de la orientación de la misma, puede ser divergente/convergente o sentido/antisentido (**Figura 2B**



B

Transcritos de lncRNAs en la base de datos GENCODE (14,880)								
Intergénicos (9,518)			Génicos (5,362)					
Misma orientación	Convergente	Divergente	Exónico (2,411)		Intrónico (2,784)		Traslapante (167)	
4,165	1,937	3,416	S	AS	S	AS	S	AS
			NA	2,411	563	2,221	52	115

Figura 2.-Anotación manual y clasificación de los lncRNAs del genoma humano. (A) Clasificación de lncRNAs basada en la intersección con el gen codificante más cercano. (B) Número de transcritos lncRNAs identificados por categoría. (S), Transcripción en dirección sentido; (AS), transcripción en dirección antisentido con respecto al gen codificante con el que traslapa (Tomada y modificada de Derrien et al. 2012).

Esta clasificación se hizo a partir del catálogo de lncRNAs más completo del genoma humano hasta el momento realizado por el Consorcio GENCODE (proyecto ENCODE, Derrien et al. 2012). Dicho catálogo contiene 9,277 genes de lncRNAs, anotados manualmente, a partir de los cuales se generan 14,880 diferentes transcritos (**Figura 2B**).

Por tanto, la transcripción ubicua del genoma produce una población compleja de

transcritos, de los cuales un gran porcentaje son lncRNAs.

A pesar de que en los últimos años ha sido evidente que el genoma contiene miles de lncRNAs, son pocos los que se han caracterizado funcionalmente. En algunos de los genomas de los organismos modelo (humano, ratón) se cuenta ya con anotaciones y catálogos detallados de lncRNAs, sin embargo, el gran reto en los subsecuentes años será determinar si la totalidad de estas moléculas son funcionales, así como la descripción de sus respectivos mecanismos de acción. De los estudios masivos de expresión génica, se dedujo que los lncRNAs presentan patrones de expresión altamente tejido-específico (Cabili et al. 2011, Derrien et al. 2012), lo que les ha valido ser considerados como elementos muy importantes en el desarrollo y la diferenciación celular. Uno de los primeros métodos bioinformáticos utilizados en la inferencia de las posibles funciones de un lncRNA es la “culpabilidad por asociación” (del inglés *“guilt by association”*). Este método ayuda a realizar hipótesis sobre las posibles funciones de un lncRNA con base en su co-expresión con genes codificantes (Guttman et al. 2009). Mediante este método se realizan correlaciones de expresión entre un lncRNA dado y un grupo de genes codificantes, asumiendo que la observación de una correlación positiva en los patrones de expresión se debe a una presunta co-regulación. Posteriormente, y con base en las funciones conocidas del grupo de proteínas co-expresadas junto con el lncRNA, se generan hipótesis sobre sus funciones potenciales, asociándolo con una o varias rutas metabólicas. Aunque la mera observación de estas correlaciones no es una prueba suficiente y directa de que el lncRNA en cuestión efectivamente tiene una función en estos procesos, este enfoque ha sido empleado en la predicción funcional de cientos de

lncRNAs (Huarte et al. 2010; Hung et al. 2011).

De manera conjunta, uno de los pasos importantes en la caracterización funcional de un gen es la realización de experimentos de perturbación, tales como experimentos de pérdida y ganancia de función, y el análisis subsecuente del fenotipo ocasionado. Sin embargo, este tipo de experimento puede tener la desventaja de que el fenotipo asociado no siempre resulta evidente (Willingham et al. 2005). Las deducciones indirectas sobre la función de un lncRNA realizadas mediante “culpabilidad por asociación” dan una idea general sobre qué fenotipos observar en los experimentos de perturbación de función, y en la mayoría de los casos sirven para generar hipótesis más directas sobre las posibles funciones de un lncRNA, ayudando en la planeación de experimentos de manera más directa (Guttman y Rinn, 2012).

El silenciamiento por RNA de doble cadena o RNA de interferencia (*RNAi*, por sus siglas en inglés: *RNA interference*) ha sido la técnica más utilizada en los experimentos de pérdida de función de lncRNAs. La realización del experimento de pérdida de función puede llevarse a cabo mediante un enfoque más global que permita analizar los efectos a gran escala, es decir, monitorear los cambios en los perfiles de expresión global mediante RNA-seq. Este tipo de estudios son los que precisamente han llevado a sugerir que algunos lncRNAs tienen un impacto global en la regulación génica (Guttman et al. 2011, regulación en *trans*, véase adelante). A continuación se expondrán diferentes tipos de mecanismos de lncRNAs propuestos a partir de diversos estudios y caracterizaciones funcionales, tanto a nivel global, así como casos particulares.

Mecanismos de acción propuestos para los lncRNAs

Aunque se ha podido determinar que muchos lncRNAs son importantes en la regulación génica, hasta hace poco se desconocían los mecanismos mediante los cuales estas moléculas llevaban a cabo su función reguladora. En la actualidad, podemos decir que el número de lncRNAs que se han estudiado y caracterizado funcionalmente es mínimo con respecto a los miles de transcritos no-codificantes presentes en el genoma (**Figura 2**). Sin embargo, este número limitado de estudios va en aumento día a día debido a la necesidad de conocer hasta qué grado este fenómeno de transcripción no-codificante ubicua es relevante en la fisiología celular.

En primer lugar, el conjunto de genes regulados por un lncRNA determinan el alcance de su función. Se define como un regulador en *cis* a aquel que ejerce su función sobre el/los gen(es) vecino(s) presente(s) en el mismo alelo a partir del cual el lncRNA se transcribe. Se asume que todos aquellos lncRNAs que no cumplen este criterio funcionan en *trans*, es decir, regulando cualquier otro conjunto de loci dentro del genoma (Guttman y Rinn, 2012). Es importante hacer notar que normalmente suele haber una co-expresión de los lncRNAs y sus genes vecinos, asumiendo mediante “culpabilidad por asociación” que los lncRNAs en cuestión podrían estar actuando en *cis* (Ebisuya et al. 2008; Ponjavic et al. 2009, Cabili et al. 2011). Sin embargo, en algunos casos se ha demostrado que esta correlación en la expresión se debe a la utilización de elementos regulatorios comunes, y no necesariamente a que el lncRNA esté funcionando en *cis* sobre sus genes vecinos (Guttman et al. 2009; Huarte et al. 2010). Por lo tanto, uno de los experimentos más utilizados para determinar el tipo de función de los lncRNAs es el experimento de pérdida

de función, el cual permite distinguir un mecanismo en *cis* de un mecanismo en *trans*. Si el experimento de pérdida de función de un lncRNA afecta los patrones de expresión normalmente observados para los genes vecinos, es muy probable que el lncRNA en cuestión sea un regulador en *cis* (Sleutels et al. 2002; Pandey et al. 2008; Koerner et al. 2009; Bertani et al. 2011). Por otro lado, si este mismo experimento no permite identificar un efecto sobre los genes vecinos, y por el contrario, muestra efectos en otros genes, el lncRNA en cuestión podría ser un regulador potencial en *trans* (Rinn et al. 2007; Khalil et al. 2009; Guttman et al. 2011). Adicionalmente, se han desarrollado métodos que permiten mapear los sitios de unión de un lncRNA en el genoma (Chu et al. 2011; Simon et al. 2011), lo cual se presenta como una herramienta útil en futuras caracterizaciones de lncRNAs para distinguir aquellos que funcionan en *cis* de los que funcionan en *trans*. Por lo tanto, los resultados obtenidos en este tipo de experimentos constituyen un buen inicio en la caracterización de los mecanismos de acción de un lncRNA y proveen evidencias que permiten la elaboración de nuevas hipótesis de trabajo.

A pesar de que sólo se conoce la función de unas cuantas decenas de lncRNAs, estos estudios han permitido proponer diversos mecanismos de acción que parecen estar conservados entre especies, lo cual indica la relevancia de los lncRNAs en la regulación génica. En particular, se sabe que muchas proteínas poseen dominios específicos que les conceden la capacidad de interactuar con RNA, permitiendo así la formación de complejos ribonucleoproteicos necesarios en la correcta regulación transcripcional (Maison et al. 2002; Wutz et al. 2002; Rinn et al. 2007; Guttman et al. 2011; Khalil et al. 2009). En varios casos, la función reguladora del lncRNA está directamente ligada al tipo de proteínas con

las que interactúa, es decir, un lncRNA regula, por ejemplo, positivamente la expresión de un gen o de un grupo de genes asociándose con proteínas o complejos proteicos involucrados en la activación transcripcional (Bertani et al. 2011; Wang et al. 2011). Se ha reportado que los lncRNAs participan tanto en el silenciamiento como en la activación de la expresión génica, asociándose con proteínas o complejos proteicos represores o activadores, respectivamente. Las particularidades de estas interacciones RNA-proteína han llevado a proponer diferentes mecanismos de acción para los lncRNAs, mismos que se explicarán más adelante.

Además, es importante señalar que algunas proteínas poseen dominios de unión a DNA y RNA (Searles et al. 2000; Lu et al. 2003; Brown 2005). Por ejemplo, se reportó recientemente que el factor multifuncional CTCF, ampliamente reconocido como una proteína de unión a DNA involucrada en diversos procesos que van desde la regulación transcripcional, *splicing* alternativo, impronta genómica, inactivación del cromosoma X, entre otros (Phillips y Corces, 2009; Holwerda y de Laat, 2013), también es capaz de unir RNA mediante su dominio RBR (por sus siglas en inglés: *RNA-binding region*) (Saldaña-Meyer et al. 2014). En este mismo trabajo se describe que CTCF regula la activación transcripcional del locus *p53* a dos niveles: (i) mediante su unión directa al DNA en la región del promotor de *p53*, y (ii) mediante la unión al RNA antisentido a *p53*, *Wrap53*. Además, se demostró que la unión a RNA a través del dominio RBR es necesaria en la formación de multímeros de CTCF, lo cual se prevee será de gran relevancia en el estudio de la formación de asas cromosomales (*loops*) mediada por CTCF, y a su vez, presupone un papel importante de los lncRNAs en la estructuración del genoma (Saldaña-Meyer et al,

2014).

Cabe resaltar que algunos trabajos han propuesto que el RNA no sólo interactúa con proteínas, sino también con otros ácidos nucleicos (con otros RNAs, y en particular con DNA) (Rinn y Chang, 2012), proponiendo que estas interacciones también podrían ser importantes en la función de un lncRNA, aunque la dificultad de su estudio *in vivo* representa un obstáculo en su descripción global (Martianov et al. 2007; Schmitz et al. 2010). Todas estas observaciones han llevado a sugerir que el RNA podría estar funcionando como una interfaz dinámica entre el DNA y complejos proteicos, cuya finalidad funcional es la correcta expresión transcripcional (Chu et al. 2011; Simon et al. 2011).

Se ha propuesto que la formación de estructuras secundarias particulares en los lncRNAs son la base principal de su interacción con las proteínas, por ello algunos grupos de investigación han comenzado a enfocarse en el desarrollo de nuevas técnicas que permitan analizar las estructuras secundarias del RNA a nivel global de una manera más confiable (Kertesz et al. 2010; Underwood et al. 2010; Wan et al. 2011). La capacidad de los lncRNAs para interactuar con proteínas les confiere diversas funciones reguladoras. A continuación se describirán algunos de los modelos de funcionamiento de lncRNAs propuestos hasta el momento:

Modelo alostérico

Los lncRNAs pueden modular la actividad reguladora de ciertas proteínas o complejos proteicos a través de la interacción física (**Figura 3**). Por ejemplo, el lncRNA SRA (por sus

siglas en inglés: *steroid receptor RNA activator*) tiene la capacidad de interactuar con la proteína multifuncional CTCF, potenciando su actividad como elemento barrera (Yao et al. 2010). Otro ejemplo de este mecanismo de acción es el lncRNA NRON, cuya interacción con el factor de transcripción NFAT (*nuclear factor of activated T-cells*) evita su acumulación dentro del núcleo y por tanto, interviene con su función transcripcional (Willingham et al. 2005).

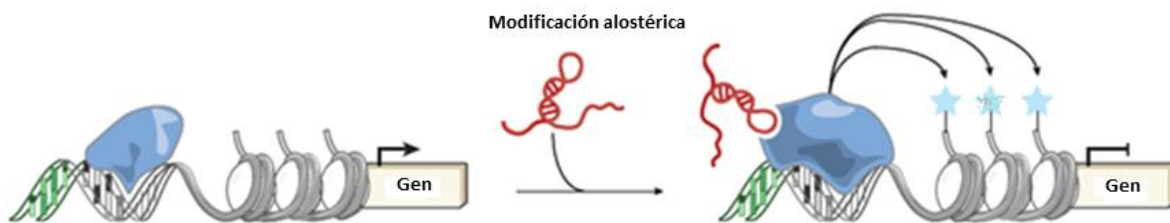


Figura 3.-Modelo del mecanismo de acción alostérico. En este modelo se propone que una proteína reguladora cambia su configuración alostérica, y por lo tanto modifica su actividad, cuando interactúa físicamente con un lncRNA (Tomada de Guttman y Rinn, 2012).

Mecanismos de señuelo o competencia endógena

Este tipo de lncRNAs, mejor conocidos como ceRNAs (*competing endogenous RNAs*), funcionan titulando o “secuestrando” a otros elementos reguladores de manera competitiva, tales como proteínas o miRNAs, y por tanto interfiriendo con su función (**Figura 4A**). Por ejemplo, el lncRNA MALAT1 interactúa con factores de procesamiento post-transcripcional ricos en residuos de serinas/argininas (proteínas SR), secuestrándolas en un compartimento particular del núcleo celular, lo cual tiene un impacto directo en el proceso de *splicing* alternativo (Tripathi et al. 2010). Por otro lado, se ha reportado que varios lncRNAs tienen la capacidad de hibridar de manera competitiva con miRNAs, lo cual

tiene una influencia directa en los niveles intracelulares de miRNAs, así como en los niveles de sus respectivos mRNAs blanco. A este último tipo de lncRNAs se les denomina “esponjas de miRNAs” (Cesana et al. 2011). Un ejemplo muy interesante de este mecanismo es el linc-RoR (*long intergenic RNA regulator of reprogramming*), el cual se une competitivamente a diferentes miRNAs cuyos blancos canónicos son los factores transcripcionales de pluripotencia Oct4, Sox2, y Nanog, previniendo así la degradación de los mRNAs asociados a estos factores transcripcionales, promoviendo el mantenimiento de la auto-renovación en las células troncales embrionarias y, por lo tanto evitando su diferenciación (Wang et al. 2013).

Modelo de andamiaje molecular

Se dice que un lncRNA actúa como un “andamio molecular” cuando tiene la capacidad de funcionar como un adaptador que atrae en proximidad espacial dos o más diferentes complejos proteicos, lo cual permite el ensamblaje de una macro-estructura con una función biológica definida (**Figura 4B**) (Spitale et al. 2011). El ejemplo clásico de andamio molecular es HOTAIR, un lncRNA de 2.2 kb codificado en el locus *HOXC*, que se expresa únicamente en células con identidades posicionales distales y posteriores (Rinn et al. 2007). HOTAIR interactúa de manera simultánea con los complejos PRC2 y LSD1-CoREST a través de diferentes dominios estructurales en su extremo 5' y 3', respectivamente (Tsai et al. 2010). Los complejos PRC2 y LSD1-CoREST son remodeladores de la cromatina, por lo que el ensamblaje de esta macro-estructura biológica regula en *trans* la represión

transcripcional de diversos genes mediante la regulación coordinada de la metilación de la H3K27 (lisina 27 de la histona H3) y la desmetilación de la H3K4me2 (lisina 4 de la histona H3). Las modificaciones post-traduccionales de histonas H3K27me3 y H3K4me2 están implicadas en la represión y activación transcripcional, respectivamente.

Mecanismo de lncRNAs como potenciadores de la transcripción

Los lncRNAs potenciadores de la transcripción (del inglés: *enhancer RNAs*), como su nombre lo indica, están involucrados en la activación génica. En esta clasificación entran tanto (a) los RNAs que se transcriben a partir de zonas genómicas conocidas como “potenciadores” (*enhancers*), como (b) aquellos que se transcriben a partir de cualquier otra región del genoma (*enhancer-like*).

Los *enhancers* o potenciadores son regiones del DNA necesarias para la activación regulada y el mantenimiento de la expresión de ciertos genes. Se dice que los potenciadores son elementos regulatorios distales ya que pueden encontrarse a cientos de kilobases (kb) de distancia de los genes que regulan (De Laat y Duboule, 2013), o incluso en cromosomas distintos (Spilianakis et al. 2005). Sin embargo, se ha propuesto que mediante la formación de asas cromosomales, estos elementos pueden entrar en contacto con los genes que regulan (**Figura 4D**). Además, los potenciadores son elementos importantes en la regulación de la expresión génica tanto espacial (tejido-específico) como temporalmente (Lai y Shiekhattar, 2014). En general, estas regiones pueden llevar a cabo su función mediante la unión de factores de transcripción involucrados en la

activación génica; además, y de manera similar a lo que sucede con los promotores, los potenciadores pueden ser identificados de manera global a través de la presencia de “marcas” propias de estos elementos, como H3K4me1, H3K27ac (acetilación de la lisina 27 en la histona H3), y la presencia de la acetiltransferasa de histonas p300 (Rinn y Chang, 2012). De manera adicional, estudios de secuenciación masiva del transcriptoma permitieron demostrar que estos elementos de regulación pueden transcribirse, produciendo lncRNAs (Kim et al. 2010; De Santa et al. 2010). Diversos estudios han propuesto que los transcritos que se generan a partir de los elementos potenciadores no son un simple *by-product* de la transcripción por la RNA Pol II. Por el contrario, se ha reportado que estos transcritos, denominados eRNAs (por sus siglas en inglés: *enhancer RNAs*), son funcionales y forman parte del mecanismo de activación transcripcional mediante el cual una región potenciadora regula a sus genes blanco. En algunos de estos casos se demostró que son las moléculas de RNA *per se*, y no el acto de la transcripción, las que son importantes en la activación transcripcional de genes (Li et al. 2013).

Por otro lado, el lncRNA *enhancer-like* HOTTIP está involucrado en la activación transcripcional de algunos genes homeóticos del cluster *HOXA* (Wang et al. 2011). HOTTIP interactúa con WDR5, proteína del complejo TrxG (Trithorax) que cataliza la H3K4me3, marca de histonas asociada con la activación transcripcional. A diferencia de los eRNAs, HOTTIP no se transcribe a partir de una región potenciadora, sin embargo la regulación que HOTTIP ejerce es dependiente de una cercanía espacial entre el locus a partir del cual HOTTIP se transcribe y los genes *HOXA* que regula, promovida por la formación de un asa cromosomal (**Figura 4D**). De esta manera, HOTTIP ayuda a reclutar al complejo trxG a los

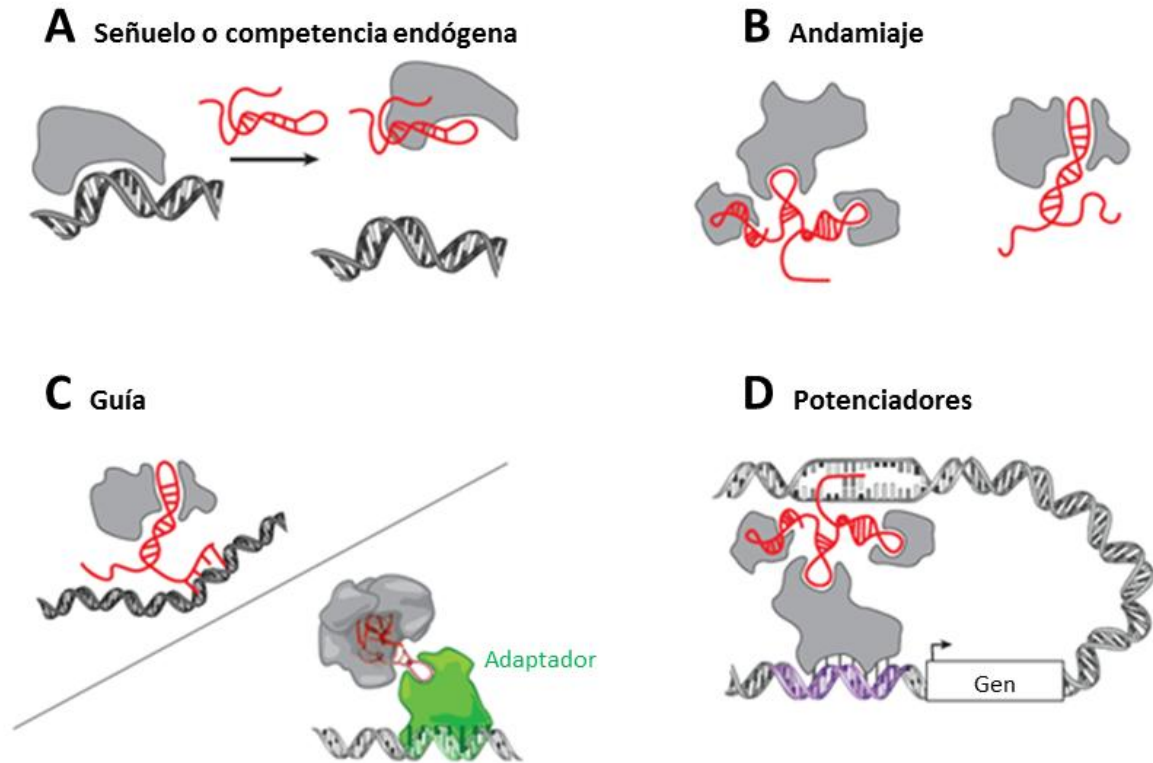


Figura 4.- Modelos funcionales de lncRNAs. (A) Los lncRNAs pueden actuar como “señuelos” o “esponjas” titulando proteínas importantes, como factores transcripcionales, o miRNAs (no se muestra en la figura), y por lo tanto evitando que lleven a cabo sus funciones reguladoras. (B) Los lncRNAs también pueden actuar como andamios moleculares, es decir, como una interfaz de contacto entre dos o más diferentes proteínas/complejos proteicos. (C) Otros lncRNAs pueden actuar como “guías”, ayudando a reclutar diversos tipos de proteínas, como remodeladores de la cromatina, a un locus de manera específica. Se ha propuesto que esto puede ocurrir a través de interacciones RNA-DNA, o a través de la interacción del lncRNA con una proteína con capacidad para unirse al DNA (*Adapter*). (D) A su vez, dos regiones linealmente distales en el genoma, pueden entrar en proximidad física mediante la formación de asas cromosomales. Se dice entonces que el lncRNA, transcrito a partir de una de esas regiones, promueve el reclutamiento *in cis* de ciertos factores o remodeladores de la cromatina, regulando así la activación transcripcional (Tomada y modificada de Rinn y Chang 2012).

genes *HOXA*, promoviendo así el mantenimiento de la marca de histonas H3K4me3 y por lo tanto, regulando *in cis* la activación génica de los mismos (Wang et al. 2011). El mecanismo de regulación de HOTTIP es análogo al mecanismo mediante el cual funcionan las regiones potenciadoras o *enhancers*, lo que le ha valido a HOTTIP ser considerado

como un lncRNA potenciador de la transcripción o *enhancer-like*.

Regulación epigenética

Las evidencias iniciales que mostraron que un RNA *per se* puede funcionar como una entidad reguladora provinieron de estudios de compensación de dosis; por un lado con la caracterización de Xist en la inactivación del cromosoma X (Penny et al. 1996; Wutz y Jaenisch, 2000), y por otro lado con el descubrimiento de diversos lncRNAs involucrados en el proceso de impronta genómica (Lyle et al. 2000; Pandey et al. 2008). Desde hace varios años se tenían evidencias indirectas de que el RNA podía influenciar la estructura de la cromatina, y por lo tanto, regular la expresión génica (Paul y Duerksen, 1975). En la actualidad, gracias a diferentes estudios, se sabe que los lncRNAs juegan un papel importante en la regulación epigenética, siendo esta función una de las mejor caracterizadas hasta el momento.

El genoma de un organismo eucarionte se organiza en una serie de estructuras que han mostrado ser fundamentales para efectuar diversas funciones, entre las cuales se encuentra la expresión regulada de genes. El empaquetamiento y compactación del material genético en el núcleo celular es posible gracias a la estructura conocida como cromatina, un conjunto de DNA y proteínas. Estas proteínas, mejor conocidas como histonas (Felsenfeld y Groudine, 2003), se organizan en octámeros formando al nucleosoma, alrededor del cual se enrollan aproximadamente 147 pares de bases (pb) de DNA. El octámero de histonas está constituido por un par de cada una de las histonas H2A,

H2B, H3 y H4. Los nucleosomas constituyen la unidad fundamental de la organización de la cromatina y representan su primer nivel de organización (**Figura 5**). El siguiente nivel de compactación es el resultado del plegamiento de los nucleosomas en una fibra compacta de 30 nm de diámetro, conocida como solenoide. Este nivel de organización se favorece mediante la unión de una quinta histona, H1, que no forma parte del nucleosoma y que interactúa con las regiones internucleosomales. La fibra de 30 nm conforma la base de niveles superiores de compactación del material genético (**Figura 5**). Esta estructura nucleoprotéica posee una organización y regulación definida dentro del núcleo celular. Dicha organización, en conjunto con la participación de elementos reguladores que se localizan en el DNA, permite una regulación fina y específica, la cual dicta los patrones de expresión para cada gen en un momento y tejido dado en los organismos pluricelulares.

Funcionalmente, el compactar el genoma en cromatina es *per se* un impedimento para la realización de diversos procesos; por ejemplo, la compactación del genoma puede generar una reducida accesibilidad a ciertas secuencias en el DNA. Afortunadamente, la célula eucariota desarrolló, en paralelo con los mecanismos de compactación, mecanismos que revierten de manera regulada dicho proceso. La estructura de la cromatina es por lo tanto dinámica y flexible, y varía dependiendo de: (a) su localización en el núcleo, así como de (b) modificaciones inducidas por diferentes tipos de proteínas conocidas genéricamente como “remodeladores de la cromatina” (Felsenfeld et al. 1996; Felsenfeld y Groudine, 2003).

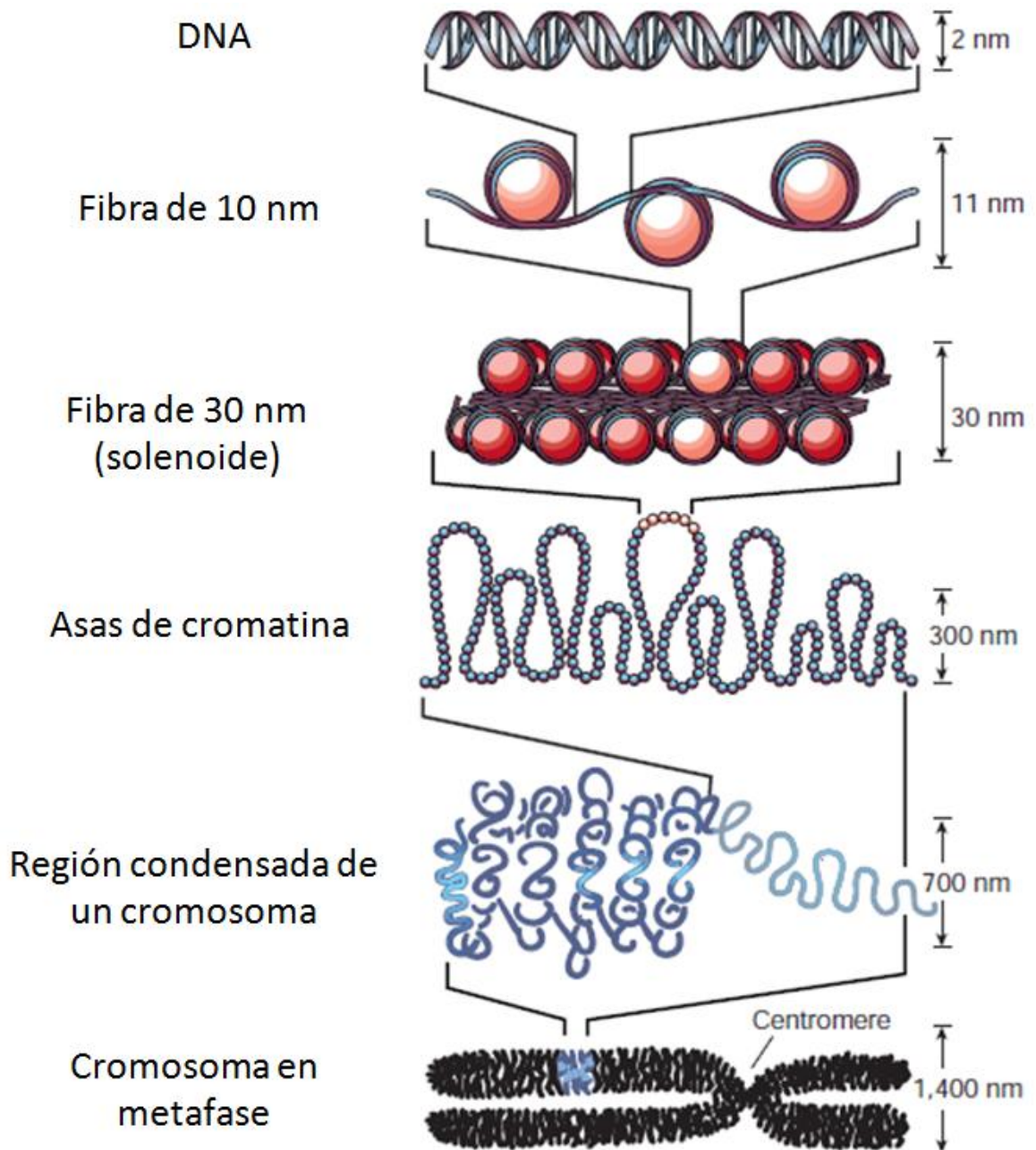


Figura 5.-Múltiples niveles de compactación del DNA. El DNA se enrolla alrededor de un octámero de histonas para dar lugar al nucleosoma. Esta estructura representa el primer nivel de organización de la cromatina y se conoce como "collar de perlas" o fibra de 10 nm. El segundo nivel de compactación (solenoid) es posible gracias a la incorporación de la histona H1 a la cromatina. Posteriormente, esta fibra forma asas de cromatina de gran tamaño, mismas que serán la base de niveles superiores de organización del genoma, hasta llegar a la formación de cromosomas metafásicos, los cuales constituyen el máximo grado de compactación (Tomada de Felsenfeld y Groudine, 2003).

Por lo tanto, la cromatina y su estructura tienen influencia en la regulación de la expresión génica. Por ejemplo, la forma más compacta de la cromatina, denominada heterocromatina, es inaccesible a la maquinaria transcripcional y por lo tanto se asocia con el silenciamiento génico. Por otro lado, aquellas regiones que muestran actividad transcripcional se asocian con zonas de la cromatina menos compactas o “abiertas” que se conocen como eucromatina (Felsenfeld y Groudine, 2003). De esta manera, se pueden distinguir dos diferentes estados estructurales de la cromatina con funciones antagónicas al nivel de la expresión génica, concluyendo que la estructura cromatínica de un dominio está íntimamente asociada a su actividad transcripcional. La transición coordinada entre un estado y otro, así como su mantenimiento a través de divisiones celulares sucesivas es posible gracias a la existencia de diversos mecanismos epigenéticos encargados de la regulación de la estructura de la cromatina.

Los complejos remodeladores de la cromatina (**Figura 6**, en azul) se expresan en todos los contextos celulares de manera ubicua, sin embargo, regulan diferentes conjuntos de genes en distintos tejidos celulares. Dado que para muchos de ellos no se han descrito dominios de unión al DNA (Müller y Kassis, 2006), no es claro cómo es que estos complejos, esenciales en el dictamen de la estructura de la cromatina, se reclutan de manera específica a los loci que regulan.

Recientemente, se han descrito diversos lncRNAs que interactúan con complejos remodeladores de la cromatina. En estos estudios, se demostró que el RNA es un elemento crucial en la localización global de los complejos remodeladores de la cromatina (Maison et al. 2002), es decir, que los lncRNAs ayudan a reclutar de manera específica, ya

sea en *cis* o en *trans*, a los complejos remodeladores de la cromatina hacia los loci o regiones que regulan (Koziol y Rinn, 2010). Por lo tanto, se ha llegado a proponer que los lncRNAs son los factores que confieren especificidad a los complejos remodeladores de la cromatina y modificadores de histonas, anclándolos a lugares específicos a lo largo del genoma (Wang y Chang, 2011), y de ahí su emergente importancia en el campo de la epigenética (**Figuras 4C y 6**).

A continuación se describirán brevemente algunos de los lncRNAs descritos recientemente en la regulación epigenética:

PcG/TrxG.- El grupo de proteínas Trithorax (TrxG) y el grupo de proteínas Polycomb (PcG) son complejos remodeladores de la cromatina que funcionan antagónicamente, es decir, el complejo TrxG se encarga de mantener un estado de expresión génica activo, mientras que las proteínas PcG están involucradas en el mantenimiento de un estado de represión génica (Ringrose y Paro, 2004). A su vez, se han caracterizado dos diferentes complejos multiproteicos Polycomb, PRC1 y PRC2 (por sus siglas en inglés: *Polycomb Repressive Complex*), ambos involucrados en el silenciamiento epigenético. Los complejos PcG y TrxG son responsables de las modificaciones post-traduccionales en las histonas H3K27me3 y H3K4me3, respectivamente; sin embargo, no se conoce en detalle el mecanismo mediante el cual PcG y TrxG son reclutados de manera específica a los genes que regulan. Esta pregunta es relevante dentro del campo epigenético si se considera que los complejos remodeladores de la cromatina son factores ubicuos, presentes en todos los contextos celulares, sin embargo regulan a un conjunto distinto de genes en diferentes tejidos celulares. Además, las proteínas PcG y TrxG no poseen dominios de unión a DNA, por lo

que no es claro cómo es que estos factores se reclutan de manera específica a los loci que regulan (Müller y Kassis, 2006). Recientemente se describieron algunos casos en los que estos remodeladores de la cromatina son reclutados de manera locus-específica gracias a su capacidad de interactuar con lncRNAs. A este mecanismo de reclutamiento se le conoce de manera general como mecanismo de “guía” (**Figura 4C**).



Figura 6.-Procesos involucrados en la regulación epigenética. En muchos casos, los lncRNAs son los elementos que confieren especificidad a los complejos remodeladores de la cromatina, reclutándolos de manera específica, a los loci que regulan. Se ha observado una desregulación en la localización de estos complejos epigenéticos (en rojo) en la ausencia de lncRNAs (mediante experimentos de pérdida de función).

Un ejemplo claro de un lncRNA involucrado en la regulación epigenética es Xist, quién es uno de los principales coordinadores del proceso de inactivación del cromosoma X (Pontier y Gribnau, 2011). Codificado en la región conocida como centro de inactivación del cromosoma X (*Xic*, por sus siglas en inglés: *X inactivation center*), la longitud de Xist varía dependiendo de la especie, y en el humano mide aproximadamente 19 kb. Xist se expresa únicamente en uno de los dos cromosomas X durante el desarrollo femenino, y su expresión está relacionada con una represión global de genes del cromosoma X a partir del cual se transcribe. Se determinó experimentalmente que una región de 1.6 kb localizada al interior de Xist, conocida como RepA, interactúa con Ezh2, una proteína del complejo PRC2, y que esta interacción coordina el reclutamiento de PRC2 al cromosoma X, siendo este evento el inicio de la inactivación, por lo que se ha propuesto que RepA podría estar funcionando como centro de nucleación del silenciamiento epigenético del cromosoma entero (Zhao et al. 2008). Tsix, transcrito no-codificante antisentido a Xist, reprime la expresión de Xist en *cis* al competir con RepA por interactuar con PRC2; de esta manera se evita el establecimiento prematuro de la marca de histonas H3K27me3 y se controla temporalmente el inicio de la inactivación del cromosoma de una forma precisa. Por lo tanto, se ha concluido que la expresión de RNAs no-codificantes en el cromosoma X es necesaria para la correcta formación de la heterocromatina a lo largo del cromosoma X inactivo (Lee, 2010).

Por otro lado, también se han descrito lncRNAs que interactúan con proteínas del complejo TrxG. Un ejemplo es *Mira*, un lncRNA de 800 pb localizado en el SDR (por sus siglas en inglés: “spacer DNA region”) entre los loci *Hoxa6* y *Hoxa7*. *Mira* activa la

transcripción de los genes homeóticos *Hoxa6/a7*, reclutando a la metiltransferasa de histonas del complejo TrxG MLL1. A su vez, la activación de *Hoxa6/a7* está relacionada con la diferenciación de las células troncales a células del mesodermo, en especial, a células germinales. Se ha propuesto que la formación de un híbrido de RNA/DNA podría ser el mecanismo mediante el cual se coordina el reclutamiento del activador epigenético MLL1 al locus *Hoxa6/a7* por el lncRNA *Mira* (Bertani et al. 2011).

Metilación del DNA.- Este proceso epigenético involucra una modificación bioquímica que consiste en agregar un grupo metilo (-CH₃) a las citosinas, normalmente en un contexto CpG. Esta modificación es mediada por metiltransferasas de DNA (DNMT, por sus siglas en inglés: *DNA methyltransferase*) y ha sido reportada en todos los vertebrados analizados, siendo una de las marcas epigenéticas más estudiadas. La metilación del DNA se asocia normalmente con regiones de heterocromatina, y por lo tanto con silenciamiento génico. La metilación del DNA puede afectar la transcripción de genes en 2 maneras: i) la presencia de los grupos metilo puede representar un impedimento estérico para la unión de factores transcripcionales al DNA (Choy et al. 2010), ii) la metilación del DNA es reconocida por proteínas con dominios de unión a CpG metilados (MBD, por sus siglas en inglés: *Methyl-CpG-Binding Domain*). Las proteínas MBD se unen a las citosinas metiladas, y ayudan posteriormente al reclutamiento de desacetilasas de histonas (HDAC, por sus siglas en inglés: *Histone Deacetylase*) y otros remodeladores de la cromatina (Ichimura et al. 2005; Reese et al. 2007; Ramírez et al. 2012), regulando la compactación de cromatina en esa región (formación de heterocromatina), y promoviendo así la represión

transcripcional de un locus. La metilación del DNA es un componente importante de diversos procesos celulares, como por ejemplo el desarrollo embrionario, la inactivación del cromosoma X, la impronta genómica y la estabilidad cromosomal (Reik et al. 2001; Sun et al. 2006; Ogawa et al. 2008; Lawson et al. 2013; Messerschmidt et al. 2014).

Recientemente se han reportado diversos trabajos que demuestran que los lncRNAs están involucrados en la metilación del DNA. Por una lado, *Kcnq1ot1* es un lncRNA de 91.5 kb involucrado en la impronta de un grupo de genes del dominio *Kcnq1*. La expresión de *Kcnq1ot1* conduce al silenciamiento transcripcional de los genes improntados de ese locus (Pandey et al. 2008). Se ha propuesto que *Kcnq1ot1* no sólo ayuda en el establecimiento de la represión transcripcional de los genes improntados en el alelo paterno, sino también en el mantenimiento de la misma a través de su interacción con la enzima DNMT1. *Kcnq1ot1* recluta a DNMT1 al dominio, y de esta manera contribuye al mantenimiento de la metilación del DNA durante divisiones celulares subsecuentes (Mohammad et al. 2010).

Otro ejemplo que demuestra la importancia de los lncRNAs en la regulación de la metilación del DNA a nivel local es el lncRNA *ecCEBPA*, que traslapa con la transcripción del gen codificante sensible a metilación *CEBPA*. El lncRNA *ecCEBPA* es capaz de modular localmente la metilación, y por lo tanto, la expresión de *CEBPA*. La DNMT1 interactúa con el lncRNA *ecCEBPA* con mayor afinidad incluso que al DNA. Esta interacción es dependiente de estructuras tallo-asa en el RNA que ayudan a “secuestrar” o titular a la DNMT1, previniendo así la metilación del locus (Di Ruscio et al. 2013), y dando lugar a una mayor transcripción del mRNA de *CEBPA*. Los ejemplos anteriormente descritos permiten

aseverar que los lncRNAs son componentes importantes en la regulación de la metilación del DNA, ya sea promoviendo o bloqueando la acción de la DNMT1.

Complejos remodeladores de la cromatina dependientes de ATP.- Además de la metilación del DNA y de las modificaciones post-traduccionales de histonas, la estructura de la cromatina depende también del posicionamiento de los nucleosomas a lo largo del DNA (**Figura 5**). Los complejos remodeladores dependientes de ATP controlan la accesibilidad a secuencias de DNA mediante la desestabilización de las interacciones entre el DNA y las histonas, favoreciendo el reposicionamiento nucleosomal, la expulsión de histonas del DNA y el intercambio de variantes de histonas (Jiang y Pugh, 2009; Hargreaves y Crabtree, 2011). Estos complejos remodeladores poseen un dominio común de ATPasa, lo cual les confiere la capacidad de alterar la estructura de la cromatina, y por lo tanto de regular la expresión génica, utilizando la energía proveniente de la hidrólisis de ATP. Es importante señalar que el efecto que este tipo de remodeladores de la cromatina tienen sobre la expresión génica puede ser tanto negativo como positivo (Vignali et al, 2000).

Existen diferentes tipos de familias de complejos dependientes de ATP. Recientemente se determinó que la actividad de los remodeladores de la cromatina pertenecientes a la familia SWI/SNF es regulada también por lncRNAs. Existen al menos 2 diferentes trabajos que describen la interacción de diferentes subunidades del complejo SWI/SNF con RNA, así como sus correspondientes efectos regulatorios.

El primer trabajo se realizó en plantas y demostró que lncRNAs producidos a partir de la RNA Pol V regulan el posicionamiento de nucleosomas reclutando al complejo SWI/SNF de manera específica a ciertos loci dentro del genoma, y promoviendo así su represión transcripcional (Zhu et al. 2013).

El segundo trabajo identificó un lncRNA que se transcribe a altos niveles en cáncer de próstata, y cuya expresión correlaciona con un mayor riesgo en la progresión de la enfermedad, así como con mayores tasas de mortalidad. Este lncRNA, denominado *SChLAP1*, interactúa directamente con SNF5, una subunidad del complejo SWI/SNF. Esta interacción provoca un decremento en la unión del complejo SWI/SNF a 30% de los promotores que regula, y en consecuencia, una disminución en la expresión de sus genes blanco (Prensner et al. 2013).

Estos hallazgos sugieren que tanto el reclutamiento como la actividad de los complejos remodeladores de la cromatina dependientes de ATP son mediados por lncRNAs, y que además, esta característica se encuentra conservada durante la evolución, por lo que podría representar un mecanismo de regulación general que aún queda por determinarse en detalle.

Dinámica nuclear.- El núcleo es el organelo celular que contiene a la información genética y posee una estructura dinámica y flexible definida espacio-temporalmente, es decir, es específica de cada tejido, así como de cada etapa de la diferenciación celular y del ciclo celular (Kosak y Groudine, 2004; Misteli, 2005). La organización tridimensional de la

cromatina al interior del núcleo juega un papel importante en la regulación de la expresión génica. De manera general y como se había mencionado previamente la cromatina se encuentra organizada en dominios de heterocromatina y eucromatina. Desde una perspectiva citológica, la heterocromatina se ha encontrado predominantemente asociada a la periferia nuclear y al nucléolo, mientras que la eucromatina tiene una localización más central al interior del núcleo (Geyer et al. 2011). Además, y de manera contraria a lo antes pensado, los cromosomas no se distribuyen de manera aleatoria al interior del núcleo. Diversas evidencias indican que el genoma de una célula eucarionte se encuentra organizado en “territorios cromosomales” en los que cada cromosoma ocupa un lugar definido al interior del núcleo (Cremer y Cremer, 2010). De manera adicional, la maquinaria transcripcional y la producción de RNA se localizan dentro de foci discretos, lo que sugiere la existencia de “fábricas transcripcionales”, es decir, sitios bien definidos dentro del núcleo celular ricos en componentes de la maquinaria transcripcional, tales como la RNA polimerasa, y en donde genes transcripcionalmente activos que se encuentran linealmente distales, o incluso en cromosomas distintos, convergen espacialmente (Osborne et al. 2004). Por lo tanto, los genes se presentan como elementos cuya posición es dinámica dentro del núcleo, moviéndose hacia adentro y hacia afuera de las fábricas transcripcionales, y en consecuencia activándose o reprimiéndose transcripcionalmente, respectivamente.

En este sentido se han identificado varios lncRNAs con actividades reguladoras al nivel de la estructura tridimensional de la cromatina. Uno de los ejemplos más recientes es el lncRNA FIRRE, cuyo locus se localiza en el cromosoma X y pertenece al grupo de

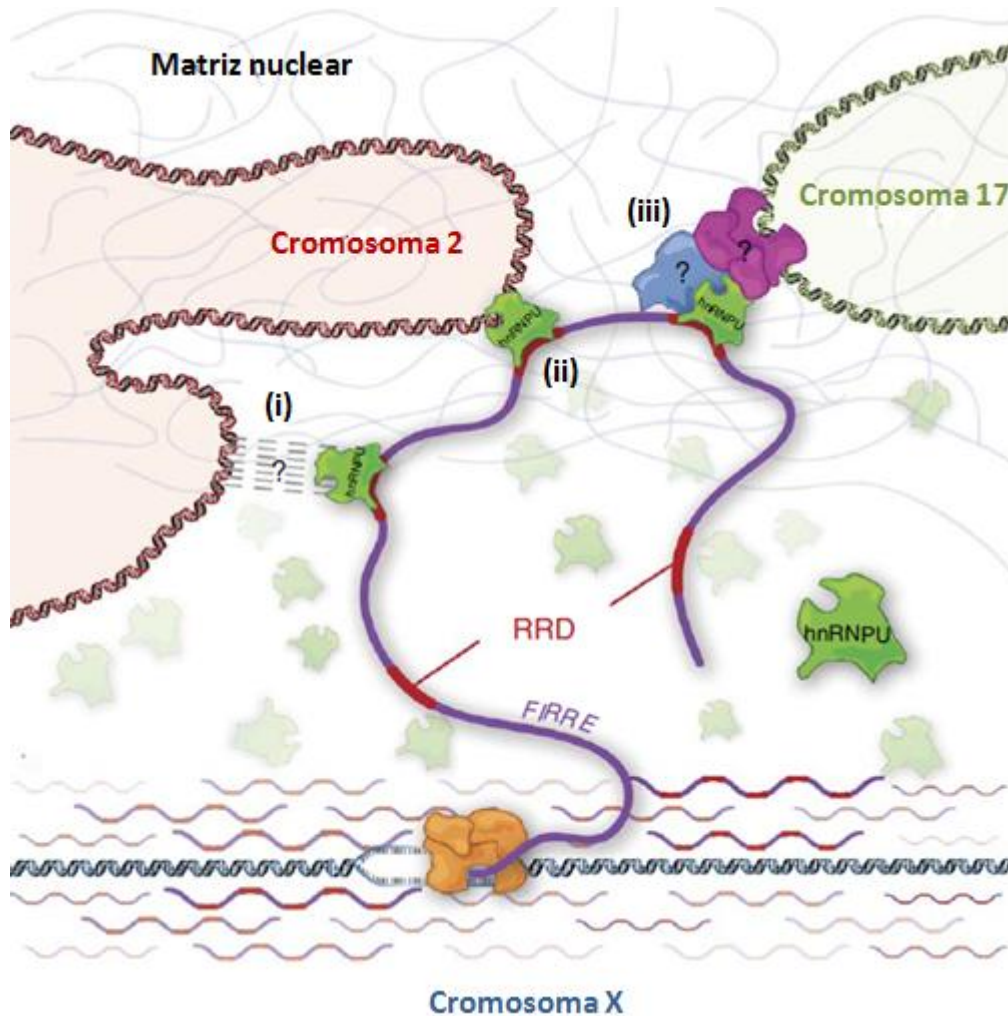


Figura 7.- Modelo de acción del lncRNA FIRRE como un factor de organización regional. El lncRNA FIRRE interactúa con la proteína hnRNPU mediante su dominio RRD (por sus siglas en inglés: *repeating RNA domain*). Se sabe que tanto FIRRE como hnRNPU son necesarios en la formación de las interacciones *trans* cromosomales entre el cromosoma X y otros cromosomas. Se han sugerido 3 diferentes posibles mecanismos que explican la naturaleza de las interacciones *trans* cromosomales mediadas por FIRRE y hnRNPU: (i) La interacción de hnRNPU con elementos adicionales de la matriz nuclear, (ii) la unión directa de hnRNPU a secuencias de adhesión a la matriz nuclear en el DNA, ó mediante (iii) la interacción indirecta con otras proteínas con dominios de unión a DNA (Tomada y modificada de Hacisuleyman et al. 2014).

genes que escapan a la inactivación del cromosoma X. En este estudio se demostró que FIRRE, en conjunto con la proteína de la matriz nuclear hnRNPU, es responsable de la formación de una estructura tridimensional que facilita la interacción en *trans* de

diferentes regiones cromosomales, por lo que se puede decir que actúa como un organizador de la arquitectura nuclear (**Figura 7**).

Ya que la formación de esta estructura es dependiente de FIRRE, se ha propuesto que una de las funciones de los lncRNAs podría consistir en reunir espacialmente genes involucrados en rutas metabólicas o procesos celulares similares, ya sea mediante la formación de estructuras que propicien interacciones tanto intercromosomales (Hacisuleyman et al. 2014) como intracromosomales mediante la formación de asas de cromatina (Zhang et al. 2014), regulando así la co-transcripción de genes de manera muy similar a lo propuesto previamente para las fábricas transcripcionales.

Todos los ejemplos expuestos anteriormente demuestran que la regulación de los loci mencionados no es el resultado de transcripción pasiva; por el contrario, las mismas moléculas de RNA no-codificante son necesarias para facilitar el reclutamiento de los remodeladores de cromatina, ayudando así en el silenciamiento/activación de estos loci.

Son contados los casos en los que se ha demostrado que el reclutamiento de algún complejo remodelador a sus loci blanco es mediado por un RNA largo no-codificante (Tsai et al. 2010); sin embargo se desconoce si este es un mecanismo general de reclutamiento que simplemente no había sido reportado anteriormente, y que además podría ser propio de todos los remodeladores de la cromatina, o solo de un subconjunto de ellos.

Conforme el poder tecnológico ha ido en aumento, las innovaciones en diversas técnicas de secuenciación han permitido el abaratamiento, así como la facilitación de la

obtención de secuencias de genomas y transcriptomas completos (así como de proteomas) que permitan abordar de una manera más global el estudio de los procesos biológicos y celulares. Con los nuevos proyectos de secuenciación masiva (en particular del GENCODE) se ha puesto en relevancia la anotación de cerca de 10,000 nuevos loci de RNA largos no-codificantes en el genoma humano (Derrien et al. 2012). Elucidar la función de todos esos nuevos RNA largos no-codificantes se presenta como el siguiente reto. Las caracterizaciones funcionales realizadas hasta el momento muestran que estos elementos participan en la regulación génica a distintos niveles, y en consecuencia, los lncRNAs están ganando importancia día a día como una nueva clase de genes (Mercer y Mattick, 2013). Algunos esfuerzos recientes en la literatura han abordado el problema de manera inicial llevando a cabo ensayos de pérdida (*Knockdown*) y ganancia de función (sobre-expresión) de los RNA largos no-codificantes de interés, e incluso algunos estudios han realizado este tipo de experimentos de manera sistemática y a gran escala (Guttman et al. 2011; Lin et al. 2014). Sin embargo, este tipo de enfoques no deja de ser indirecto, por lo que la validación funcional de cada uno de estos elementos, así como la elucidación de sus mecanismos de acción, se muestran como necesarias. En todos los ejemplos de caracterizaciones funcionales mencionados anteriormente se demostró que los RNA no-codificantes son entidades moleculares con funciones reguladoras. En nuestro grupo de investigación, estamos interesados particularmente en la caracterización funcional de un lncRNA presente en el locus α -globina de pollo.

1.1. ANTECEDENTES GENERALES

El transporte de oxígeno es esencial en el desarrollo de cualquier organismo pluricelular, sin embargo, los requerimientos de este gas durante cada etapa del desarrollo son diferentes. Para poder trasladar el oxígeno por medio de la sangre a lo largo del organismo se requiere de la hemoglobina, una proteína sintetizada en los eritrocitos (Palis y Segel, 1998). La hemoglobina es una proteína tetramérica compuesta de dos diferentes tipos de cadenas polipeptídicas denominadas α y β (**Figura 8B**). Cada cadena es producida a partir de dominios multigénicos localizados en diferentes cromosomas, pero co-regulados a lo largo del desarrollo (**Figura 8A y 8C**).

Durante el desarrollo, el genoma se somete a una compleja red de regulación transcripcional regida por diversos factores, tanto genéticos como epigenéticos, con el fin de establecer patrones de expresión propios que definen la identidad celular. Durante el desarrollo, las células madre hematopoyéticas se comprometen a la producción de células con un linaje definido, proceso conocido como diferenciación celular. Entre los diferentes linajes a los que las células madre hematopoyéticas dan origen, se encuentra el linaje eritroide (**Figura 9**; Ferreira et al. 2005). Una vez determinado el linaje eritroide, la diferenciación celular avanza hacia estadios eritroides maduros. La culminación de este proceso es la generación de células terminalmente diferenciadas, en este caso, los eritrocitos (**Figura 12**). Este proceso de diferenciación celular se conoce como eritropoyesis (Hattangadi et al. 2011).

La producción de sangre es similar para todos los vertebrados y se da en etapas secuenciales durante el desarrollo embrionario, distinguiéndose dos fases: la primitiva (fase transitoria) y la definitiva (fase permanente, adulta). Se denomina eritropoyesis primitiva al proceso de producción de eritrocitos que tiene lugar en etapas tempranas del

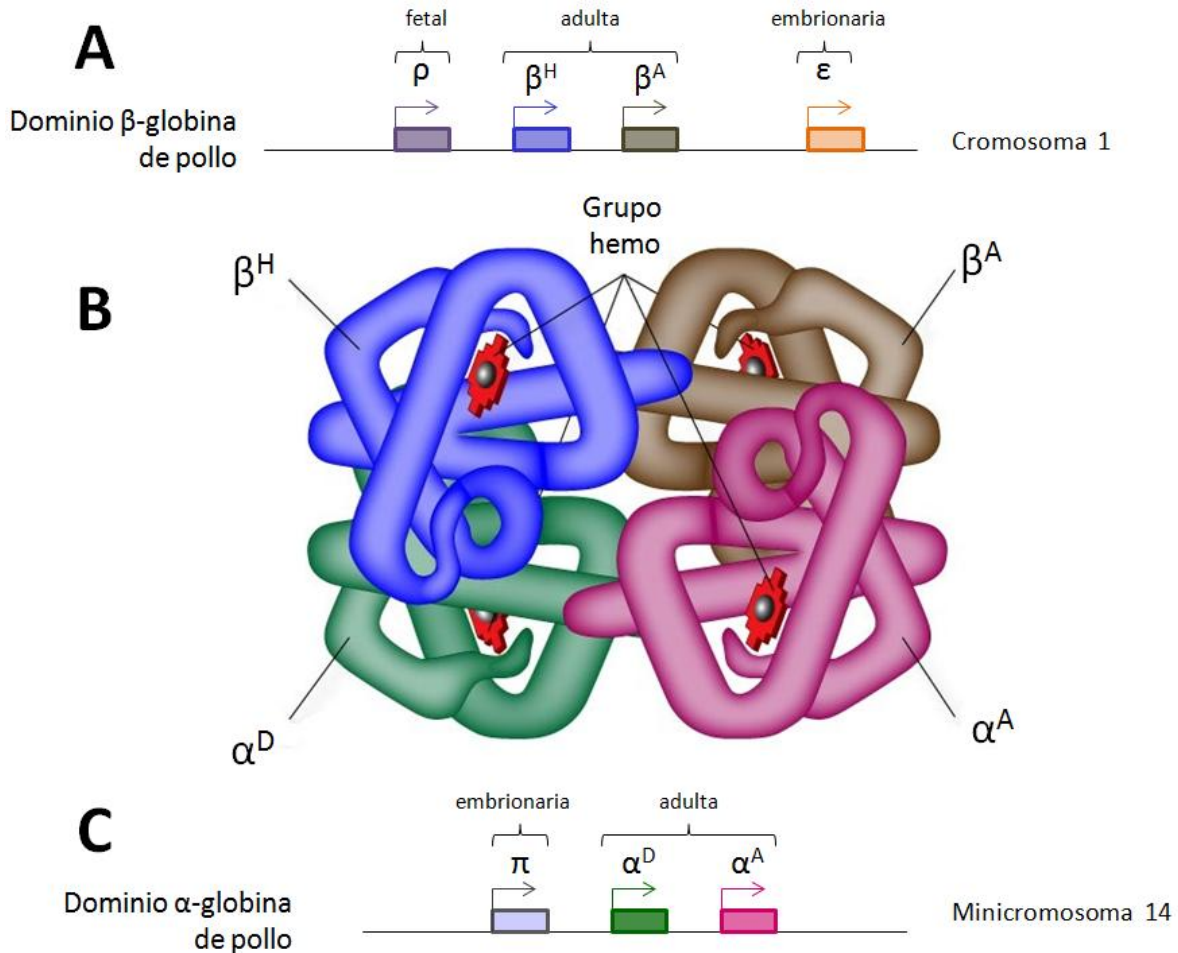


Figura 8.- Estructura de la hemoglobina y dominios multigénicos α y β -globina de pollo. (A) El dominio β -globina se localiza en el cromosoma 1 y está compuesto por el gen ϵ de expresión embrionaria, el gen ρ de expresión fetal, y los genes β^A y β^H de expresión adulta. (C) El dominio α -globina se localiza en el minicromosoma 14 y se compone del gen embrionario π , y los genes de expresión adulta α^D y α^A . (B) La estructura tetramérica que se muestra en la figura correspondería a la hemoglobina de pollo que se expresa en la etapa adulta, la cual se conforma de 2 cadenas polipeptídicas α y 2 cadenas polipeptídicas β .

desarrollo, y que se define por la expresión de las globinas embrionarias (Stamatoyannopoulos, 2005). A su vez, la eritropoyesis definitiva reemplaza a la eritropoyesis primitiva en etapas posteriores o tardías del desarrollo embrionario, y se caracteriza por la expresión de las globinas adultas.

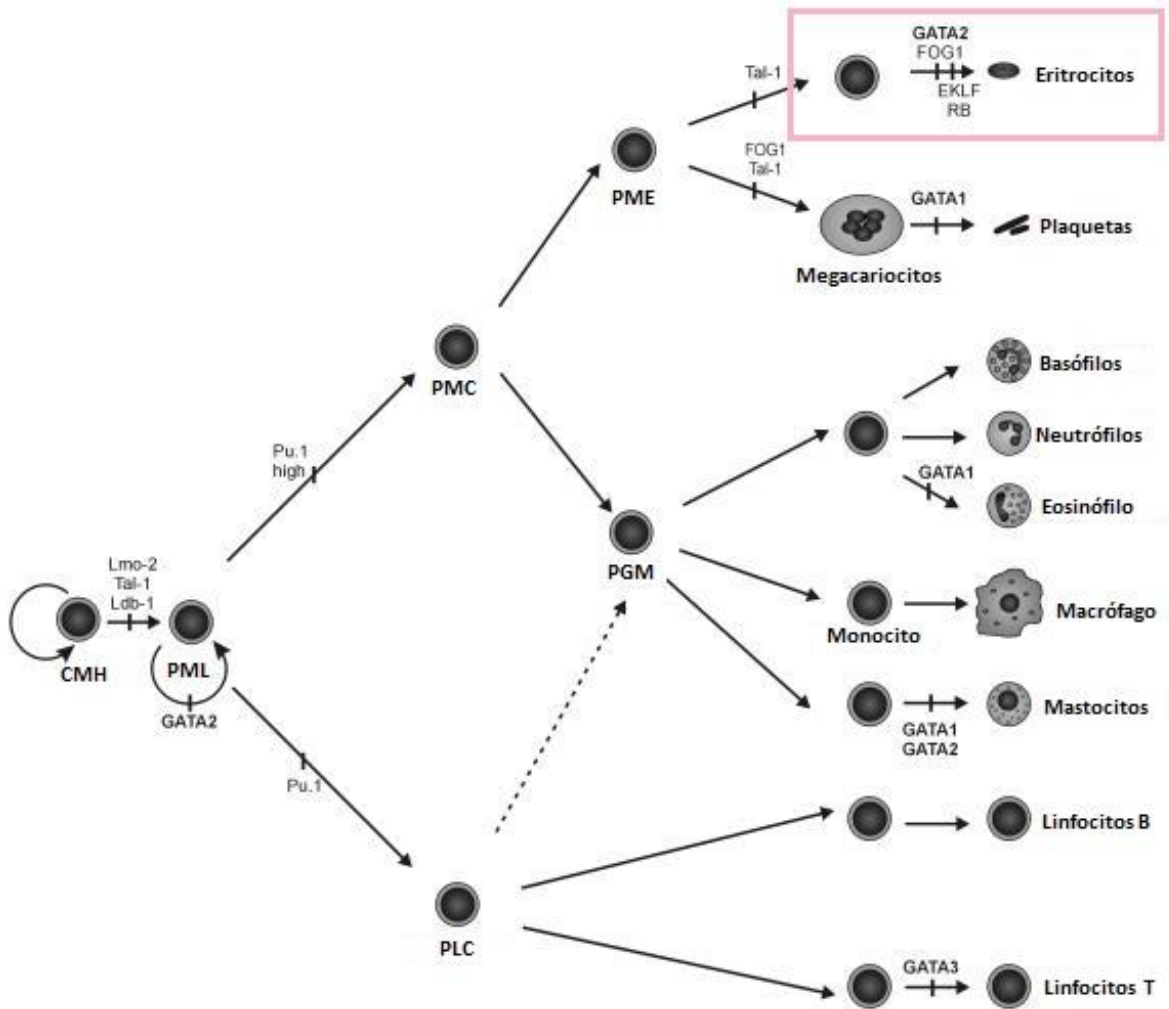


Figura 9.- Determinación de los diferentes linajes celulares sanguíneos durante la hematopoyesis. La hematopoyesis se entiende como el conjunto de diferentes procesos de diferenciación celular encargados de la producción de las células sanguíneas a partir de las células madre hematopoyéticas (CMH). En las flechas de la figura se indican algunos de los factores transcripcionales más importantes en el proceso de determinación y diferenciación celular durante la hematopoyesis. PML: Progenitores multilíneaje; PMC: Progenitor Mielode Común; PLC: Progenitor Linfode Común; PME: Progenitor Megacariocítico-Eritroide; PGM: Progenitor Granulocito-Monocito (Tomada y modificada de Ferreira et al. 2005)

Como mencionamos anteriormente, la producción de hemoglobina se realiza a partir de la síntesis de cadenas polipeptídicas independientes, codificadas a partir de diferentes dominios multigénicos. En el contexto de esta tesis, nos enfocaremos exclusivamente en el dominio α -globina de pollo. En la **Figura 10** se presenta el dominio α -globina de pollo:

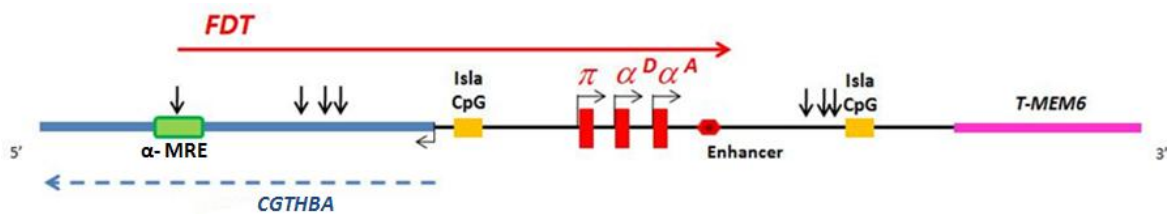


Figura 10.- Estructuración del dominio α -globina de pollo En rojo se encuentran los genes α -globina, el gen π codifica para la globina que se expresa durante la etapa embrionaria, mientras que α^D y α^A corresponden a los genes de expresión adulta. Las flechas verticales en color negro corresponden a sitios de hipersensibilidad a la DNasa I. La orientación de la transcripción de los genes se indica con las flechas horizontales.

Durante el desarrollo embrionario del pollo, los eritrocitos primitivos aparecen en la circulación entre los días 4 y 5 post-fecundación, y expresan la globina embrionaria π (**Figura 10** y **11**). A partir de los días 6 y 7 post-fecundación, el linaje adulto inicia de manera específica la transcripción de los genes α^D y α^A (Felsenfeld, 1993). Conforme se avanza en el desarrollo, se da una inversión en los patrones de expresión: el gen embrionario π se silencia en etapas tardías del desarrollo, mientras que las globinas adultas α^D y α^A incrementan sus niveles. A partir del día 10 del desarrollo de pollo, el linaje adulto de eritrocitos se impone en el organismo de manera definitiva (**Figura 11**).

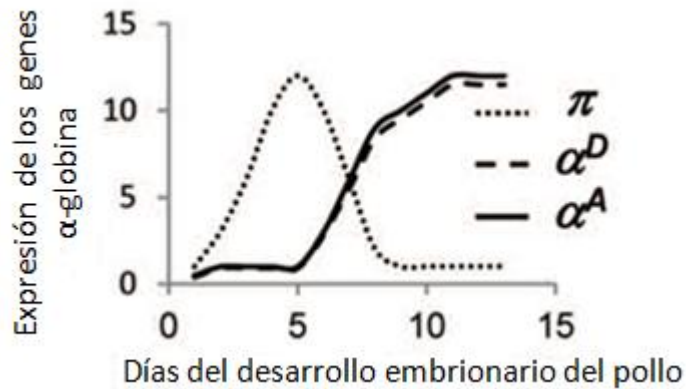


Figura 11.- Expresión diferencial de los genes α-globina de pollo durante el desarrollo embrionario. En la gráfica se muestran los niveles de expresión de las α-globina en eritrocitos de pollo obtenidos a diferentes días post-fecundación (Tomada y modificada de Valdes-Quezada et al. 2013, ANEXO 1).

El reemplazo de la eritropoyesis embrionaria o primitiva por la eritropoyesis definitiva o adulta a lo largo del desarrollo embrionario de pollo está definido a nivel molecular por un “recambio” en la expresión de los genes globina (**Figura 11**) (conocido en inglés como *hemoglobin switching*) (Palis y Segel, 1998; Baumann y Dragon, 2005). A su vez, la expresión de los genes globina se encuentra regulada a través del proceso de diferenciación eritroide. Para este fin, en el laboratorio se cuenta con una serie de líneas celulares aisladas a partir de tejido adulto de pollo (**MATERIALES Y MÉTODOS**) que mimetizan la diferenciación eritroide adulta (Beug et al. 1979 y 1982; Graf, 2002). Estas líneas celulares están comprometidas a un linaje eritroide adulto, y se encuentran arrestadas en diferentes estadios de la eritropoyesis (**Figura 12**). En este sistema celular, se ha observado que la expresión de los genes α-globina adultos, α^D y α^A, se presenta en la etapa final de la diferenciación eritroide, que a su vez es un símil de eritrocitos maduros de 10 días (**Figura 12**, células **HD3 inducidas** a diferenciación). La expresión de la globina

embrionaria π no se detecta en ninguna de las líneas celulares utilizadas en el sistema de la **Figura 12**. Lamentablemente, la falta de un sistema análogo para el estudio de la diferenciación eritroide primitiva no nos permite establecer si este proceso es similar durante la diferenciación eritroide embrionaria.

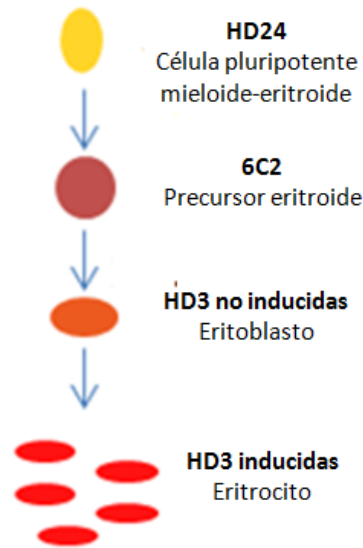


Figura 12.- Tipos celulares eritroides utilizados para el estudio de la diferenciación eritroide adulta de pollo. En el esquema se indican algunos de los estadios más importantes en la diferenciación eritroide, así como las líneas celulares con las que se cuenta en el laboratorio (en negrita) para su estudio. Las flechas indican el flujo de un estado menos diferenciado a uno más diferenciado. Todos los precursores eritroides mostrados presentan niveles basales de expresión de los genes α -globina, con excepción de las células HD3 inducidas a diferenciación, las cuales incrementan considerablemente los niveles de expresión de los genes adultos α^D y α^A únicamente.

Por lo tanto, los genes globina representan un paradigma en la regulación génica ya que están sujetos a una fina regulación transcripcional, tanto a través del desarrollo embrionario de pollo (**Figura 11**), como a través de la diferenciación eritroide (**Figura 12**). Es importante señalar que los dominios α y β -globina de pollo han sido objeto de estudios muy importantes en la regulación diferencial de genes ya que los eritrocitos de pollo, a

diferencia de los eritrocitos de la mayoría de los mamíferos, no carecen de núcleo, lo que ha permitido el abordaje de preguntas a nivel transcripcional con enfoques tanto genéticos como epigenéticos.

El dominio α -globina de pollo se localiza en el minicromosoma 14 y se extiende a lo largo de una región de ~40 kb. En el extremo 5' del dominio se han descrito diferentes sitios de hipersensibilidad a la DNasa I (**Figura 10**). La hipersensibilidad se entiende como una mayor propensión de esa zona al corte por parte de la DNasa I, lo que se interpreta como la presencia de una estructura cromatínica laxa en esa región. Se sabe que la presencia de estos sitios correlaciona con elementos reguladores de la actividad transcripcional, muy probablemente debido a la unión de diferentes factores, como por ejemplo, factores transcripcionales. A su vez, los promotores de los genes α -globina de pollo poseen una organización modular y contienen la información necesaria para su expresión durante las diferentes etapas de la diferenciación eritroide (Felsenfeld, 1993); sin embargo, requieren del elemento potenciador localizado en el extremo 3' del dominio (*enhancer*, **Figura 10**) para alcanzar los niveles de expresión adecuados (Rincón-Arano et al. 2005; Escamilla-Del-Arenal y Recillas-Targa, 2008). En el extremo 5' del dominio se encuentra el gen *C16orf35/CGTHBA*, el cual se transcribe en dirección opuesta a los genes α -globina y cuyo promotor se identificó como una isla CpG (región genómica con una alta frecuencia de sitios CpG) situada 4 kb río arriba del gen de expresión embrionaria π . Se ha propuesto que el dominio α -globina de pollo está embebido en una región con una estructura cromatínica constitutivamente abierta, misma que parece estar conservada a lo largo de la evolución (Razin et al. 2003; Brown et al. 2006). Nuestro grupo de investigación

se ha dedicado durante años a la elucidación de los mecanismos moleculares responsables de la regulación de los genes α -globina durante el desarrollo de pollo y la diferenciación eritroide, con un énfasis particular en los mecanismos involucrados en el recambio de la hemoglobina.

Con respecto a la regulación de los genes α -globina de pollo en el contexto del desarrollo, resultados recientes de nuestro grupo de investigación utilizando la técnica de 3C (del inglés: *Chromosome configuration capture*), sugieren que el elemento potenciador (*enhancer 3'*, **Figura 10**) interactúa con los promotores de los genes π y α^D desde etapas tempranas del desarrollo embrionario (eritrocitos de 5 días post-fecundación), y no así con el gen α^A (**Figura 13**). Se ha propuesto que el contacto físico entre el promotor de π , el promotor de α^D y el potenciador 3' es promovido por la formación de asas cromosomales (**Figura 13**), lo cual es consistente con el mecanismo de funcionamiento previamente reportado para los potenciadores (Song et al. 2007). Además, esta estructura (o como se le denomina en inglés: *hub*) parece ser específica del linaje eritroide (Gavrilov y Razin, 2008). Con respecto a la expresión de los genes globina, se ha propuesto que en etapas tempranas del desarrollo embrionario (5 días), la expresión del gen π es promovida por la presencia de factores transcripcionales específicos para ese estadio, con una ausencia concomitante de factores específicos para los genes adultos que únicamente se encuentran presentes en la etapa adulta (Valdes-Quezada et al. 2013, **ANEXO 1**). De esta manera, los eritrocitos circulantes en las etapas embrionarias tempranas expresan preferencialmente al gen π . Conforme se avanza en el desarrollo, el recambio en la expresión de las globinas es favorecido por el silenciamiento de la globina embrionaria π ,

con un aumento en paralelo de la expresión de las globinas adultas α^D y α^A (etapa de 10 días del desarrollo). Los resultados de este trabajo también sugieren que la metilación del DNA, en particular al nivel del promotor y del cuerpo del gen π , es el mecanismo involucrado en el establecimiento y mantenimiento del silenciamiento transcripcional de la globina embrionaria (**Figuras 11 y 13**). Además, la unión del factor CTCF a un sitio cercano del promotor del gen α^D parece delimitar este sub-dominio de silenciamiento génico, funcionando como una barrera o límite entre el nuevo sub-dominio con el gen π silenciado, y el gen ahora transcripcionalmente activo α^D . Adicionalmente, la metilación del gen π , y en consecuencia su silenciamiento transcripcional, se ven acompañados por su disociación del *hub* (**Figura 13**), es decir, el contacto observado en etapas tempranas entre π y el potenciador 3' se pierde en etapas tardías del desarrollo (Valdes-Quezada et al. 2013, **ANEXO 1**).

Por otro lado, en el contexto de la diferenciación eritroide adulta también se reportó la formación de esta estructura tri-dimensional o *hub* (Gavrilov y Razin, 2008). En este caso, los experimentos de 3C se realizaron en células HD3 (no expresan los genes globina) y HD3 inducidas a diferenciación (expresión de α^D y α^A). Dado que los resultados obtenidos en un experimento de 3C son poblacionales, lo que se concluyó es que la formación de este *hub* es dependiente del estadio de la diferenciación celular, con la presencia de *hubs* “inmaduros” y poco frecuentes en el estadio de eritoblastos, y *hubs* completamente formados y en mayores frecuencias en el estadio terminal de diferenciación (HD3 diferenciadas = símil de eritrocitos terminalmente diferenciados). En el *hub* descrito en este trabajo se observa nuevamente que el potenciador 3' interactúa

con el gen α^D , y no así con π ni α^A . Dado que la expresión de las α -globina adultas correlaciona con la formación de este *hub*, esto sugiere fuertemente que esta estructura es importante para la activación transcripcional de los genes globina.

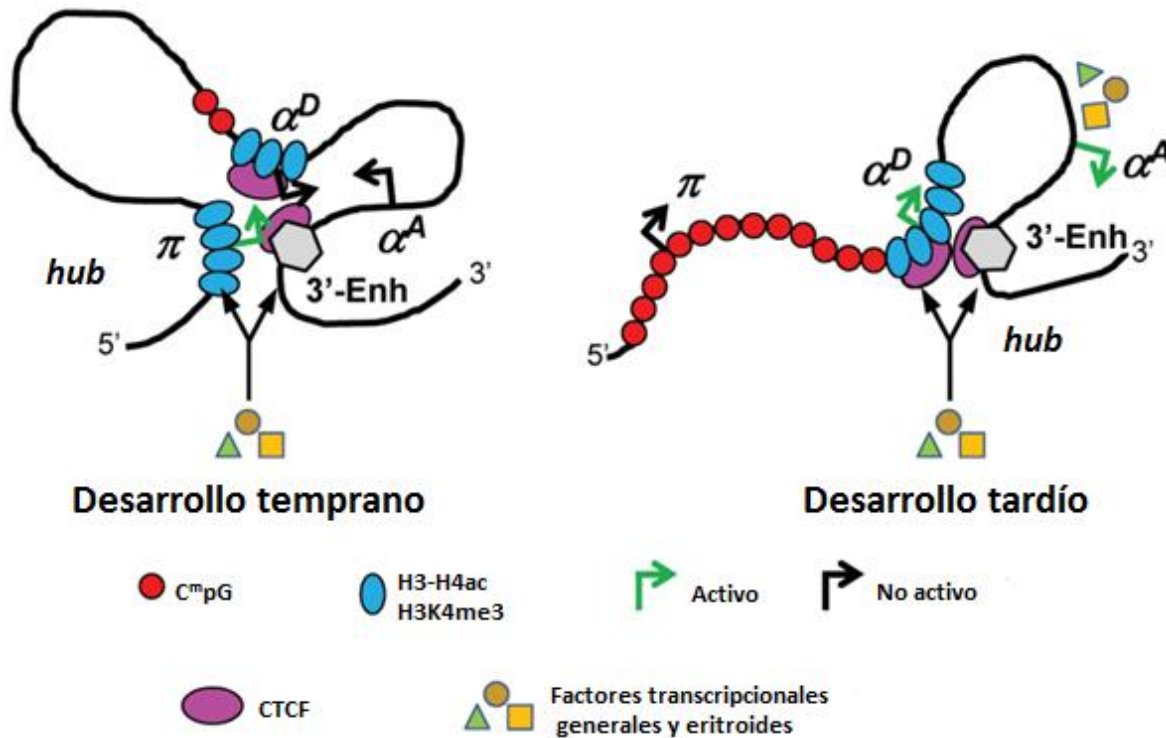


Figura 13.- Conformación topológica del dominio α -globina de pollo durante el desarrollo. En etapas tempranas del desarrollo (5 días) los genes π y α^D contactan al potenciador 3' (3'-Enh), formando una estructura conocida como *hub*. En esta etapa se expresa preferencialmente π . En etapas tardías del desarrollo, una vez que el recambio de hemoglobina ha ocurrido y que la eritropoyesis definitiva se ha impuesto, el gen π se encuentra silenciado vía metilación del DNA ($C^m pG$), disociándose del *hub*. En esta etapa se expresan únicamente las globinas adultas α^D y α^A .

Por otro lado, un estudio reciente estableció la presencia de transcripción no-codificante que corría por las zonas intergénicas del dominio α -globina de pollo. Se ha propuesto que este transcrito se extiende a lo largo de 33 kb, cubriendo la totalidad del dominio (Razin et al. 2004), e incluso, traslapando con los genes globina, por lo que se le llamó FDT (por sus

siglas en inglés: *Full Domain Transcript*). En la **Figura 10** se muestra la localización propuesta para el FDT sobre el dominio α -globina de pollo. Sin embargo, hasta el momento no se contaba ni con la caracterización detallada de este transcrito, ni con pruebas experimentales que nos permitieran concluir que este RNA es funcional. Dada la creciente importancia de los lncRNAs en la regulación génica, en nuestro grupo de investigación nos sentimos particularmente atraídos hacia la posibilidad de que este elemento pudiera ser funcional, en particular, dentro del contexto de la regulación de los genes α -globina de pollo.

1.2. ANTECEDENTES PARTICULARES

Con los reportes iniciales del FDT y la relevancia creciente de los lncRNAs en la regulación génica, en nuestro grupo de trabajo se inició un proyecto de investigación dedicado a la caracterización de este transcrito. El objetivo central del proyecto era demostrar si se trataba de un lncRNA importante para la regulación de sus genes vecinos, las α -globinas. Cabe resaltar que esta hipótesis inicial presupone un modelo de regulación en *cis* dado que parecía una pregunta más sencilla de abordar en comparación con el modelo de regulación en *trans*. Los resultados de este proyecto inicial sugieren la presencia de un RNA largo no-codificante de ~23 kb, al cual se nombró lncRNA- α GT por sus siglas en inglés: *long non-coding RNA α -globin transcript* (Arriaga-Canon et al. 2014, **ANEXO 2**). En este mismo trabajo se encontró que el lncRNA- α GT se transcribe en la misma dirección con respecto a los genes α -globina de pollo, y a su vez, muestra un patrón de expresión

específico del linaje eritroide.

Caracterización de la transcripción a lo largo del dominio α -globina de pollo

Con el objetivo de definir la zona abarcada por este transcrito (el inicio y el final de su transcripción) así como la hebra de DNA a partir de la cual era sintetizado (transcripción en dirección sentido o antisentido), se realizaron RT-PCRs de cadena específica a partir de RNA total aislado de eritrocitos de pollo de 5 y 10 días post-fecundación (*5dRBC* y *10dRBC*, respectivamente, por sus siglas en inglés: *Red Blood Cells*) con pares de oligos (*primers*) diseñados en distintas zonas a lo largo del dominio α -globina de pollo (**Figura 14**). Cabe resaltar que se diseñaron 14 diferentes *primers* (P1-P14), algunos de ellos ubicados en zonas intergénicas (P9-P14, ver **Figura 14**), y se detectó transcripción en todas las regiones analizadas por RT-PCR, aún en aquellas zonas que no coincidían con ninguna anotación génica previa (**Figura 14**).

Con este experimento se observó transcripción en dirección antisentido (**Figura 14**, flechas azules) de la zona P1 a P8, lo cual concuerda con la transcripción del gen *C16orf35/CGTHBA*. Sin embargo, a partir de la zona P6 (33 kb) se empieza a detectar señal de transcripción en dirección sentido (**Figura 14**, flechas rojas), tanto en una región que traslapa con el gen *C16orf35/CGTHBA* (*primers* P6-P8), como en regiones intergénicas río abajo (P9-P14). La transcripción sentido se extiende a lo largo de ~23 kb, aunque en este punto no se puede concluir que la transcripción detectada en diferentes regiones del dominio α -globina corresponda a un transcrito continuo.

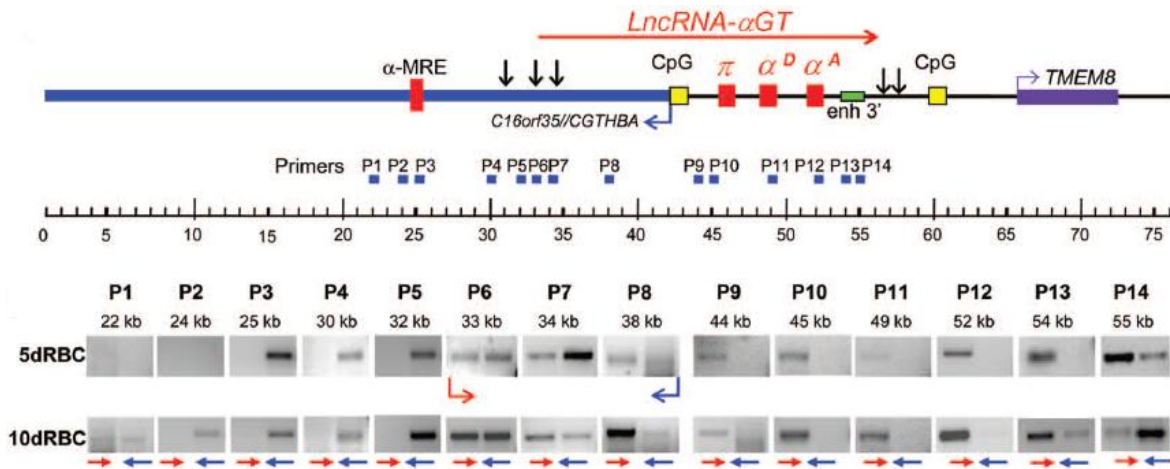


Figura 14.- RT-PCR cadena específica en diferentes regiones del dominio en 5dRBC y 10dRBC. Se diseñaron los *primers* P1-P14 a lo largo del dominio (cajas azules) cuyas posiciones se indican en la figura de arriba. El esquema del dominio α -globina de pollo se encuentra a escala (unidad = 1 kb), la cual está indicada en la parte inferior del esquema. Se extrajo RNA total de eritrocitos de pollo en las etapas de 5 y 10 días del desarrollo embrionario (5dRBC y 10dRBC, respectivamente), mismo RNA que fue sometido a RT-PCR cadena específica, lo cual permitió definir la presencia de transcritos en diferentes regiones intergénicas del dominio (P1-P14) así como su direccionalidad. La presencia de transcripción sentido y antisentido se encuentra representada por flechas rojas y azules, respectivamente. El inicio de la transcripción en dirección sentido se encuentra esquematizado con una flecha roja, alrededor de la kilobase 33 (figura tomada de Arriaga-Canon et al. 2014, **ANEXO 2**).

Para este punto se ha analizado ya la expresión en zonas intergénicas, aunque únicamente de manera cualitativa, en eritrocitos de pollo provenientes tanto de la etapa primitiva (expresión predominante del gen π) así como de la etapa definitiva de la eritropoyesis (expresión predominante de los genes adultos α^D y α^A). Se pensó en ampliar este análisis hacia otros tejidos y líneas celulares, con la intención de analizar dos aspectos: i) si esta transcripción era propia del tejido eritroide, y en cuyo caso estudiar ii) si esta transcripción podría tener alguna implicación directa en la diferenciación eritroide. Como podremos recordar, en el laboratorio contamos con líneas celulares arrestadas en diferentes estadios de la diferenciación eritroide. Lo que se hizo fue extrapolar el análisis de la **Figura 14** a las

líneas celulares eritroides HD24, 6C2, HD3 y dif-HD3 (representando en el orden mencionado los siguientes estadios de la diferenciación: progenitor mieloide-eritroide-> precursor eritroide-> eritoblasto-> eritrocito, respectivamente) utilizando como controles negativos (células no eritroides) a la línea linfóide DT40 y a células primarias de cerebro de pollo (obtenidas a partir de cerebros de embriones de pollo a 10 días del desarrollo). Además es importante recordar que en el modelo de diferenciación eritroide antes mencionado, la expresión de los genes α -globina no es detectable en células HD24, 6C2 ni HD3. Sin embargo, la inducción de la diferenciación en las células HD3 conlleva a un incremento en la expresión de las globinas adultas α^D y α^A . En la **Figura 15** se muestran únicamente los resultados obtenidos de la RT-PCR cadena específica en las zonas P4, P9 y P10 (ver **Figura 14**), siendo P4 correspondiente a la transcripción del gen *C16orf35/CGTHBA*, y P9 y P10 a la transcripción observada previamente en la zona intergénica entre *C16orf35/CGTHBA* y el gen π .

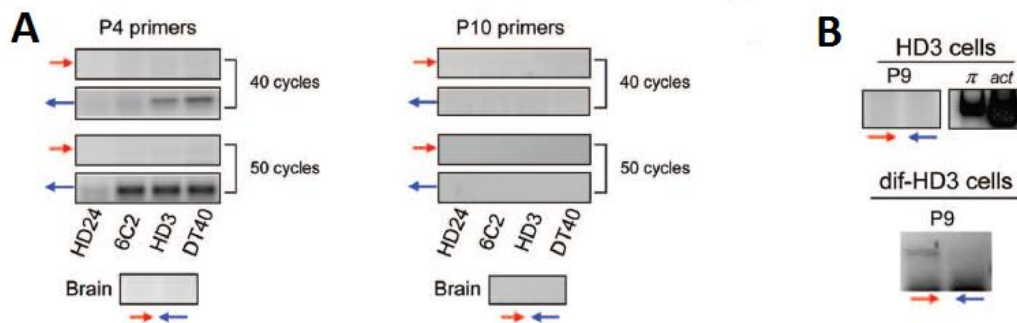


Figura 15.- RT-PCR cadena específica en diferentes estadios de la diferenciación eritroide. (A) RT-PCR cadena específica a 40 y 50 ciclos de amplificación en las zonas P4 y P10 (**Figura 14**) con RNA total de cerebro de pollo (*Brain*) y de las líneas celulares eritroides HD24, 6C2, HD3 y linfóide DT40. (B) RT-PCR cadena específica radioactiva en la zona P9 en células HD3 y HD3 inducidas a diferenciación (dif-HD3). De nuevo, las flechas azules y rojas indican transcripción antisentido y sentido, respectivamente (Tomada de Arriaga-Canon et al. 2014).

La figura muestra que el transcrito de *C16orf35/CGTHBA* se detecta en muy bajas cantidades en todas las líneas celulares analizadas (**Figura 15A**, amplificación de PCR tardía a 40 y 50 ciclos); sin embargo, en la zona P9, la transcripción en dirección sentido fue únicamente perceptible en las células HD3 diferenciadas (**Figura 15B**). Cabe resaltar que en cerebro no se detectó transcripción en ninguna de las dos zonas analizadas. Estos resultados sugieren que la transcripción en dirección sentido previamente observada en eritrocitos (*RBCs*) corresponde a transcritos propios de las etapas finales de la diferenciación eritroide, y por lo tanto correlaciona con la expresión de los genes adultos α^D y α^A .

Por otro lado, la detección de transcritos de alto peso molecular por ensayos de northern blot (**Figura 16**) utilizando distintas sondas a lo largo del dominio sugiere la existencia de transcritos de gran tamaño (mayores a 10 kb) que en principio abarcan gran parte del dominio α -globina de pollo y cuyo tamaño exacto es difícil de estimar (Arriaga-Canon, et al, 2014, **ANEXO 2**). En la **Figura 16** se muestran únicamente los resultados de northern blot obtenidos con las sondas A y C. Esta fue la primera evidencia obtenida en nuestro laboratorio de que la transcripción intergénica observada previamente (**Figura 14**) correspondía en realidad a un mismo transcrito de gran tamaño (Arriaga-Canon et al. 2014, **ANEXO 2**).

Determinación del potencial codificante y de un promotor específico del linaje eritroide

Por definición un lncRNA es un RNA de tamaño mayor a los 200 nucleótidos que no posee capacidad alguna para traducirse a proteína, razón por la cual demostrar que un RNA

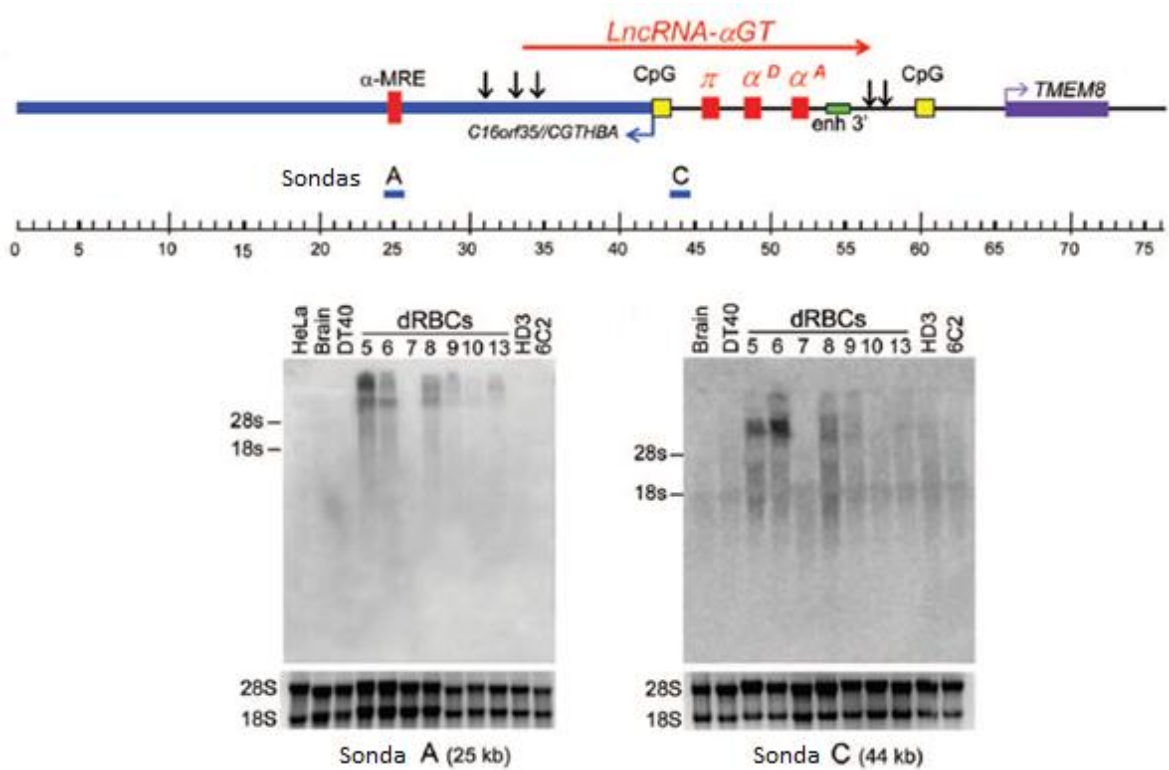


Figura 16.- Ensayos de Northern Blot en eritrocitos de pollo (RBCs) de diferentes etapas y en distintas líneas celulares. En la figura se muestran únicamente dos de las sondas utilizadas (*probes* A y C) y su localización en el dominio α-globina de pollo. La hibridación observada con la sonda A se atribuye al gen antisentido *C16orf35/CGTHBA* y corresponde, como se esperaba, con la detección de transcritos de alto peso molecular (El tamaño del 28S es aproximadamente 5 kb). Por otro lado, la sonda C corresponde a una región intergénica entre *C16orf35/CGTHBA* y el gen π , y el resultado sorprendentemente apunta a que la transcripción previamente observada en esa zona (**Figura 14**) corresponde a un transcrito largo (Figura tomada de Arriaga-Canon et al. 2014).

carece de potencial codificante se muestra como imperativo en la caracterización de un lncRNA. Por lo tanto para la evaluación del potencial codificante del RNA largo observado previamente en el dominio α-globina de pollo se utilizó el programa *Coding Potential Calculator* (CPC), el cual es capaz de discriminar con mucha precisión entre un mRNA que codifica para proteína y un verdadero RNA no-codificante (**Figura 17**). En este análisis usamos como control las secuencias de los genes codificantes *C16orf35/CGTHBA*, *TMEM8*,

π , α^D y α^A ; además, las secuencias de *Gomafu*, *Xist* y *HOTAIR* fueron utilizadas como controles validados de lncRNAs. La predicción arrojada por este programa indica que este RNA largo no parece tener potencial codificante, por lo que podría estar funcionando como un RNA largo no-codificante (**Figura 17B**, Arriaga-Canon et al. 2014, **ANEXO 2**).

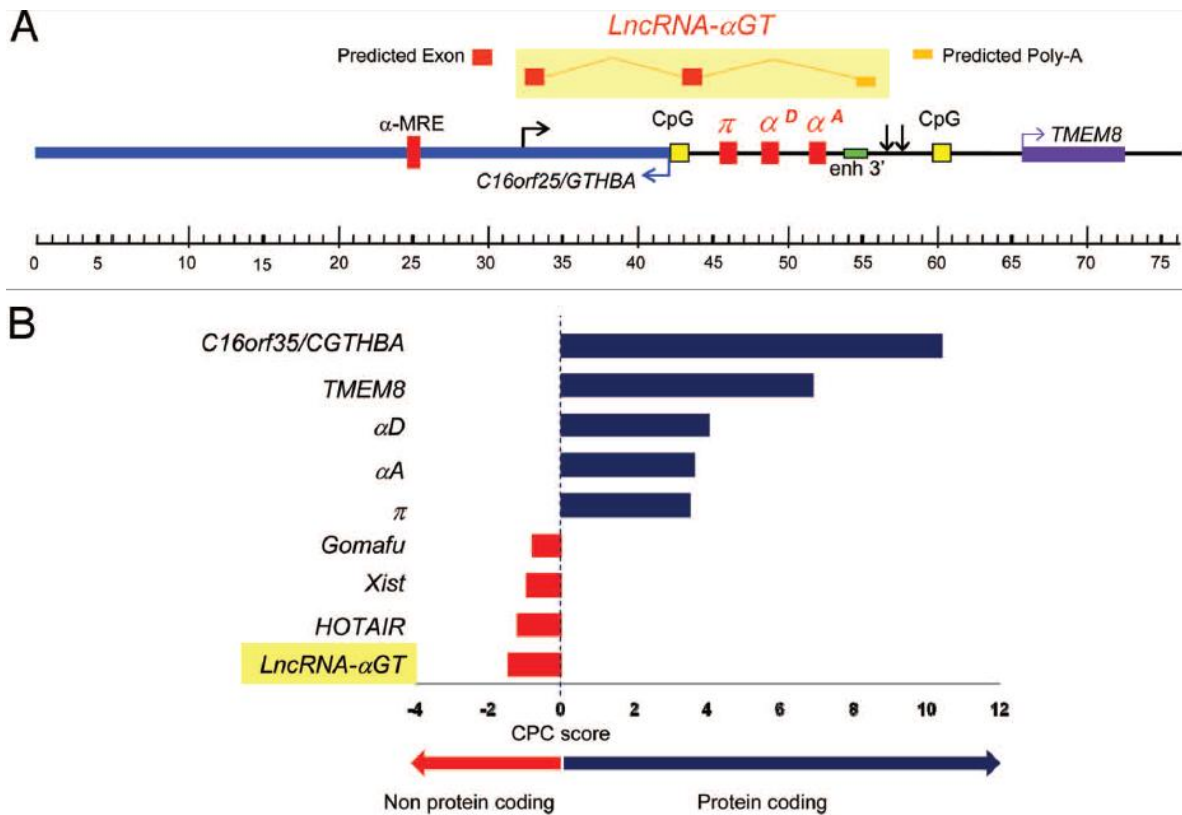


Figura 17.- Predicción *in silico* de la estructura de lncRNA- α GT y de su potencial codificante. (A) Análisis *in silico* de la estructura exón-intrón de lncRNA- α GT con los programas SSPNN y NetGen2. **(B)** Evaluación con con el programa CPC del potencial codificante de lncRNA- α GT. Se analizó la secuencia de 23 kb que es en principio abarcada por lncRNA- α GT. Todas las secuencias analizadas tienen un *score* basado en su potencial para codificar proteína: <-1, secuencia no-codificante bona fide; de -1 a 0, secuencia no-codificante débil; de 0 a 1, secuencia codificante débil; y >1, secuencia codificante bona fide (Tomada de Arriaga-Canon et al. 2014).

No obstante, y como hemos mencionado antes, los lncRNAs comparten ciertas características con los mRNAs. Ambos, tanto mRNAs como lncRNAs, poseen una

estructura exón-intrón y por lo tanto están sometidos a procesamiento o *splicing*, motivo por el cual mediante un análisis *in silico* con los programas SSPNN y NetGen2 se predijo que el lncRNA- α GT posee 2 exones (**Figura 17A**).

Con base en los experimentos anteriores, el inicio de la transcripción en dirección sentido se ubica entre las kilobases 32 y 33 del dominio α -globina (**Figura 14**). Pensando que en esta posición pudiera ubicarse el inicio de la transcripción del RNA largo no-codificante, se decidió analizar esa zona con la finalidad de encontrar evidencia en favor de la presencia de un promotor para este lncRNA en esa región. Una vez que mediante ensayos de RT-PCR cadena específica se delimitó de manera más precisa el inicio de la transcripción sentido (Arriaga-Canon et al. 2014), se decidió analizar la actividad promotora de un fragmento de DNA de 1 kb utilizando a la luciferasa como gen reportero (**Figura 18**). El resultado indica que el fragmento de DNA de 1 kb analizado (**Figura 18A**, región representada con una caja roja alrededor de la kilobase 32 del dominio) es capaz de promover la transcripción de la luciferasa cuando se clona en dirección 5'-3' en un contexto eritroide (línea celular HD3), y no así cuando se clona en la dirección contraria 3'-5' (**Figura 18B**). Los resultados muestran además que la actividad promotora de este fragmento de DNA parece ser específica del tejido eritroide, ya que el mismo experimento se realizó en fibroblastos de pollo, en donde se observó que este fragmento de DNA no conlleva a la expresión de la luciferasa con respecto al control (**Figura 18C**).

En resumen, proponemos que existe un RNA largo no codificante (**Figuras 16 y 17**) que se extiende a lo largo de 23 kb en el dominio α -globina de pollo (**Figura 14**), al cual se denominó lncRNA- α GT (por sus siglas en inglés: *long noncoding RNA α -globin transcript*).

A su vez, lncRNA- α GT presenta un patrón de expresión eritroide (**Figuras 15, 16 y 18**) que correlaciona con la expresión de los genes adultos α^D y α^A en un modelo de diferenciación eritroide (**Figura 15**).

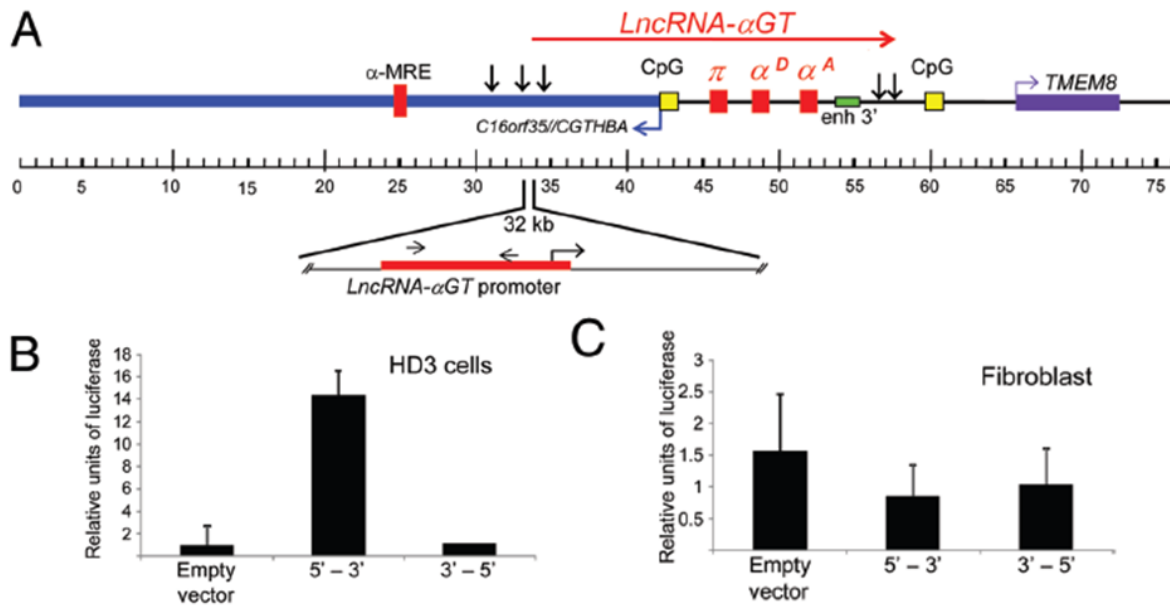


Figura 18.- Ensayo funcional de la región río arriba del inicio de la transcripción de lncRNA- α GT. (A) Esquema que muestra la posición del fragmento de 1 kb (caja roja alrededor de la kilobase 32) que fue clonado en dirección 5'-3' y 3'-5' río arriba del gen de la luciferasa con la finalidad de evaluar su capacidad promotora. (B) Transfección transitoria en células HD3 de los plásmidos con luciferasa como gen reportero analizando la capacidad promotora del fragmento de DNA de 1 kb con respecto al control (*Empty vector*, en el control no se clonó ningún fragmento río arriba de la luciferasa). (C) Mismo experimento realizado en cultivos primarios de fibroblastos de pollo (Tomada de Arriaga-Canon et al. 2014).

2. PLANTEAMIENTO DEL PROBLEMA

La presencia de un gen no-codificante, *lncRNA- α GT*, en el dominio α -globina de pollo abre nuevas preguntas, sobre todo, al nivel de su función. Cabe resaltar que existen algunos ejemplos en la literatura de transcritos no-codificantes largos que funcionan en *cis* regulando la expresión de sus genes vecinos (Bertani, et al, 2011; Wang, et al, 2011). Por lo tanto, deseamos entender la posible función del *lncRNA- α GT* en la regulación de los genes α -globina de pollo. Como se demostró previamente mediante RT-PCR en eritrocitos de pollo de 5 y 10 días del desarrollo, el *lncRNA- α GT* parece expresarse en una etapa de eritropoyesis tanto temprana como definitiva. Estos datos nos llevan a hipotetizar que si el transcrito se expresa durante esas etapas es porque juega un papel en ambas. Sin embargo, esta posibilidad queda abierta para ser comprobada. Además, la expresión del *lncRNA- α GT* en eritrocitos de distintas etapas del desarrollo fue analizada cualitativamente, lo que no nos permite realizar conjeturas más sólidas sobre su patrón de expresión durante el desarrollo de pollo desde un punto de vista cuantitativo. Debido a que se desconoce si *lncRNA- α GT* tiene un papel importante en el contexto del dominio α -globina de pollo, una valoración de su posible contribución es necesaria.

3. HIPÓTESIS

El RNA largo no-codificante α GT (lncRNA- α GT) regula la expresión de los genes α -globina de pollo.

4. OBJETIVO GENERAL

Determinar si existe una relación entre lncRNA- α GT y la expresión de los genes α -globina tanto en el contexto del desarrollo de pollo, como en la diferenciación eritroide.

OBJETIVOS PARTICULARES

- 1.- Analizar la expresión del lncRNA- α GT mediante RT-PCRs cuantitativas en distintas etapas del desarrollo del pollo.
- 2.- Estudiar el efecto de la pérdida de función de lncRNA- α GT en la expresión de los genes α -globina de pollo en células HD3 diferenciadas.
- 3.- Evaluar el efecto de la ganancia de función del lncRNA- α GT en la expresión de los genes α -globina de pollo en células HD3.

5. MATERIALES Y MÉTODOS

Cultivo celular

La línea celular HD3 corresponde a eritoblastos de pollo adulto transformados con el virus de la eritoblastosis aviar (Beug et al. 1979). Las células HD3 se cultivan en DMEM adicionado con 2% de suero de pollo y 8% de SFB (suero fetal bovino) en una atmósfera de 5% de CO₂ a 37°C. A su vez, se utilizaron cultivos primarios de eritrocitos de pollo, obtenidos mediante punción de la vena apical de embriones de pollo incubados durante distintos días, con el fin de obtener glóbulos rojos de las etapas del desarrollo requeridas (embriones de *Gallus gallus domesticus*, Alpes, Puebla, México).

Plásmidos

Para la transfección estable en la línea celular HD3 se realizó una construcción a partir del plásmido pSilencer3.1-H1 puro (Applied Biosystems) clonando el shRNA diseñado en contra de lncRNA- α GT bajo el control del promotor H1 RNA pol III por medio de una restricción BamHI/HindIII (shRNA α GT). Esta construcción posee resistencia a puromicina. Adicionalmente, y como control del experimento, se utilizó la transfección del vector pSilencer3.1-H1 puro “vacío” (Ctrl.). Para la transfección transitoria en HD3 se realizaron las siguientes construcciones: a partir del vector pcDNA3.1/Hygro (Invitrogen) se clonaron los fragmentos de 1 y 2 kb en dirección sentido (+) y antisentido (-) bajo el control del promotor CMV en los sitios BamHI y NotI, respectivamente. Como control del experimento (MOCK), se utilizó la transfección del vector pcDNA3.1/Hygro/lacZ

(Invitrogen).

Inmunoprecipitación de la cromatina (ChIP)

Los eritrocitos de pollo una vez colectados se lavan 2 veces con PBS suplementado con 2% de SFB. Se fijan las células en PBS con formaldehído al 1% y se incuban a temperatura ambiente durante 10 minutos. Se añade glicina a una concentración final de 125 mM para detener la fijación y se lavan las células dos veces con PBS frío. Posteriormente, las células se resuspenden en buffer de lisis (1% SDS, 5mM EDTA, 50 mM) suplementado con diversos inhibidores de proteasas (SIGMA). Se sónica la muestra con el fin de fragmentar la cromatina en fragmentos de 300-600 pb (este rango se evaluó por electroforesis antes de proseguir con el experimento). Cada reacción de inmunoprecipitación se realiza en un volumen final de 1 mL en un buffer que contiene 1% de Tritón X-100, 2 mM de EDTA, 20 mM de Tris-HCl pH 8.1, 150 mM de NaCl y diversos inhibidores de proteasas, y se utilizan 800 µg de cromatina y 3 µg de anticuerpo por reacción. La inmunoprecipitación (cromatina+anticuerpo) se incuba toda la noche a 4°C en un agitador giratorio. Los anticuerpos utilizados fueron anti-H3K4me3 (Millipore), anti-H3K27me3 (Abcam). A su vez, se utilizaron como controles inmunoprecipitaciones con anticuerpos inespecíficos, IgG de conejo (Millipore) e IgG de ratón (Millipore). Posteriormente, se agregaron a cada muestra 50 µL perlas magnéticas recubiertas de proteína G (Invitrogen). Las perlas se recuperaron por centrifugación durante 1 minuto a 2400 rpm a 4°C. El material recuperado con las perlas se purificó utilizando proteinasa K, y el DNA recuperado en cada

reacción se analizó mediante PCR cuantitativa (*Platinum Taq DNA polymerase*, Invitrogen) utilizando SYBR Green (SIGMA). Los enriquecimientos se calcularon mediante el método $\Delta\Delta C_t$ normalizando contra una región de no enriquecimiento (H1) y contra el control de inmunoprecipitación inespecífica (IgG).

Extracción de RNA y RT-PCR cuantitativa (qRT-PCR)

Para todos los análisis de expresión de este trabajo se procedió a extraer RNA total de las células de interés (eritrocitos de diferentes etapas del desarrollo embrionario, células HD3 y HD3 diferenciadas) resuspendiendo en 1 mL de Trizol (Invitrogen). El RNA obtenido se trató con DNasa I (*RQ1*, Promega) previamente a su utilización en la Reverso Transcripción. Para preparar cDNA, se partió de 1 μ g de RNA utilizando “*random primers*” (*Reverse Transcription System*, Promega). El cDNA fue analizado mediante PCR cuantitativa (*Platinum Taq DNA polymerase*, Invitrogen) utilizando SYBR Green (SIGMA). En el caso de los análisis de expresión en eritrocitos, se utilizó como normalizador al RNAr 28S, y para las HD3 y HD3 diferenciadas se utilizó ACTB (β -actina). Esta diferencia en el normalizador utilizado para cada condición se debe a que los cálculos de expresión son más precisos a una menor diferencia de C_t entre el gen blanco y el normalizador. Si se comparan los niveles de expresión de los genes α -globina adultos entre eritrocitos y células HD3 diferenciadas, observamos que a pesar de que la inducción produce un incremento de los mismos con respecto a las células HD3 no inducidas, la expresión en eritrocitos sigue siendo por mucho, mayor a la observada en células HD3 diferenciadas.

Esto significa que aunque el 28S es un buen normalizador dentro del contexto de eritrocitos, no necesariamente funcionará como un buen normalizador dentro de otro contexto (HD3) en dónde los niveles de expresión difieran en mayor medida.

Diseño de un shRNA específico en contra del lncRNA- α GT

A continuación, hablaremos del diseño del siRNA específico para lncRNA- α GT, tanto del procedimiento que se siguió, como de las consideraciones que se tomaron en cuenta:

- En primer lugar, se procedió a la elección de la zona blanco. Como podremos recordar, el análisis *in silico* predijo un procesamiento del lncRNA- α GT en dos exones (**Figura 17A**). Ya que el primer exón traslapa con el gen antisentido *C16orf35/CGTHBA*, garantizar el diseño de un shRNA específico en contra del lncRNA- α GT (y por lo tanto, dirigido de manera específica a la cadena sentido) es difícil. Por lo tanto, se decidió diseñar un shRNA dirigido en contra del segundo exón predicho, cuya naturaleza es intergénica. Este exón putativo mide 1,103 nucleótidos, y sus coordenadas dentro del dominio α -globina de pollo son 43,691-44,794 (NCBI AY016020).
- Existen hoy en día múltiples programas bioinformáticos para el diseño de siRNAs. Estos programas predicen varios posibles siRNAs a partir de una secuencia genómica de interés. Los programas utilizados fueron: *siRNA at Whitehead* del MIT (<http://jura.wi.mit.edu/bioc/siRNAext/>) y *BlockIt RNAi Designer* de Invitrogen (<https://rnaidesigner.invitrogen.com/rnaiexpress/>). Se dio como *input* a ambos programas

la secuencia del exón 2, antes especificada, los cuales predijeron muchos siRNAs que fueron sometidos a varios filtros para una elección final. Como primer paso, se trabajó con todos aquellos siRNAs predichos por ambos programas.

- Para la elección del siRNA más adecuado se tomaron en consideración algunas características de secuencia propias de los siRNAs. La primera de ellas fue “la regla de la termodinámica asimétrica” entre hebras (Khvorova et al. 2003; Schwarz et al. 2003). En general se ha visto que la hebra con el extremo 5’ menos estable (lo cual se atribuye a apareamientos más débiles) es preferencialmente ensamblada en el complejo RISC (Hutvagner, 2005). Por lo tanto, en el filtro de los candidatos obtenidos previamente *in silico* se dio prioridad a aquellos con un extremo 5’ enriquecido en nucleótidos A+U (posición 1 - 7) (Ui-Tei et al. 2004; Pei y Tuschl 2006).
- De manera congruente con el paso anterior, múltiples análisis de secuencia de siRNAs han identificado de manera independiente nucleótidos individuales ubicados de manera preferencial en ciertas posiciones. Por ejemplo: se ha observado que la mayoría de los siRNAs contienen una A ó U en la posición 1 (Reynolds et al. 2004; Jagla et al. 2005) y en la posición 10 (Takasaki et al. 2004; Shabalina et al. 2006).
- El contenido de G+C en los siRNAs suele ser de 30%-52% (Chalk et al. 2004; Reynolds et al. 2004). Se ha visto que contenidos de G+C muy bajos correlacionan con siRNA dúplex poco estables y con una afinidad de apareamiento menor hacia los transcritos blanco. Por otro lado, contenidos de G+C muy altos pueden impedir el ensamblaje de una cadena guía en el complejo RISC.

- De manera adicional, la presencia de estructuras secundarias locales en los transcritos blanco puede restringir la accesibilidad de [RISC + cadena guía] a la zona complementaria, y por lo tanto disminuir la eficiencia del *knockdown* (Luo y Chang, 2004; Schubert et al. 2005). Por esta razón, este criterio se muestra como un filtro importante dentro de nuestro análisis, motivo por el cual se analizaron *in silico* las posibles estructuras secundarias presentes en la secuencia blanco, que en nuestro caso es la región que corresponde al exón 2 de lncRNA- α GT predicho previamente. Este análisis se llevó a cabo utilizando el algoritmo PITA (Kertesz et al. 2007) (http://genie.weizmann.ac.il/pubs/mir07/mir07_prediction.html). Este programa acepta como *inputs* a la secuencia blanco de nuestro interés (exón 2) y a una lista de siRNAs candidatos (los que obtuvimos previamente con los dos programas de diseño de siRNA). Después realizó predicciones de estructuras secundarias a lo largo del exón 2, lo cual permite predecir regiones de accesibilidad dentro del mismo, y en función de esto, ordenó la lista de siRNAs candidatos del más al menos efectivo.

Al final, dentro de una lista considerable de siRNAs candidatos en contra del lncRNA- α GT, se escogió únicamente a uno de ellos. Todos los criterios anteriormente mencionados se tomaron en cuenta para la elección y finalmente el siRNA elegido satisfizo cada uno de ellos. La secuencia de siRNA elegida fue predicha por ambos programas bioinformáticos previamente utilizados, con un extremo 5' enriquecido en A+U, siendo el nucleótido de la posición 1 una A, y el nucleótido de la posición 10 un U. Su contenido de G+C es de 42.86%, lo cual indica que está dentro del rango aceptable. Además, este siRNA alcanzó los mejores *scores* con el algoritmo PITA. La secuencia del siRNA elegido es: 5'-

AUGAAGCUGUGAGAUGUCC-3'.

Transfecciones en HD3

Todas las transfecciones se hicieron por lipofección (Invitrogen). Por cada transfección se utilizaron 1×10^6 células que se resuspendieron en 800 μ L de DMEM sin suero y se colocaron en placas de 6 pozos. Se transfectó en todos los casos 1 μ L de plásmido. En el caso de las transfecciones estables para el experimento de pérdida de función, las construcciones a transfectar (vector vacío y vector shRNA α GT) fueron previamente linearizadas con la enzima de restricción XmnI. Una vez transfectados los plásmidos, se dejó incubar a 37°C durante 6 horas antes de inactivar con suero. Posteriormente, se dejaron recuperando las células durante 48 horas antes de ser sometidas a selección con puromicina (SIGMA) a una concentración final de 1 μ g/mL. En el caso de las transfecciones transitorias, los plásmidos a transfectar (pcDNA3.1_1kb(+), pcDNA3.1_1kb(-), pcDNA3.1_2kb(+), pcDNA3.1_2kb(-), pcDNA3.1/lacZ) no se linearizaron previamente y a las 48 horas de recuperación se extrajo RNA para los análisis de expresión.

Diferenciación de células HD3

Las células HD3 se encuentran arrestadas en el estadio de eritoblasto, pero dicho arresto puede ser revertido mediante la inactivación del virus de la eritoblastosis aviar incubando a 42°C durante 48 horas. Adicionalmente, la eficiencia de la diferenciación puede incrementarse suplementando el medio de cultivo con el fármaco ISO-H-7 a una

concentración final de 20 μ M (SIGMA) y HEPES 10 mM (pH 8.0), en una atmósfera con 1% de CO₂ (Nicolas et al. 1991).

Tabla de secuencias de oligonucleótidos

Región amplificada		Secuencia 5'-3'	Tamaño del amplicón
(1)Promotor α^D	Forward	CTGCTCTCAGCATTGCACAG	129pb
	Reverse	GGGCTGGTGGCTGGTGG	
(2)Promotor lncRNA- α GT	Forward	CTTCACAACATTGGCCTGGTC	224 pb
	Reverse	GGTGTCTCGGAAGCTGAC	
(3)P8	Forward	AGAAAGGGACATTCACGTCC	161 pb
	Reverse	GCACATGAAGAAACCTGTTGC	
(4)HI	Forward	GGAACAAGTTGGCAAGGTCCTAT	145 pb
	Reverse	TCTTCTGCCCTGCCCGTAT	

(5)mRNA π (oligos exón-exón)	Forward	TCACTGGAGAGGCTTTTTGCC	227 pb
	Reverse	GTGGGAAAGCAGCTTGAAGTT	
(6)mRNA α^D (oligos exón-exón)	Forward	GCTCTGACTAGGATGTTCCACC	225 pb
	Reverse	CTGCGACAACAGCTTGAATTG	
(7)mRNA α^A (oligos exón-exón)	Forward	CCCTGGAAAGGATGTTCCACC	224 pb
	Reverse	GGCCCAGGAGTTTGAAGTTG	
(8)lncRNA-αGT (región P9, Figura 14)	Forward	GCGAGACACGGCAGGACAGG	143 pb
	Reverse	GCACCTCAGTGCAGGCTGTGT	
(9)28S	Forward	CCTAGCAGCCGACTTAGAACTG	327 pb
	Reverse	GCTCAACAGGGTCTTCTTCC	
(10)ACTB	Forward	CCCAGACATCAGGGTGTGATGG	120 pb
	Reverse	CCAGTTGGTGACAATACCGTGTTCC	
(11)Región 1 kb	Forward	AGCACCCGCAGTGAGGGCTTA	974 pb

	Reverse	CTGGAGCTCCTGGCCTACGC	
(12)Región 2 kb	Forward	ACATGCATCCATGGCACACA	2104 pb
	Reverse	CGTGGCTGTGCAATGCTGA	
(13)2 kb qPCR	Forward	GTTCCACCTACACCTTCG	159 pb
	Reverse	TGAGCAGTGACAGTCACAGG	
(14)shRNA contra lncRNA-αGT	Forward	GATCCGGACATCTCACAGCTTCA TTTCAAGAGAATGAAGCTGTGAG ATGTCCTTTTTTGAAA	63 pb
	Reverse	AGCTTTTCCAAAAAAGGACATCT CACAGCTTCATTCTTTGAAATGA AGCTGTGAGATGTCCG	

6. RESULTADOS

6.1. Expresión de lncRNA- α GT y su correlación con la estructura local de la cromatina durante el desarrollo del pollo

Estudios recientes sobre los lncRNAs han puesto en evidencia que: i) los lncRNAs presentan patrones de expresión muy bien definidos espacio-temporalmente, es decir, que en muchos casos su expresión es tejido-específica, además de estar acotada a una etapa del desarrollo y/o a un estadio en particular tanto de la diferenciación como de la división celular; ii) los lncRNAs se expresan en muy bajos niveles. Dado que se ha observado la presencia del lncRNA- α GT en varias etapas del desarrollo de pollo mediante RT-PCR convencional, lo que se busca ahora es realizar un análisis cuantitativo de su expresión durante el desarrollo de pollo que nos permita conocer en detalle sus niveles de expresión.

Para abordar estas preguntas aislamos eritrocitos de pollo a partir de sangre circulante de huevos de 5, 6, 7, 8 y 10 días post-fecundación, y posteriormente se procedió a la extracción de RNA total con trizol. Después se realizó una RT con *random primers* y el cDNA obtenido se utilizó como templado para la qPCR (*quantitative PCR*). Se utilizó como gen normalizador al 28S (oligos **(9)**, consultar tabla de oligonucleótidos en la sección de **Materiales y métodos**), y como control se analizó simultáneamente la expresión de las globinas π y α^D (oligos **(5)** y **(6)**, respectivamente, Materiales y métodos) en 5, 7 y 10 días del desarrollo embrionario. Los resultados obtenidos se muestran en la **Figura 19**.

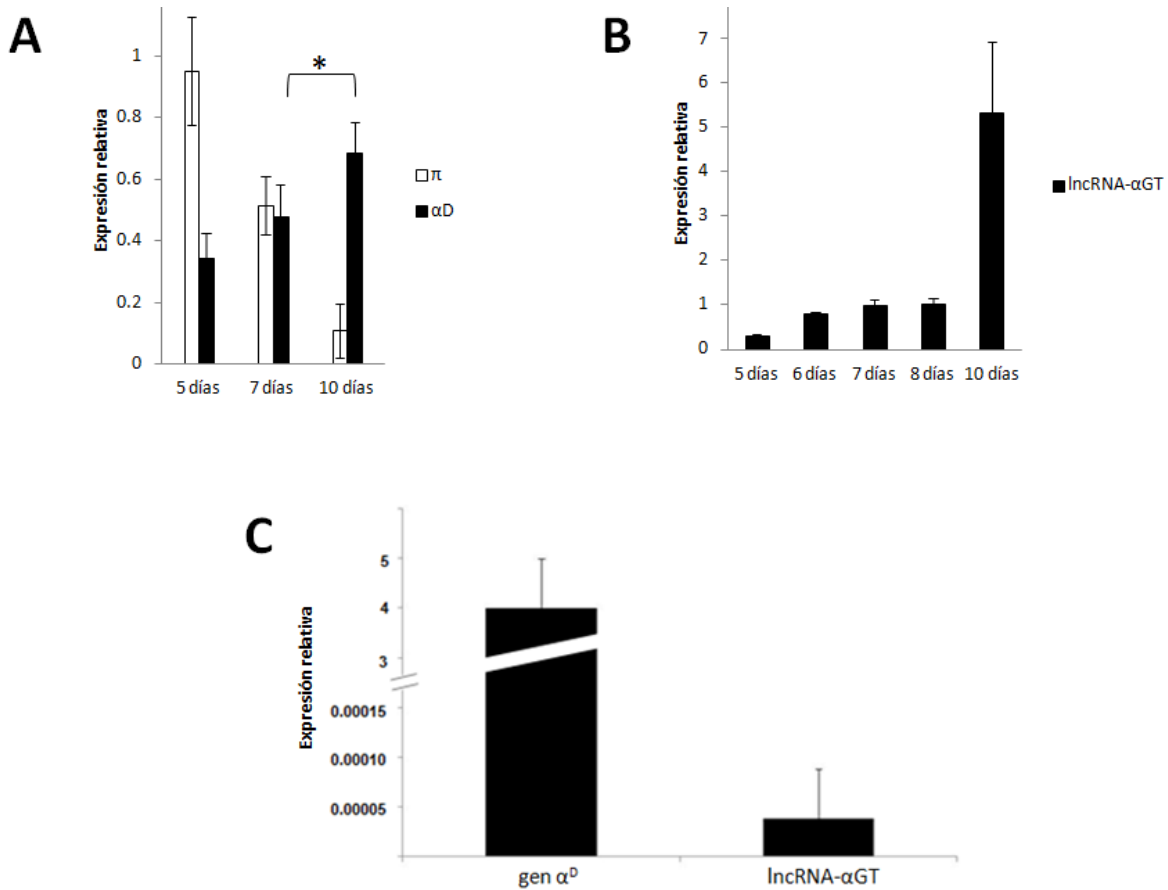


Figura 19.- Expresión de lncRNA- α GT en eritrocitos durante el desarrollo embrionario de pollo. (A) Expresión de las globinas π y α^D en eritrocitos de 5, 7 y 10 días del desarrollo analizada por qRT-PCR. **(B)** Expresión de lncRNA- α GT (oligos **(8)**, Materiales y métodos) durante diferentes días del desarrollo embrionario analizada por qRT-PCR. **(C)** Expresión relativa de α^D y lncRNA- α GT en eritrocitos de 10 días. Mediante este experimento se estimó que lncRNA- α GT se expresa alrededor de $\sim 130,000$ veces menos que el mRNA de α^D . En estas gráficas se muestra la expresión promedio y las barras de error corresponden a las desviaciones estándar de dos diferentes experimentos. * indica un valor $P < 0.05$ (**Figura 19A**).

En la **Figura 19A** se muestra que la globina embrionaria π presenta niveles altos de expresión en eritrocitos de 5 días, mismos que disminuyen a lo largo del desarrollo de pollo. En la figura es claramente apreciable el momento de recambio (7 días), etapa que se caracteriza por la transición de eritropoyesis embrionaria a eritropoyesis definitiva, y que al nivel de expresión génica se entiende como el reemplazo de las globinas

embrionarias, en este caso π , por las globinas adultas α^D y α^A . Por otro lado observamos que lncRNA- α GT niveles bajos de expresión en etapas más tempranas del desarrollo embrionario (de 5 a 8 días), y que posteriormente sus niveles aumentan considerablemente en eritrocitos de 10 días (**Figura 19B**). A su vez, se determinó que los niveles de expresión de lncRNA- α GT son bajos en comparación con los de α^D , ya que en la etapa de 10 días, lncRNA- α GT es $\sim 130,000$ veces menos abundante que el mRNA codificante α^D (**Figura 19C**). Este cálculo se obtiene dividiendo la expresión promedio del mRNA de α^D entre la expresión promedio del lncRNA- α GT obtenidas en la etapa de eritrocitos de 10 días.

Con el objetivo de validar estos datos, así como la caracterización del promotor de lncRNA- α GT hecha previamente (ver **Antecedentes particulares**, Arriaga-Canon et al. 2014, **ANEXO 2**), se decidió analizar el potencial de activación del promotor durante el desarrollo. Diferentes estudios en la literatura han documentado que las modificaciones post-traduccionales de las histonas están íntimamente relacionadas con la actividad promotora. En particular, la presencia de trimetilación en la lisina 4 de la histona H3 (H3K4me3) en el promotor de un gen se asocia con un estado de activación transcripcional, e incluso, el enriquecimiento de esta modificación post-traduccionales de histonas ha sido un criterio ampliamente utilizado en la identificación y anotación a nivel global de promotores activos (Schneider, et al, 2004; Bernstein, et al, 2005; Pokholok, et al, 2005). Como se mencionó antes, se procedió a analizar el estado de los promotores de lncRNA- α GT y α^D mediante el análisis del enriquecimiento de la H3K4me3. Esto se realizó por CHIP (por sus siglas en inglés: *Chromatin immunoprecipitation*), método utilizado para

conocer los sitios genómicos ocupados por una proteína de nuestro interés mediante una inmunoprecipitación de la cromatina con el anticuerpo anti-H3K4me3. A su vez, se decidió incluir en el análisis a una región no promotora para fines comparativos (oligos P8 **(3)**, ver Materiales y métodos). En la **Figura 20A** se muestra la ubicación en el dominio α -globina de cada una de las zonas analizadas, así como sus correspondientes enriquecimientos en la marca de histona H3K4me3 (**Figura 20B**). El enriquecimiento de esta marca se evaluó mediante PCR cuantitativa (qPCR) normalizando contra una zona de no enriquecimiento; en este caso, se utilizó a la zona HI del dominio β -globina de pollo que se ubica dentro de una región de heterocromatina (oligos **(4)**, Materiales y métodos), por lo que se espera que la H3K4me3 no esté enriquecida. Además, el enriquecimiento se mide a partir del “ruido de fondo” del experimento, que en este caso se dedujo a partir de la inmunoprecipitación de la cromatina (ChIP) con un anticuerpo inespecífico (para este experimento se utilizó IgG de conejo). Estos dos últimos controles (normalización en contra de la zona HI y sustracción del ruido de fondo) son necesarios para calcular el enriquecimiento de la marca de histonas H3K4me3 en una zona determinada mediante qPCR. Además, en este experimento se analizó también el enriquecimiento de otra marca de histonas, la trimetilación de la lisina 27 en la histona H3 (H3K27me3), modificación propia de heterocromatina facultativa (**Figura 20C**). La marca H3K27me3 es antagónica a la H3K4me3, y su enriquecimiento, por ejemplo, en un promotor, se asocia con silenciamiento génico (Li et al. 2007). Cada una de las gráficas mostradas en las **Figuras 20B y 20C** contiene una “línea de base” o umbral (valor de enriquecimiento = 1). Todos los valores observados por encima de la línea punteada son considerados como enriquecidos,

ya sea en mayor o menor medida, con respecto al ruido de fondo. Todos los resultados mostrados en la **Figura 20** corresponden a ensayos de ChIP realizados en eritrocitos de embriones de pollo de 10 días.

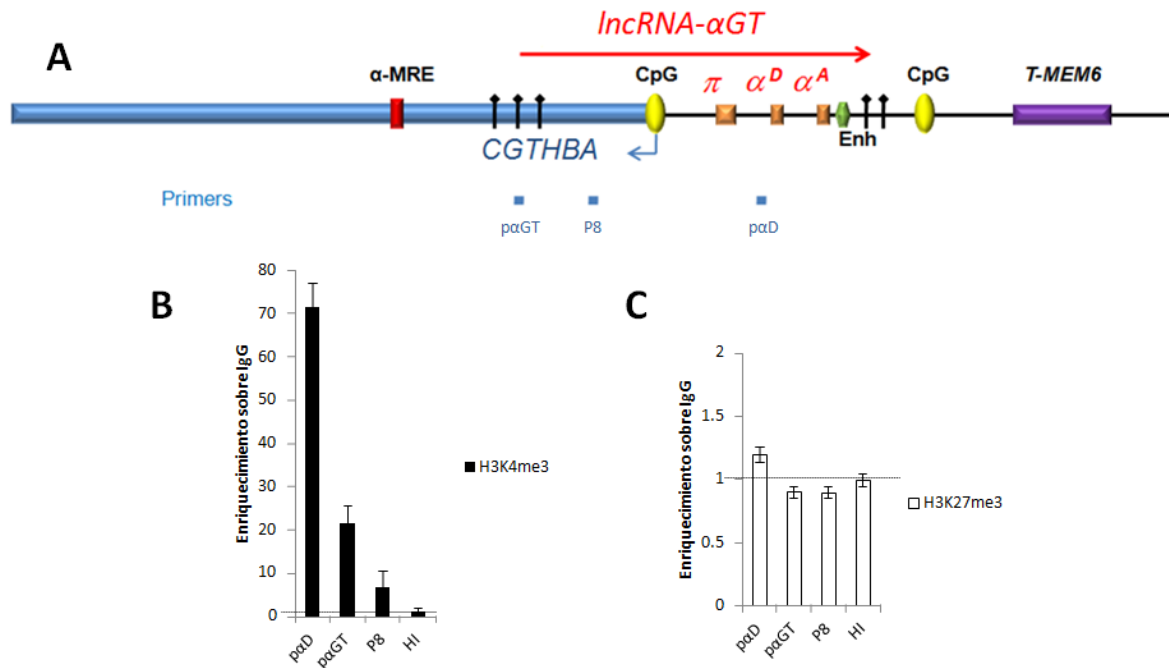


Figura 20.- Enriquecimiento de H3K4me3 y H3K27me3 en eritrocitos de 10 días del desarrollo. (A) Esquema del dominio α -globina de pollo que muestra la localización de las diferentes zonas analizadas (“primers”, cajas azul claro): α GT, P8 y α D (oligos (2), (3) y (1), respectivamente, consultar en Materiales y métodos). (B) ChIP-qPCR con el anticuerpo anti-H3K4me3 en eritrocitos de 10 días del desarrollo. Los datos se normalizaron contra la zona HI (región de heterocromatina del dominio β -globina de pollo), misma que se incluyó en la gráfica con fines comparativos. (C) ChIP-qPCR con el anticuerpo anti-H3K27me3 en eritrocitos de 10 días del desarrollo.

Los resultados de este experimento son congruentes con la previa caracterización del promotor del *IncRNA- α GT* (ver **Figura 18**), ya que se observó un enriquecimiento considerable de la modificación de histonas propia de promotores activos H3K4me3 en el α GT con respecto a HI (región de heterocromatina) y P8 (región no promotora, **Figura 20B**). A su vez, y también de manera congruente con los análisis de expresión previos

(Figura 19A), se observó un enriquecimiento de la histona H3K4me3 aún mayor en el promotor del gen α^D . Por otro lado, el ChIP para la marca H3K27me3, asociada a heterocromatina facultativa, reveló que esta modificación post-traducciona l en las histonas presenta niveles mínimos de enriquecimiento en todas las zonas analizadas (Figura 20C).

Por otra parte, el enriquecimiento de la histona H3K4me3 en el promotor de lncRNA- α GT indica que se trata de un promotor activo, dato congruente con los resultados previos que indican que es justo en la etapa de 10 días en dónde se observa una mayor transcripción de lncRNA- α GT (Figura 19B). Asimismo, se analizó el enriquecimiento de la histona H3K4me3 en estas mismas regiones en etapas previas del desarrollo (eritrocitos de 5 y 7 días; Figura 21).

Por una parte, lo que observamos al comparar 3 diferentes etapas del desarrollo de pollo es que, si bien la marca de histona H3K4me3 parece estar enriquecida en todas las etapas y en todas las zonas analizadas con respecto al ruido de fondo (línea punteada), las zonas p α GT y p α D muestran un comportamiento dinámico a lo largo del desarrollo (Figuras 21A y 21C). Asimismo, en etapas más tempranas del desarrollo (5 y 7 días) el promotor de lncRNA- α GT muestra un enriquecimiento basal de la marca de histona H3K4me3, mismo que se dispara considerablemente en el día 10 del desarrollo (Figura 21A). Por otro lado, el promotor del gen α^D muestra niveles similares de enriquecimiento en la histona H3K4me3 tanto en el día 5 (eritropoyesis primitiva, niveles de expresión de α^D bajos) como en el día 10 (eritropoyesis definitiva, nivel máximo de expresión de α^D), así como una disminución notoria en 7 días (Figura 21C). Esto contrasta con los enriquecimi-

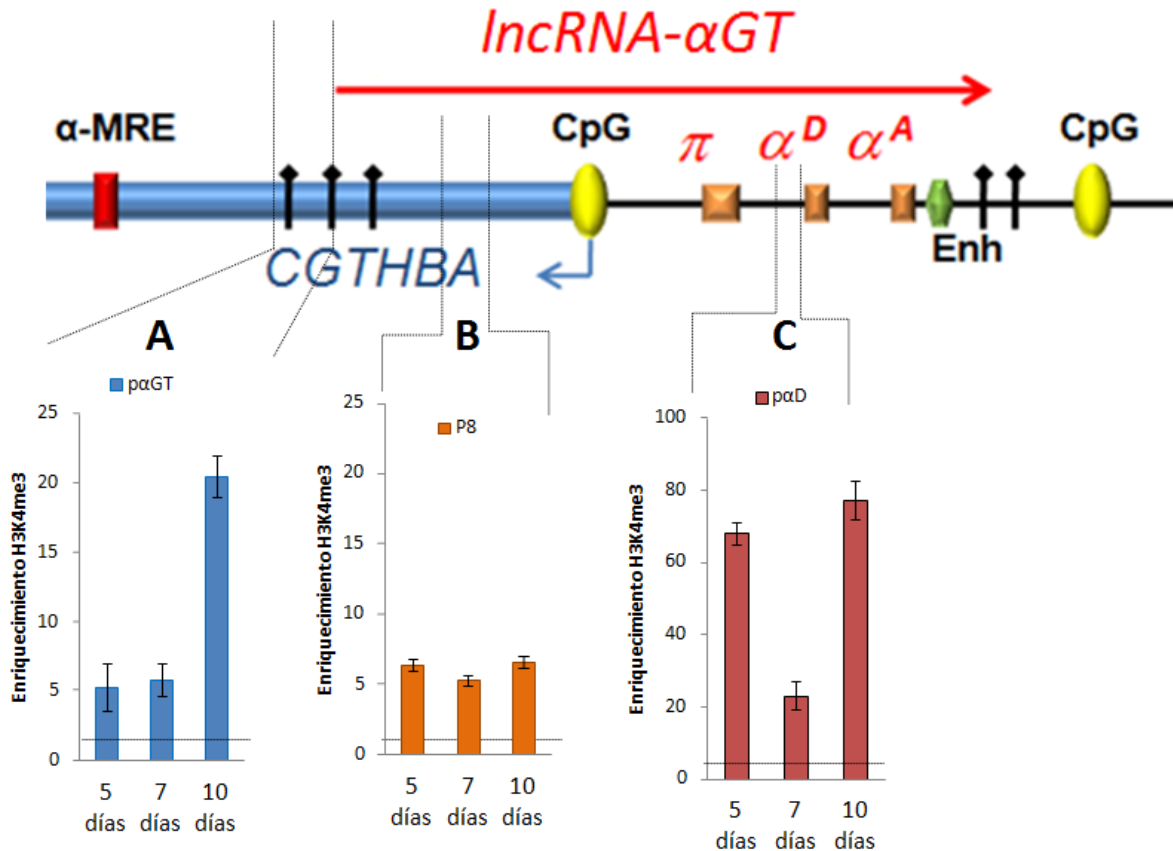


Figura 21.- Enriquecimiento de la H3K4me3 en los promotores de *IncRNA- α GT* y α^D durante el desarrollo de pollo. Se realizó ChIP utilizando un anticuerpo anti-H3K4me3 en eritrocitos de 5, 7 y 10 días del desarrollo, y se analizó el enriquecimiento de la histona H3K4me3 en los promotores de (A) *IncRNA- α GT* ($\rho\alpha$ GT) y (C) α^D ($\rho\alpha$ D) mediante qPCR. (B) En el análisis se incluyó a la zona P8 como un control negativo, ya que esta región no coincide con ningún promotor anotado previamente.

Entos de la marca de histona H3K4me3 observados para la zona control P8, en donde los enriquecimientos para la histona H3K4me3 no solo fueron mínimos (Figura 20B), sino además se mantuvieron constantes a lo largo del desarrollo (Figura 21B), lo que refuerza la idea de que $\rho\alpha$ GT, al igual que $\rho\alpha$ D, es un promotor regulado de manera diferencial a lo largo del desarrollo.

Por lo tanto, los resultados obtenidos en esta primera parte del proyecto permiten concluir que el promotor del *IncRNA- α GT* ($\rho\alpha$ GT) es activo mayormente a los 10 días del

desarrollo embrionario, ya que el enriquecimiento observado para la histona H3K4me3 es mayor en esa etapa (**Figura 21A**). Cabe resaltar que este resultado es consistente con la expresión del transcrito, para la cual se observó que hay una mayor cantidad de lncRNA- α GT en eritrocitos de 10 días (**Figura 19B**). Por otra parte, aunque los niveles de expresión del gen α^D en 5 días son menores con respecto a sus niveles de expresión en eritrocitos de 10 días, su promotor ($p\alpha D$) muestra niveles altos de enriquecimiento para la histona H3K4me3 en 5 días (**Figura 21C**). Cabe resaltar que este último dato, a pesar de no correlacionar con el patrón de expresión observado para el gen α^D durante el desarrollo, es consistente con otros resultados de nuestro grupo y otros que llevaron a proponer que el dominio α -globina de pollo posee una configuración de la cromatina constitutivamente abierta (Valdes-Quezada et al. 2013, **ANEXO 1**)

6.2. Determinación de la función del lncRNA- α GT

6.2.1. Ensayo de pérdida de función del lncRNA- α GT en células de pollo HD3 diferenciadas

Con base en los datos obtenidos previamente en este proyecto se puede deducir que si lncRNA- α GT lleva a cabo alguna función de relevancia en los eritrocitos muy probablemente sea en la etapa de la eritropoyesis definitiva (eritrocitos de 10 días de desarrollo), ya que es en esta etapa dónde se demuestra que tanto los niveles de expresión, así como la actividad promotora del lncRNA- α GT son mayores (**Figuras 19B y 21A**). Además, y de manera congruente con estos resultados, la correlación positiva entre

la expresión del lncRNA- α GT y las globinas adultas se había observado previamente en el contexto de la diferenciación eritroide, particularmente, en la etapa terminal (células HD3 diferenciadas, símil de eritrocitos adultos terminalmente diferenciados, **Figura 15B**). Debido a su localización genómica y tamaño (parece abarcar gran parte del dominio α -globina, incluso traslapando con genes codificantes), así como su expresión espacio-temporal (expresión propia de los eritrocitos adultos), el lncRNA- α GT se presenta como un elemento importante en el estudio del dominio, tanto a nivel de la regulación como estructural. Con base en las evidencias previas, podemos sugerir que la posible función del lncRNA- α GT podría estar relacionada de manera específica con las α -globina adultas α^D y α^A . Para determinar si existe una contribución de lncRNA- α GT en la expresión de los genes α -globina, se realizaron experimentos de pérdida de función del lncRNA- α GT. En este trabajo de tesis se realizó únicamente el experimento de pérdida de función (*knockdown*) en el contexto de la diferenciación eritroide, en particular, en las células HD3 inducidas a diferenciación. En cuanto a la técnica de *knockdown* se optó por un silenciamiento mediado por RNA de doble cadena, mejor conocido como RNA de interferencia (*RNAi*). Para ello se diseñó un shRNA (por sus siglas en inglés: *short hairpin RNA*) específico en contra del lncRNA- α GT, mismo que se clonó en un vector de sobre-expresión (pSilencer3.1/H1-puro) para ser transfectado de manera estable en la línea celular HD3. Adicionalmente, se utilizó como control de este experimento la transfección de un plásmido vacío.

Knockdown del lncRNA- α GT mediante RNAi

Antes de proseguir, recordemos de manera general cómo funciona el sistema de RNA de interferencia. El RNAi es un proceso conservado entre diferentes organismos, cuya finalidad es reducir significativamente la expresión de un gen de manera secuencial específica a través de RNA de doble cadena (Meister y Tuschl, 2004). Los siRNAs son RNAs dúplex o de doble cadena que miden 21-23 nucleótidos. Las hebras que conforman el dúplex son perfectamente complementarias entre ellas; además este RNA bicatenario presenta 2 nucleótidos adicionales no complementarios en el extremo 3' de ambas hebras. Este dúplex de RNA es un intermediario natural del procesamiento de RNAs más largos de doble cadena (*dsRNA*, por sus siglas en inglés: *double-stranded RNA*) por endonucleasas de la familia RNAsa III. Sin embargo, los siRNAs pueden ser también introducidos de forma exógena en las células utilizando métodos de transfección, por lo que pueden sintetizarse químicamente con base en un diseño previo que permita que la secuencia del siRNA sea totalmente complementaria a la del mRNA de nuestro interés, garantizando un efecto de silenciamiento específico para ese gen. El mecanismo general del RNA de interferencia se explica en la **Figura 22**. A partir del RNA dúplex, una de las dos hebras es elegida de manera preferencial para ser ensamblada en un complejo proteico conocido como RISC (por sus siglas en inglés: *RNA-induced silencing complex*). La hebra incorporada a RISC es conocida como “cadena guía”, mientras que la hebra complementaria a ésta se denomina “cadena pasajera”. Es importante resaltar que virtualmente cualquiera de las dos cadenas podría ser ensamblada a RISC, sin embargo, estudios han demostrado que en general suele existir una preferencia por alguna de ellas.

Posteriormente, la cadena pasajera es destruida una vez que la cadena guía se incorpora a RISC. Así, RISC bloquea la expresión de un gen promoviendo, mediante un apareamiento que puede ser perfecto o parcial (microRNAs), la degradación de transcritos complementarios a la cadena guía (Sontheimer, 2005).

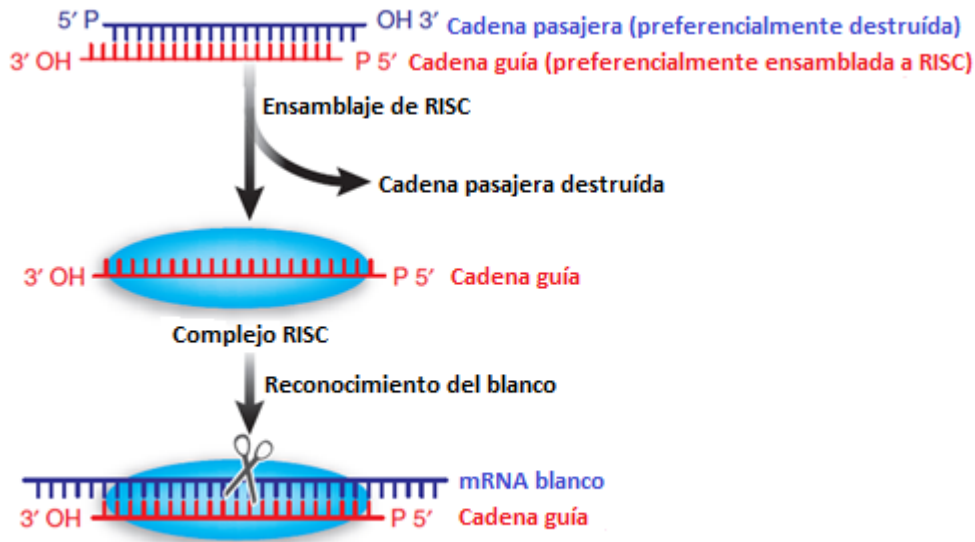


Figura 22.- Silenciamiento génico mediado por siRNAs. La asimetría en la secuencia primaria de cada una de las hebras que conforman el dúplex dicta cual es la cadena que será preferencialmente cargada en el complejo RISC. De esta manera, la cadena guía (siRNA) es ensamblada en el complejo RISC. El siRNA, a través de un apareamiento por complementariedad, promueve el reconocimiento específico de RNAs blanco. Una vez que el siRNA y su RNA diana se han apareado, el complejo RISC induce una escisión en el RNA blanco, la cual es interpretada por la maquinaria celular como una señal de degradación (Figura tomada de Pei y Tuschl, 2006).

En el contexto de esta tesis, es importante contar con un sistema de *knockdown* estable y eficiente en contra del lncRNA- α GT, por lo que se requerirá de una fuente continúa que provea a los siRNAs necesarios. Por esta misma razón, se optó por la utilización de un shRNA clonado en un vector de sobre-expresión que se transfectó de manera estable en

las células HD3, ya que los siRNAs pueden ser generados a partir del procesamiento de shRNAs expresados de manera ectópica (Bernards et al. 2006; Chang et al. 2006). Los shRNAi (*short hairpin RNA interference*) son RNAs pequeños con estructura tallo-asa que se mimetizan con intermediarios naturales del proceso de RNAi, y en consecuencia son procesados a siRNAs por la RNAsa III Dicer (**Figura 23D**). La estructura general de una secuencia de shRNAi se muestra en la **Figura 23A**. Para la generación del shRNA se mandaron a sintetizar dos oligonucleótidos de DNA complementarios entre sí de 63 nt (**Figura 23B**). Estos oligonucleótidos se aparean para formar DNA de doble hebra y posteriormente se clonan bajo un promotor de interés. En nuestro caso, se utilizó el promotor de expresión constitutiva H1 RNA pol III (**Figura 23C**). Una vez transfectados de manera estable, la producción del shRNA, y por lo tanto del siRNA en contra del lncRNA- α GT, será constante.

Una vez con el shRNA en contra del lncRNA- α GT clonado en el vector de expresión (shRNA α GT), se procedieron a realizar las transfecciones estables en las células HD3 por medio de lipofección. Una vez seleccionadas las células HD3 mediante puromicina, se indujeron a diferenciación con el fármaco ISO-H-7, incubando a 42° durante 48 horas. Posteriormente, se extrajo RNA total de estas células y se procedió a realizar una Transcripción reversa (RT) con *random primers*. A su vez, con el cDNA obtenido, se evaluó la transcripción, tanto del lncRNA- α GT como de los genes α -globina de pollo mediante PCR cuantitativa (qPCR). Los resultados obtenidos se normalizaron mediante el método de $\Delta\Delta C_t$ y se muestran en la **Figura 24**.

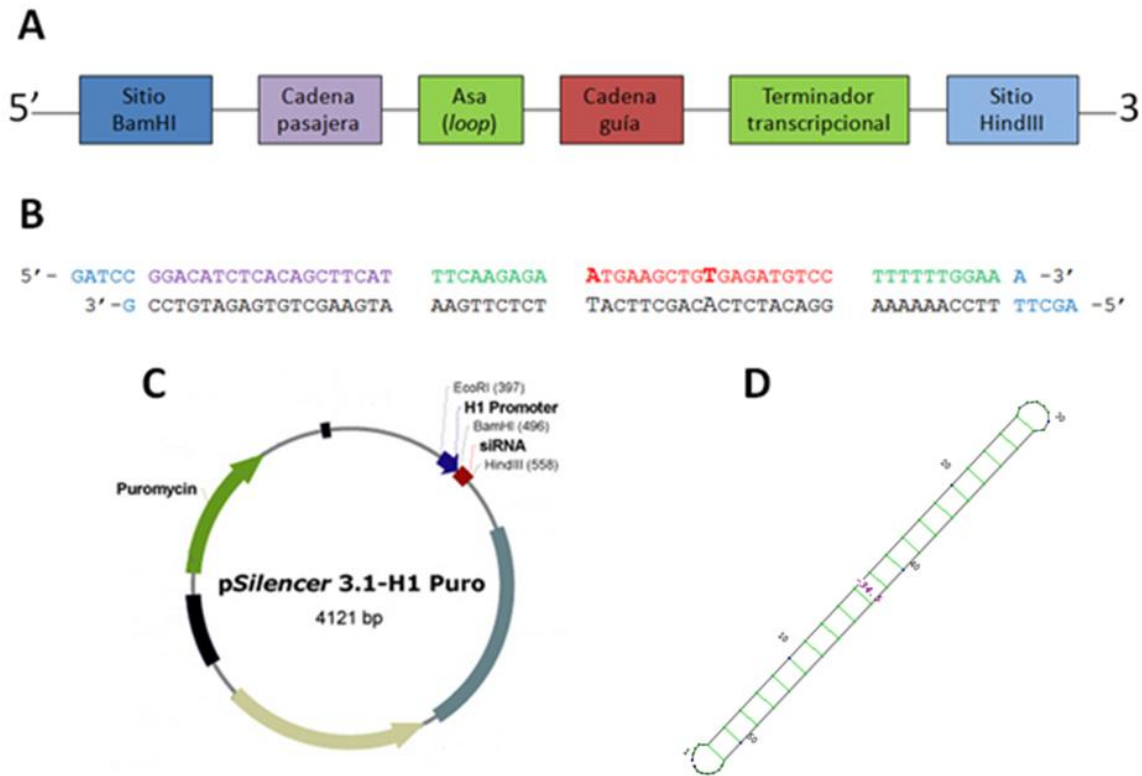


Figura 23.- Estructura y secuencia del shRNAi diseñado en contra del lncRNA- α GT. (A) Estructura general de un shRNAi. La cadena guía es la secuencia reverso complementaria de la cadena pasajera; además es muy importante enfatizar la presencia de un *loop* entre ellas con una secuencia no complementaria, de tal manera que se promueva la formación de una estructura secundaria tallo-asa. (B) Secuencia de los 2 oligonucleótidos mandados a sintetizar para la generación del shRNA (secuencia **(14)**, consultar “Tabla de oligonucleótidos” en Materiales y métodos). (C) Vector de sobre-expresión de shRNAi. La secuencia mostrada en la **Figura 23B** se clonó en los sitios BamHI-HindIII. (D) Estructura secundaria tallo-asa predicha bioinformáticamente a partir de la secuencia de la hebra 5'-3' mostrada en la **Figura 23B**.

Los resultados muestran que el *knockdown* en contra del lncRNA- α GT funciona (shRNA α GT). Además y de manera relevante, la disminución del lncRNA- α GT afecta la expresión de la globina adulta α^D , pero no la expresión de α^A y π (**Figura 24**), lo cual sugiere que el transcrito lncRNA- α GT es un elemento importante dentro de la regulación de la expresión del gen adulto α^D .

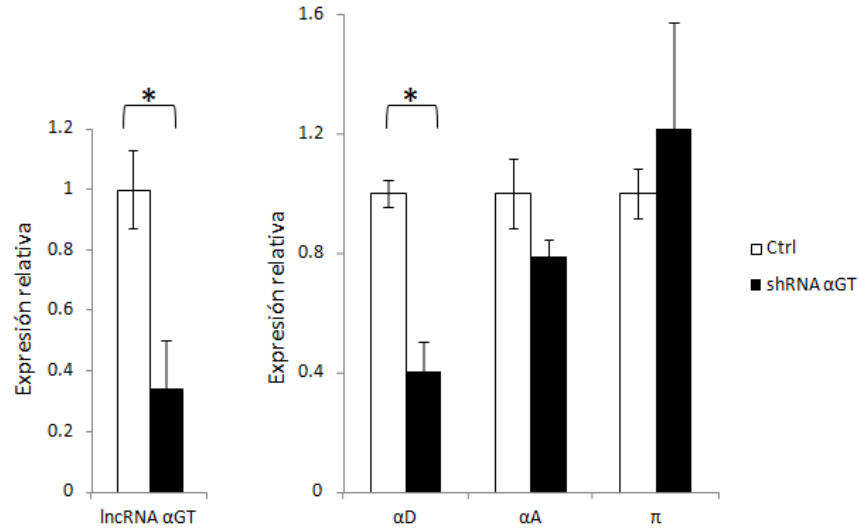


Figura 24.- Knockdown del lncRNA-αGT en células HD3 diferenciadas. Análisis de qRT-PCR para estudiar la expresión de lncRNA-αGT, α^D , α^A , y π (con oligos (8), (6), (7) y (5), respectivamente. Consultar en Materiales y métodos) en el contexto *knockdown* de lncRNA-αGT (shRNA αGT) con respecto al contexto control (Ctrl). En estas gráficas se muestra la expresión promedio y las barras de error corresponden a las desviaciones estándar de tres experimentos independientes. * indica un valor $P < 0.05$

6.2.2. Ensayo de ganancia de función del lncRNA-αGT en células HD3

Los resultados previos del experimento de *knockdown* de lncRNA-αGT pusieron en evidencia que este transcrito es necesario para la completa activación transcripcional del gen α^D durante la diferenciación eritroide adulta. Con el fin de realizar un experimento análogo al experimento de pérdida de función, que al mismo tiempo nos permita abordar la posible función del lncRNA-αGT desde otra perspectiva, se decidió llevar a cabo un experimento de ganancia de función o sobre-expresión del lncRNA-αGT.

Ya que la expresión del lncRNA-αGT se observa únicamente en la etapa final de la diferenciación eritroide adulta (células HD3 diferenciadas, **Figura 15B**) y está involucrada de manera específica con la activación transcripcional del gen α^D , se analizaron los efectos

de una expresión ectópica del lncRNA- α GT sobre la transcripción del gen α^D en células HD3 no inducidas a diferenciación (un estadio previo, en dónde no hay expresión ni de los genes globina, ni de lncRNA- α GT). En nuestro caso específico el tamaño del lncRNA- α GT (~23kb) implica un problema en la realización de la sobre-expresión de lncRNA- α GT. Debido a que este lncRNA es muy grande, una estrategia de clonación del transcrito completo resulta muy difícil. Por ello, se decidió elegir zonas específicas del transcrito de tamaños experimentalmente manejables, y para la elección de las mismas se tomaron en cuenta los siguientes criterios de exclusión/inclusión, respectivamente:

i) Se eliminaron del análisis aquellas zonas que traslapan con otros genes codificantes del dominio α -globina de pollo. Por lo tanto, de acuerdo a la **Figura 25**, las zonas intergénicas que podrían tomarse en cuenta para la sobre-expresión (cajas en azul) son la 7, 8, 9, 10 y 11.

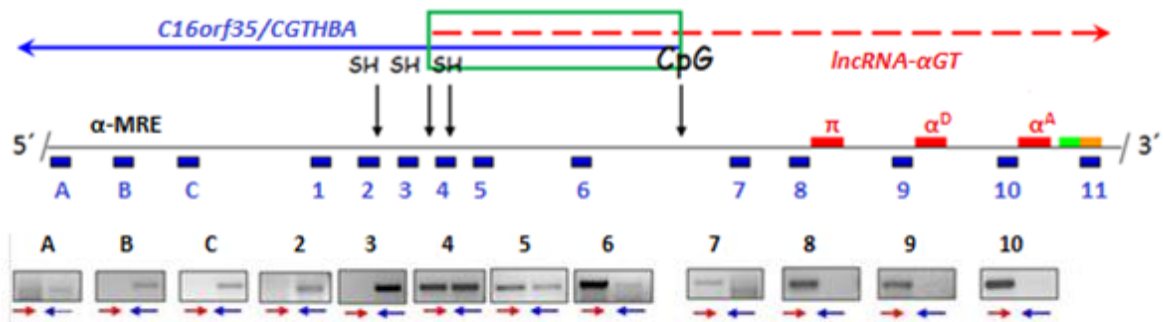


Figura 25.- RT-PCR cadena específica en diferentes zonas a lo largo del dominio α -globina de pollo. Arriba se muestra un esquema del dominio α -globina de pollo. Las cajas azules representan las zonas amplificadas mediante RT-PCR cadena específica.

ii) Las zonas a elegir tendrían que ser en principio importantes en la función del RNA largo. Hasta el momento, no se conocen en detalle los mecanismos mediante los cuales los

lncRNAs llevan a cabo su función; sin embargo, se han descrito múltiples interacciones RNA-proteínas que parecen ser importantes y necesarias para su correcto funcionamiento (Rinn et al. 2007; Khalil et al. 2009; Guttman et al. 2011). En particular, y con base en algunos ejemplos de la literatura, se ha propuesto que algunos lncRNAs presentan estructuras modulares (Guttman y Rinn, 2012). De acuerdo con este modelo, y de manera similar a lo que ocurre convencionalmente con las proteínas, un lncRNA contiene diferentes “dominios” o “módulos funcionales” bien definidos, que actúan de manera independiente pero que en conjunto son necesarios para la función global del lncRNA en cuestión (Wutz et al. 2002; Zappulla y Cech, 2004; Tsai et al. 2010). Estos dominios importantes para la función de un lncRNA suelen estar definidos por la presencia de motivos particulares, mismos que le confieren al lncRNA la capacidad, entre otras cosas, de interactuar con proteínas (Di Ruscio et al. 2013; Ilik et al. 2013), para en conjunto llevar a cabo funciones específicas. Por lo tanto, la identificación de dominios funcionales en un lncRNA es un paso importante para la descripción de sus mecanismos de acción *in vivo*, por lo que se piensa que la sola sobre-expresión de aquellos fragmentos de RNA correspondientes a dominios funcionales dentro de un lncRNA puede ser suficiente para recapitular su función. Análisis bioinformáticos previos del laboratorio (*catRAPID web server*: http://service.tartaglialab.com/page/catrapid_group) predicen que la zona intergénica entre π y α^D (**Figura 25**) presenta un mayor potencial de interacción con proteínas (ej: MLL y CTCF) con respecto a la zona intergénica entre α^D y α^A (**Figura 26**), lo cual posiciona a la zona intergénica π - α^D como una región del lncRNA- α GT que podría estar interactuando con proteínas, y por lo tanto, como una zona importante para el

desempeño de su función. Las zonas intergénicas π - α^D y α^D - α^A miden aproximadamente 2 kb (**Figura 26**).

iii) Adicionalmente, para este experimento de sobre-expresión se eligió clonar un fragmento de DNA de 1 kb localizado en la zona 7 del dominio α -globina de pollo (caja azul, **Figura 25**). A diferencia de nuestras otras opciones de fragmentos a ser sobre-expresados (zonas 8, 9, 10 y 11, **Figura 25**), la zona 7 no corresponde a ningún elemento de regulación propio de los genes α -globina de pollo, es decir, no corresponde a ningún promotor (zonas 8, 9 y 10) o *enhancer* (zona 11); sin embargo, el experimento previo de pérdida de función reveló que la zona 7 pudiera ser una región importante funcionalmente para el lncRNA-GT, ya que, no solo coincide con el exón 2 predicho bioinformáticamente (**Figura 17A**), sino además el shRNAi previamente utilizado en contra de lncRNA- α GT se diseñó en esta misma zona.

Finalmente, se decidió sobre-expresar dos fragmentos:

- 1) Fragmento "A": Región de 1 kb situada en la zona intergénica río arriba del gen π (**Figura 27A**). Esta región se amplificó con los oligonucleótidos (**11**) (consultar en tabla de Materiales y métodos).
- 2) Fragmento "B": Región intergénica entre π - α^D que mide 2 kb (**Figura 27A**). Esta región se amplificó con los oligonucleótidos (**12**) (consultar en tabla de Materiales y métodos).

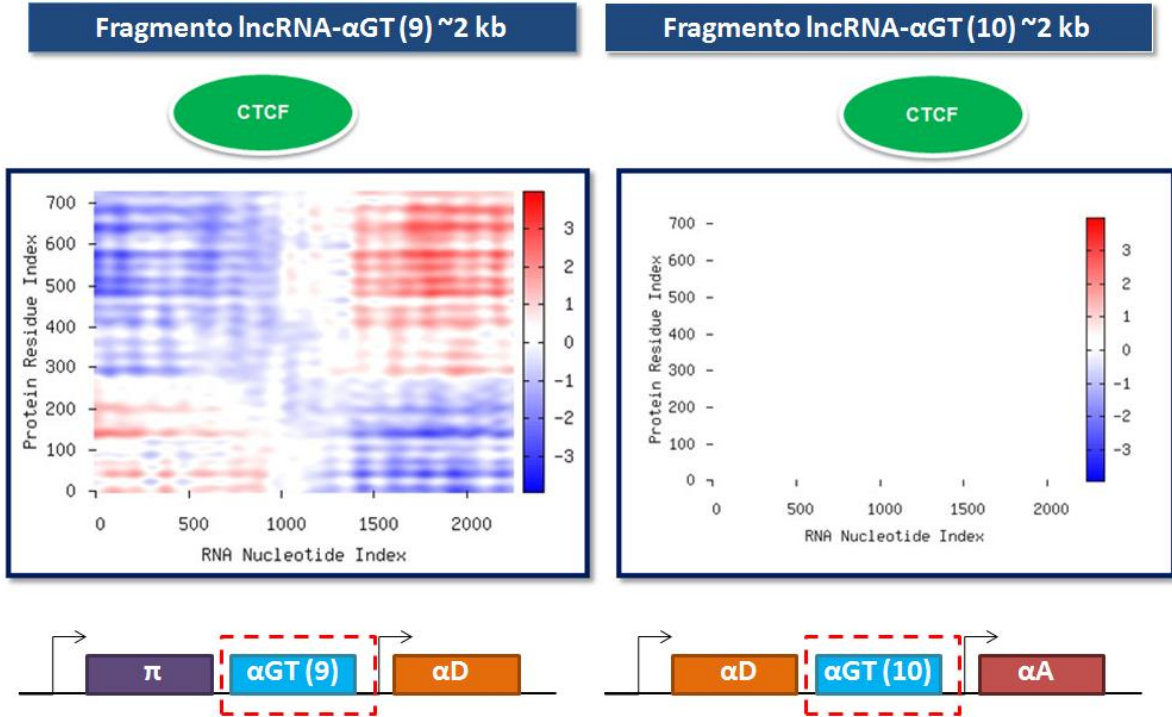
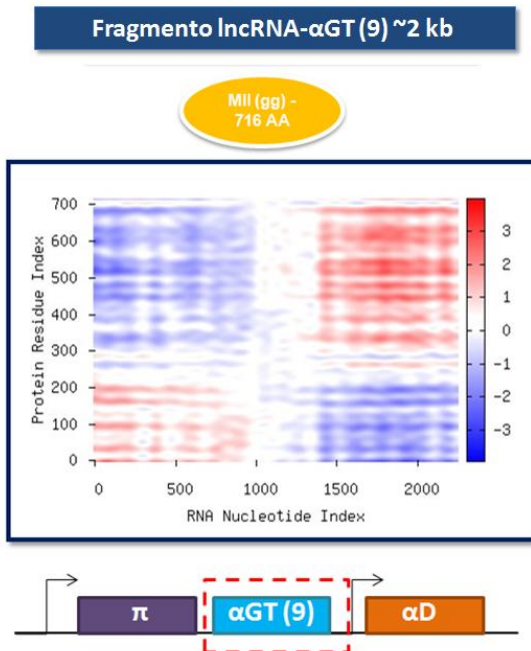
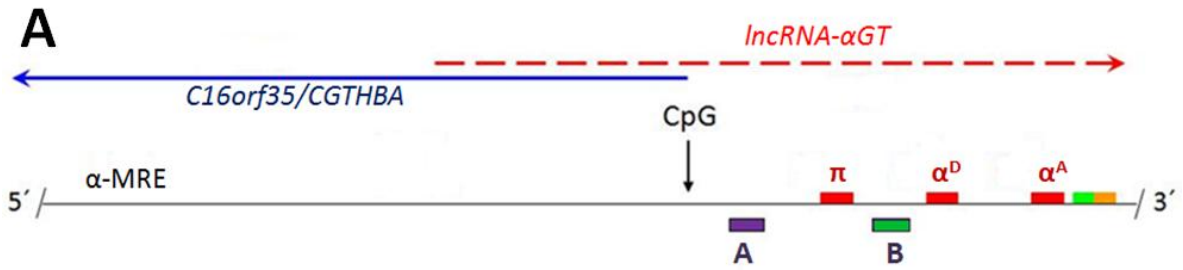
A**B**

Figura 26.- Predicción bioinformática de posibles interacciones RNA-proteína. (A) Interacción entre lncRNA- α GT y CTCF. Izquierda: predicción de interacción entre la proteína CTCF y la región intergénica π - α^D (zona 9, **Figura 25**). Derecha: predicción de interacción entre la proteína CTCF y la región intergénica α^D - α^A

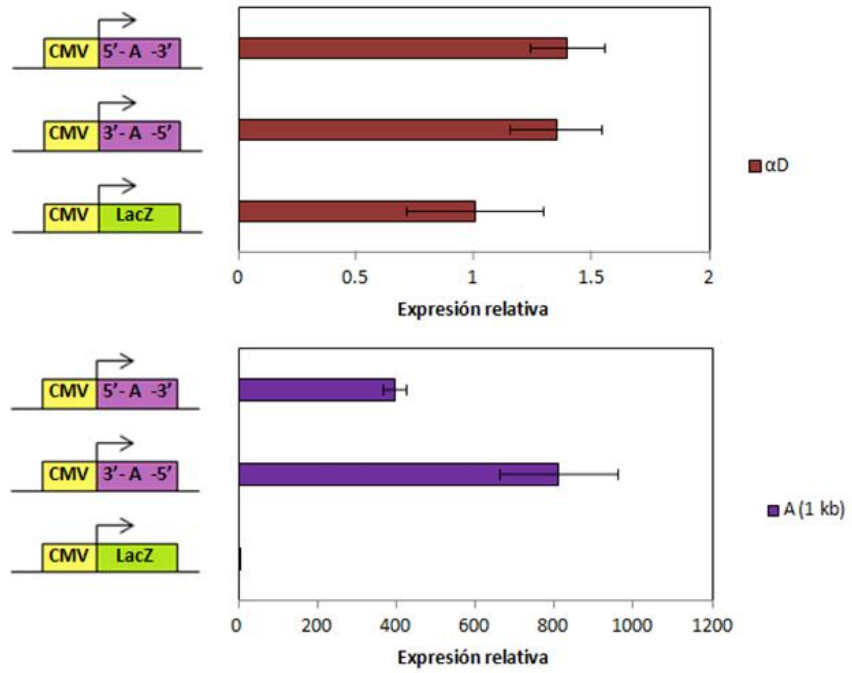
(zona 10, **Figura 25**). **(B)** Interacción entre lncRNA- α GT y la proteína MLL. Debajo de cada gráfica (*heat-map*) se muestra la zona genómica cuyo potencial de interacción fue analizado. La escala de tonos rojos indica una mayor propensión de interacción RNA-proteína en esa zona, mientras que la azul indica una menor propensión. La ausencia de color se interpreta como una incapacidad del programa para predecir una interacción.

Para la sobre-expresión, estas zonas se clonaron bajo un promotor CMV y se transfectaron de manera transitoria en las células HD3 (**Figura 27B, C y D**). Los resultados de este experimento indican que la sobre-expresión de los fragmentos “A” y “B” no parecen influir de manera significativa en la transcripción del gen α^D (La expresión del gen α^D se analizó con los oligos **(6)**).



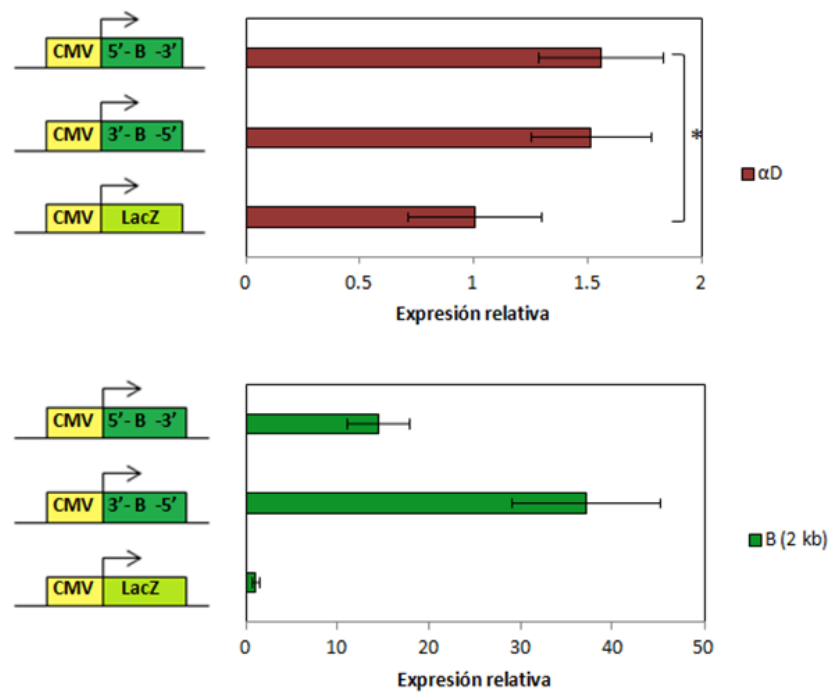
B

Fragmento A
(1 kb, intergénica
CGTHBA-π)



C

Fragmento B
(2 kb, intergénica
π-α^D)



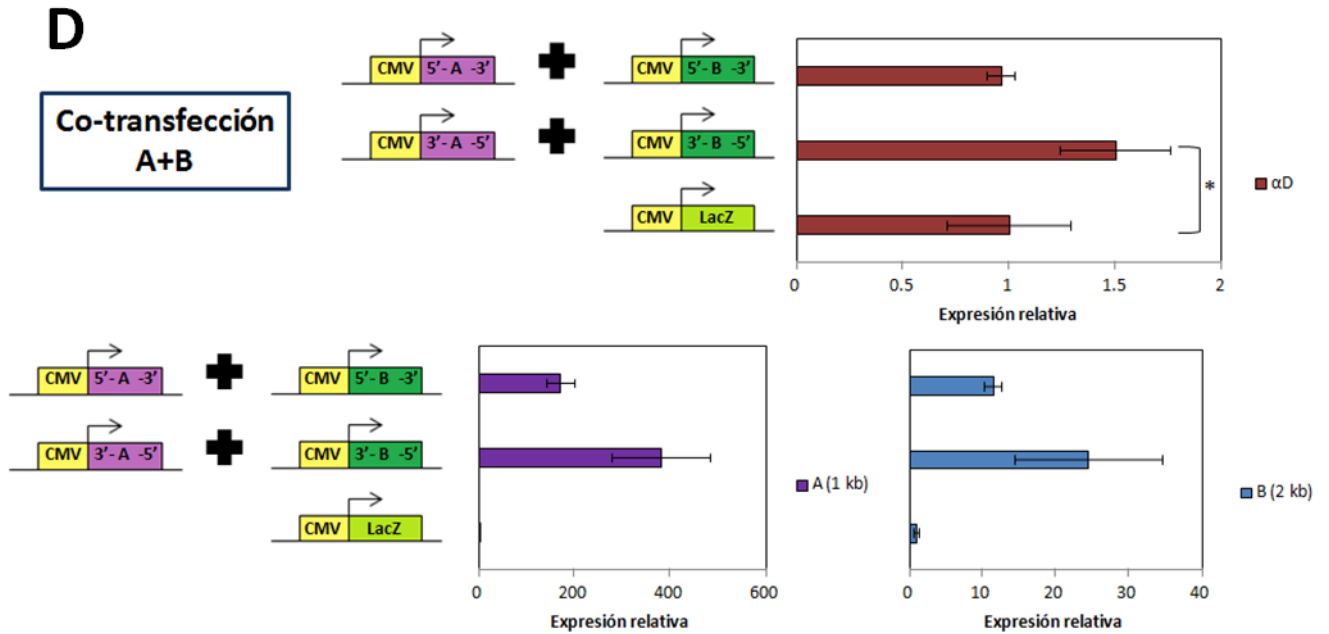


Figura 27.- Sobre-expresión de los fragmentos de DNA intergénicos de 1 y 2 kb en células HD3. (A) Esquema del dominio α -globina de pollo que muestra la localización de los fragmentos de DNA “A” y “B” (cajas en color morado y verde, respectivamente) que se clonaron en el vector de sobre-expresión pcDNA3.1/Hygro(+) bajo el promotor CMV, y posteriormente se transfectaron de manera transitoria en las células HD3. Estos fragmentos de DNA miden 1 y 2 kb, respectivamente. **(B)** Transfección transitoria del fragmento “A”. Como controles negativos se utilizaron las construcciones con el fragmento “A” clonado en dirección opuesta (3’-5’) y *LacZ* (MOCK). **(C)** Transfección del fragmento de 2 kb. **(D)** Co-transfección de los fragmentos de DNA de “A” y “B”. La expresión se analizó mediante qRT-PCR, con una normalización tipo $\Delta\Delta C_t$ tomando la condición MOCK (sobre-expresión de *LacZ*) como referencia. En estas gráficas se muestra la expresión promedio y las barras de error corresponden a las desviaciones estándar de dos diferentes experimentos. * indica un valor $P < 0.05$.

7. DISCUSIÓN

El “recambio de hemoglobina” durante el desarrollo embrionario tiene su base en la expresión diferencial de los genes globina y constituye un ejemplo muy claro de regulación transcripcional diferencial mediada por mecanismos tanto genéticos como epigenéticos. En particular, el dominio α -globina de pollo ha sido muy utilizado como modelo de estudio de diversos mecanismos que permiten una correcta expresión de los genes en su interior. Esfuerzos a lo largo de muchos años han permitido describir de manera más detallada cómo es que los genes globina logran presentar patrones de expresión altamente específicos en tiempo y espacio, los cuales conllevan a una producción adecuada de hemoglobina dependiente de la etapa del desarrollo. Adicionalmente, es importante resaltar que la expresión de los genes globina se regula también a lo largo de la diferenciación eritroide. En consecuencia, no es sorprendente que la expresión de los genes α -globina esté sujeta a una regulación altamente controlada a distintos niveles. Por ello, la descripción de los diferentes elementos reguladores involucrados en este proceso, los cuales van desde factores transcripcionales específicos, aspectos epigenéticos, estructurales y elementos genómicos tanto locales (promotores) como distales (*enhancers*), es necesaria para entender su función. Por lo tanto, la caracterización previa de un RNA largo no-codificante, lncRNA- α GT, que se transcribe a partir del dominio α -globina de pollo pone en evidencia la existencia de un elemento genético adicional y abre la puerta a nuevas preguntas debido a la relevancia creciente de los lncRNAs en la regulación génica (Arriaga-Canon et al. 2014, **ANEXO 2**).

Inicialmente y basados en observaciones de RT-PCR y Northern blots, se observó

que lncRNA- α GT se expresa, en el contexto del desarrollo de pollo, en etapas tempranas (5 días) y tardías (10 días) del desarrollo embrionario. Por otro lado, en el contexto de la diferenciación eritroide adulta, se había demostrado que el lncRNA- α GT se expresa únicamente en las etapas terminales. Cabe resaltar que esta observación se hizo en las células HD3 diferenciadas, cuya inducción promueve únicamente la expresión de las globinas adultas α^D y α^A , por lo que se dice que las HD3 diferenciadas son un símil de los eritrocitos de 10 días. Esta observación permitió describir una correlación entre la expresión del lncRNA- α GT y las globinas adultas α^D y α^A en el contexto de la diferenciación eritroide adulta. Lamentablemente, la falta de un sistema *in vitro* que pudiera mimetizar la diferenciación eritroide propia de las etapas embrionarias no nos permitió hacer estas observaciones en el contexto de la eritropoyesis primitiva. A pesar de ello, en el laboratorio nos encontramos desarrollando una estrategia para aislar y mantener en cultivo células troncales embrionarias de pollo. Uno de nuestros objetivos es que una vez obtenidas estas células podamos diferenciarlas al linaje hematopoyético, y posteriormente al linaje eritroide, y de esta manera recapitular el proceso embrionario de eritropoyesis primitiva, seguido de análisis detallados que nos permitan estudiar, por ejemplo, si lncRNA- α GT es importante en la regulación del gen embrionario π durante la diferenciación eritroide primitiva.

Por lo tanto, se decidió analizar la expresión del lncRNA- α GT de manera cuantitativa, lo que nos permitió concluir que lncRNA- α GT presenta niveles basales de expresión en las etapas tempranas del desarrollo, y posteriormente incrementa considerablemente en el día 10 del desarrollo (**Figura 19B**), etapa en la cual la

eritropoyesis adulta se ha establecido definitivamente, lo cual indica que las globinas adultas α^D y α^A han alcanzado sus niveles máximos de expresión, mismos que se mantendrán durante el resto de la vida del pollo (Groudine y Weintraub, 1981).

A su vez, los análisis de CHIP corroboraron que:

i) El inicio de transcripción del lncRNA- α GT previamente caracterizado mediante RT-PCR cadena específica presenta características propias de un promotor fidedigno regulado diferencialmente a lo largo del desarrollo del pollo. Por una parte, se determinó que el enriquecimiento en la marca de histona H3K4me3 asociada al promotor del lncRNA- α GT ($\rho\alpha$ GT) es mayor en comparación con los controles utilizados P8 (zona del dominio α -globina no promotora) y HI (región de heterocromatina del dominio β -globina) (**Figura 20B**). A su vez, el enriquecimiento en la marca de histona H3K4me3 asociado a $\rho\alpha$ GT muestra un comportamiento dinámico a través del desarrollo (**Figura 21A**), lo que sugiere que $\rho\alpha$ GT es un promotor regulado de manera específica, tanto espacial (linaje eritroide; Arriaga-Canon et al. 2014), como temporalmente (**Figura 21A**).

ii) La marca de histonas H3K4me3 se encuentra mayormente enriquecida en el promotor del lncRNA- α GT en la etapa de 10 días. Ya que la H3K4me3 es una marca de histonas característica de promotores activos, por lo tanto resulta congruente observar un enriquecimiento mayor en $\rho\alpha$ GT justo en la etapa en que el lncRNA- α GT se expresa más. Esta correlación positiva entre enriquecimiento de la marca de histona H3K4me3 en el promotor y niveles de expresión es consistente con resultados de otros grupos que demuestran que la actividad promotora, no solo está íntimamente relacionada con las

modificaciones post-traduccionales de histonas, sino además parece depender de la abundancia de las mismas, por lo que podríamos decir que un enriquecimiento mayor de la marca H3K4me3 en un promotor (o de cualquier otra modificación post-traduccionales de histonas asociada a promotores activos, como por ejemplo: la acetilación) se traduce en niveles transcripcionales más altos (Heintzman, et al, 2007).

Sin embargo, no se observa esta misma correlación para el gen adulto α^D . Ya que los niveles de expresión de α^D son mayores en 10 días con respecto a 5 días (**Figura 19A**), podríamos esperar que el enriquecimiento de la marca H3K4me3 en el promotor del gen α^D fuera también mayor en 10 días. Para nuestra sorpresa, encontramos que este promotor presenta también niveles altos de la marca H3K4me3 en 5 días del desarrollo (**Figura 21C**), lo cual es congruente con resultados previos de nuestro grupo y otros que sugieren que la expresión del gen adulto α^D es dependiente de muchos otros factores que citaremos a continuación. En primer lugar, ha sido bien descrito que el dominio α -globina se encuentra en un contexto cromatínico constitutivamente abierto (Razin et al. 2003; Valdes-Quezada et al. 2013), por lo que la expresión diferencial de los genes α -globina a lo largo del desarrollo parece estar controlada por factores específicos para cada estadio. De esta manera, la expresión del gen embrionario π en las etapas tempranas del desarrollo es controlada por factores transcripcionales específicos para π , y a su vez la expresión de los genes adultos se mantiene como mínima debido a la ausencia de factores específicos para su transcripción que solo están presentes en la etapa de la eritropoyesis definitiva (Valdes-Quezada et al. 2013). Además, resultados previos del laboratorio sugieren que un *enhancer* del dominio interactúa con los promotores de los genes π y α^D desde los 5 días

del desarrollo embrionario; sin embargo, a los 10 días del desarrollo la interacción con π se pierde, mientras que la interacción con α^D se mantiene (ver **Figura 13**). Este fenómeno se acompaña del silenciamiento del gen π , mismo que es mediado por la metilación del DNA a nivel del promotor y del cuerpo del gen (Valdes-Quezada et al. 2013). Esto último ha llevado a proponer que el silenciamiento de π es necesario para la activación de α^D y α^A en la etapa adulta, y que durante la eritropoyesis temprana la expresión de π es predominante y prioritaria sobre la de α^D y α^A , razón por la cual la transcripción del gen α^D es mínima en 5 días (**Figura 19A**), a pesar de que el promotor parece estar “abierto” en esa misma etapa (**Figura 21C**).

Asimismo, los resultados de expresión mediante ensayos de qRT-PCR sugieren que el lncRNA- α GT es 130,000 veces menos abundante que el mRNA del gen α^D en eritrocitos de 10 días, resultado congruente con lo reportado anteriormente, en donde se afirma que los RNA largos no-codificantes se expresan a muy bajas concentraciones con respecto a los genes codificantes (Ravasi et al. 2006; Cabili et al. 2011, Guttman y Rinn, 2012). Además, cabe resaltar que los niveles de expresión del gen α^D y del lncRNA- α GT se correlacionan nuevamente con los respectivos enriquecimientos de la marca en la histona H3K4me3 en sus promotores (**Figuras 19C y 20B**), ya que el enriquecimiento de la marca H3K4me3 en el α GT, si bien es considerable y suficiente para considerarlo como un promotor activo, es menor que el observado en el α^D al día 10 del desarrollo.

Por otro lado, sabemos que no todos los transcritos de naturaleza no-codificante son necesariamente moléculas de RNA funcionales. Existen en la literatura muchos ejemplos de transcripción intergénica en donde se describió que el proceso de

transcripción puede ser en sí mismo el responsable de cambios a nivel de la cromatina e influye en la presencia de ciertos factores transcripcionales, los cuales a su vez permiten la activación o represión de genes vecinos (Martens et al. 2004; Schmitt et al. 2005). Sin embargo, la técnica de RNAi, que promueve la degradación de RNA posterior a su transcripción, permite distinguir entre una molécula de RNA funcional y el acto de la transcripción como los responsables de un proceso regulador dado. En nuestro caso, el *knockdown* contra el lncRNA- α GT conlleva a una disminución en la transcripción del gen α^D en células HD3 diferenciadas, por lo que podemos concluir que el transcrito del lncRNA- α GT *per se* es necesario en la activación y mantenimiento de la transcripción del gen α^D en el contexto de la diferenciación eritroide. Por otra parte, el experimento de ganancia de función indica que la sobre-expresión de los fragmentos de DNA de 1 y 2 kb correspondientes al lncRNA- α GT no tiene efecto alguno sobre la transcripción del gen α^D .

A pesar de que el experimento de ganancia de función no fue tan informativo como el de pérdida de función, creemos que la técnica en sí podría no ser la adecuada para nuestro estudio. Por una parte, la elección de los fragmentos de DNA a sobre-expresar pudo no haber sido la indicada, por lo que futuros experimentos de sobre-expresión utilizando diferentes zonas correspondientes al lncRNA- α GT podrían realizarse para corroborar si estas diferentes regiones influyen en la expresión del gen α^D , y por lo tanto son más importantes para la función reguladora del lncRNA- α GT. Por otra parte, sabemos que el lncRNA- α GT parece estar involucrado en la regulación de su gen vecino α^D , lo que nos hace pensar en un mecanismo de acción en *cis*. Un lncRNA que funciona en *cis* coordina la atracción, reclutamiento y ensamblaje de complejos reguladores

importantes de manera específica debido a que su extremo 3' permanece anclado a su locus de origen a través de la RNA Pol II durante el acto de la transcripción, lo que promueve que el lncRNA funcione como una “guía” específica de complejos reguladores hacia su locus de origen o hacia loci vecinos (Lee, 2009). Esta propiedad es única de los lncRNAs que actúan en *cis*, lo que lleva a deducir que para que un lncRNA lleve a cabo su función reguladora, es indispensable que se encuentre en cercanía con los loci que regula, por lo que la expresión ectópica de los fragmentos del lncRNA- α GT (1 y 2 kb) podría explicar por qué no se observó un efecto sobre la transcripción del gen α^D en el experimento de ganancia de función, y por lo tanto, ser congruente con la propuesta de que el lncRNA- α GT funciona en *cis* (Wang et al. 2011). Sin embargo, en la literatura ya se han reportado experimentos de sobre-expresión en donde se observó que las moléculas de RNA, generadas a partir de la expresión ectópica de un lncRNA previamente caracterizado como un regulador en *cis*, tienen la capacidad de encontrar a sus sitios blanco dentro del genoma, y recapitular su función reguladora (Martianov et al. 2007; Schmitz et al. 2010; Jeon y Lee, 2011), lo que deja abierta la posibilidad de que muchos lncRNAs que funcionan en *cis*, pudieran potencialmente funcionar también en *trans* regulando muchos otros blancos dentro del genoma (Rinn y Chang, 2012). Por lo tanto, análisis más detallados serán necesarios en un futuro para determinar si el lncRNA- α GT funciona en *cis* o en *trans*, así como una descripción más detallada sobre su mecanismo de acción. Cabe resaltar que las transfecciones del experimento de ganancia de función fueron transitorias, lo que deja abierta la posibilidad de que en la transfección estable, cuya eficiencia es mayor, pudieran observarse resultados diferentes. La elección del

método transitorio se tomó ya que se buscaba comprobar si el lncRNA- α GT era un elemento necesario y suficiente para la activación de α^D , asumiendo que en caso de funcionar como un activador de α^D durante la diferenciación eritroide, una sobreexpresión transitoria podría ser suficiente para observar un efecto sobre la activación transcripcional del gen α^D . Sin embargo, los resultados aquí reportados sugieren que lncRNA- α GT es un elemento necesario, más no suficiente para la activación de la expresión del gen α^D durante la diferenciación eritroide.

Este mecanismo de regulación de genes globina por lncRNAs podría encontrarse conservado en otras especies, ya que al menos en ratón y humano se ha descrito la presencia de transcritos a lo largo del dominio β -globina (Bastos y Aviv, 1997; Ashe et al. 1997; Gribnau et al. 2000).

8. CONCLUSIÓN

Hasta el momento, a pesar de que la función del lncRNA- α GT no es clara, podemos concluir debido al patrón temporal de expresión observado en el contexto del desarrollo del pollo que el transcrito largo no-codificante juega un papel en la eritropoyesis definitiva, y por el contrario, no parece ser necesario en las etapas más tempranas del desarrollo. Dado que el gen α^D comienza a incrementar sus niveles de expresión desde los 7 días del desarrollo, podemos descartar que el lncRNA- α GT regule etapas tempranas de la activación de los genes α -globina adultos α^D y α^A . Si el lncRNA- α GT tiene un papel en la regulación de los genes α^D y α^A , muy probablemente sea en una etapa posterior y el mecanismo queda por determinarse. Por otro lado, en el contexto de la diferenciación eritroide adulta, se demostró que el lncRNA- α GT es necesario para la activación del gen α^D .

9. PERSPECTIVAS

Futuros estudios de caracterización funcional serán necesarios para determinar el mecanismo de acción a través del cual el lncRNA- α GT regula la expresión del gen globina adulto α^D en el contexto de la diferenciación eritroide adulta.

En primer lugar, quedan por redondearse y repetirse los experimentos de sobre-expresión. Recordemos que la sobre-expresión analizada en este trabajo se realizó transitoriamente, por lo que proponemos que la transfección estable de estas construcciones podría tener efectos diferentes. Además, en lugar de analizarse

únicamente lo que sucede en las células HD3, se puede analizar los efectos asociados a la sobre-expresión en el contexto de HD3 inducidas a diferenciación.

Por otro lado sería interesante determinar si este transcrito interactúa con alguna proteína, ya sea mediante *pulldowns* con sondas de RNA biotiniladas o mediante inmunoprecipitación de RNA. En el laboratorio nos gustaría estudiar si existe una interacción entre lncRNA- α GT y la proteína CTCF, ya que recientemente se demostró que CTCF posee un dominio de unión a RNA (Meyer et al. 2014), y además CTCF es un factor crucial en la regulación del dominio α -globina de pollo.

Un aspecto muy interesante que atañe a la configuración tridimensional del dominio α -globina de pollo es la formación de un *hub* “activo” de cromatina propio de la etapa final de la diferenciación eritroide. La adquisición de esta estructura topológica correlaciona con la expresión de los genes adultos α^D y α^A , así como con la de lncRNA- α GT. Ya que existen antecedentes de lncRNAs involucrados en la estructuración de la cromatina (Zhang et al. 2014), sería interesante demostrar si lncRNA- α GT es necesario en el establecimiento y mantenimiento del *hub*. Dado que la expresión del lncRNA es tardía en el contexto del desarrollo de pollo, el lncRNA- α GT podría estar involucrado en el mantenimiento de la expresión del gen α^D a lo largo de la vida del pollo.

10.REFERENCIAS

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 alfa-globin domain. *Epigenetics* 9:173-181

Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P., Proudfoot, N.J. (1997). Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev.* 11:2494-2509

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136:215-233

Baumann, R., Dragon, S. (2005). Erythropoiesis and red cell function in vertebrate embryos. *Eur. J. Clin. Invest.* 35:2-12

Bastos, R.N., Aviv, H. (1977). Globin RNA precursor molecules: Biosynthesis and processing in erythroid cells. *Cell* 11: 641-650

Bernards, R., Brummelkamp, T.R., Beijersbergen, R.L. (2006). shRNA libraries and their use in cancer genetics. *Nat. Methods* 3:701-706

Bernstein, E., Allis, C.D. (2005). RNA meets chromatin. *Genes Dev.* 19:1635-1655

Bernstein, B.E. et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120:169-181

Bertani, S., Sauer, S., Bolotin, E., Sauer, F. (2011). The noncoding RNA *Mistral* activates Hoxa6 and Hoxa7 expression and stem cell differentiation by recruiting MLL1 to chromatin. *Mol. Cell* 43:1040-1046

Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, et al. (2004). Global identification of human transcribed sequences with genome tiling arrays. *Science* 306:2242-2246

Beug, H., Doederlein, G., Freudenstein, C., Graf, T. (1979) 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.* 1:195-207

Beug, H., A. von Kirchbach, Doderlein, G., Conscience, J.F., Graf, T. (1979). Chicken hematopoietic cells transformed by seven strains of defective avian leukemia viruses display three distinct phenotypes of differentiation. *Cell* 18:375-390

Beug, H., Doderlein, G., Freudenstein, C., & Graf, T. (1982). Erythroblast cell lines transformed by a temperature sensitive mutant of avian erythroblastosis virus: A model system to study erythroid differentiation in vitro. *Journal of Cellular Physiology Supplement*, 1, 195-207

Brown, R.S. (2005). Zinc finger proteins: getting a grip on RNA. *Curr Opin Struct Biol* 15: 94-98

Brown, J.M., Leach, J., Reittie, J.E., Atzberger, A., Lee-Prudhoe, J., Wood, W.G., Higgs, D.R., Iborra, F.J., Buckle, V.J. (2006). Coregulated human globin genes are frequently in spatial proximity when active. *J. Cell. Biol.* 172:177-187

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25:1915-1927

Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, et al. (2005). The transcriptional landscape of the mammalian genome. *Science* 309:1559-63

Cesana M, Cacchiarelli D, Legnini I, Santini T, Sthandier O, Chinappi M, Tramontano A, Bozzoni I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* 147:358-69

Chalk, A.M., Wahlestedt, C., Sonnhammer, E.L. (2004). Improved and automated prediction of effective siRNA. *Biochem. Biophys. Res. Commun.* 319:264-274

Chang, K., Elledge, S.J., Hannon, G.J. (2006). Lessons from nature: microRNA-based shRNA libraries. *Nat. Methods* 3:707-714

Choy, M.K., Movassagh, M., Goh, H.G., Bennett, M.R., Down, T.A., Foo, R. (2010). Genome-wide conserved consensus transcription factor binding motifs are hyper-methylated *BMC Genomics*.11: 519

Chu, C., Qu, K., Zhong, FL., Artandi, S.E., Chang, H.Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol Cell*. 44:667-678

Clark, M.B., Mattick, J.S. (2011). Long noncoding RNAs in cell biology. *Semin. Cell Dev. Biol.* 22:366-76

Cremer, T., Cremer, M. (2010). Chromosome territories. *Cold Spring Harb Perspect Biol.* 2:a003889

Crick, F. (1970). Central dogma of molecular biology. *Nature*. 227:561-563

Derrien, T., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res.* 22:1775-1789

Di Ruscio, A., Ebralidze, A.K., Benoukraf, T., Amabile, G., Goff, L.A., Terragni, J., Figueroa, M.E., De Figueiredo Pontes, L.L., Alberich-Jorda, M., Zhang, P., Wu, M., D'Alò, F., Melnick, A., Leone, G., Ebralidze, K.K., Pradhan, S., Rinn, J.L., Tenen, D.G. (2013). DNMT1-interacting RNAs block gene-specific DNA methylation. *Nature* 503:371-380

Dinger, M.E., Amaral, P.P., Mercer, T.R., Pang, K.C., Bruce, S.J., et al. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* 18:1433-45

Dinger, M.E., Amaral, P.P., Mercer, T.R., Mattick, J.S. (2009). Pervasive transcription of the eukaryotic genome: Functional indices and conceptual implications. *Brief Funct Genomics Proteomics* 8: 407-423.

Djebali, S., Davis, C.A. et al. (2012). Landscape of transcription in human cells. *Nature* 489:101-108

Ebisuya, M., Yamamoto, T., Nakajima, M., Nishida, E. (2008). Ripples from neighbouring transcription. *Nat. Cell Biol.* 10:1106-1113

Escamilla-Del-Arenal, M., Recillas-Targa, F. (2008). GATA-1 modulates the chromatin structure and activity of the chicken α -globin 3' enhancer. *Mol Cell Biol* 28:575-586

Felsenfeld, G. (1993). Chromatin structure and the expression of globin-encoding genes. *Gene* 135:119-124

Felsenfeld, G., Boyes, G., Chung, J., Clark, D., Studitsky, V. (1996). Chromatin structure and gene expression. *Proc. Natl Acad. Sci. USA* 93:9384-9388

Felsenfeld, G., Groudine, M. (2003). Controlling the double helix. *Nature* 421:448-453

Ferreira, R., Ohneda, K., Yamamoto, M., Philipsen, S. (2005). GATA1 function, a paradigm for transcription factors in hematopoiesis. *Mol. Cell. Biol.* 25:1215-1227

Garber, M., Grabherr, MG., Guttman, M., Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* 8:469-477

Gavrilov, A.A., Razin, S.V. (2008). Spatial configuration of the chicken α -globin gene domain: immature and active chromatin hubs. *Nucleic Acids Res.* 36:4629-4640

Geyer, P.K., Vitalini, M.W., Wallrath, L.L. (2011). Nuclear organization: taking a position on gene expression. *Curr Opin Cell Biol.* 23:354-359

- Graf, T. (2002). Differentiation plasticity of hematopoietic cells. *Blood* 99:3089-3101
- Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R., Fraser, P. (2000). Intergenic transcription and developmental remodeling of chromatin subdomains in the human β -globin locus. *Mol. Cell*, 5:377-386
- Groudine, M., Weintraub, H. (1981). Activation of globin genes during chicken development. *Cell* 24:393-412
- Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464:1071-1076
- Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458:223-227
- Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., et al. (2010). Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* 28:503-510
- Guttman, M. et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477:295-300
- Guttman, M., Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature* 482:339-346
- Hacisuleyman, E., Goff, L.A., Trapnell, C., Williams, A., Henao-Mejía, J., Sun, L. et al. (2014). Topological organization of multichromosomal regions by the long noncoding RNA Firre. *Nat. Struct. Mol. Biol.* 21:198-206
- Hargreaves, D.C., Crabtree, G.R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.* 21:396-420

Hattangadi, S.M., Wong, P., Zhang, L., Flygare, J., Lodish, H.F. (2011). From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. *Blood*. 118:6258-6268

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., Wang, W., Weng, Z., Green, R.D., Crawford, G.E., Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39:311-318

Higgs, D.R., William, G.W. (2008). Long-range regulation of a globin gene expression during erythropoiesis. *Curr. Opin. Hematol.* 15:176-183

Holwerda, S.J.B., de Laat, W. (2013). CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philos Trans R Soc Lond B Biol Sci* 368:20120369

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., et al. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142:409-419

Hung, T., Wang, Y., Lin, M.F., Koegel, A.K., Kotake, Y., et al. (2011). Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat. Genet.* 43:621-29

Hutvagner, G. (2005). Small RNA asymmetry in RNAi: function in RISC assembly and gene regulation. *FEBS Lett.* 579:5850-5857

Ichimura, T., Watanabe, S., Sakamoto, Y., Aoto, T., Fujita, N., et al. (2005). Transcriptional repression and heterochromatin formation by MBD1 and MCAF/AM family proteins. *J. Biol. Chem.* 280, 13928-13935

Ilik, I.A., Quinn, J.J., Georgiev, P., Tavares-Cadete, F., Maticzka, D., Toscano, S., Wan, Y., Spitale, R.C., Luscombe, N., Backofen, R., Chang, H.Y., Akhtar, A. (2013). Tandem Stem-Loops in roX RNAs Act Together to Mediate X Chromosome Dosage Compensation in *Drosophila*. *Mol. Cell* 51:156-173

Ingram, V.M. (1972). Embryonic red cell formation. *Nature* 235, 338-339

Ioudinkova, E., Razin, S.V., Borunova, V., de Conto, F., Rynditch, A., Scherrer, K. (2005) RNA-dependent nuclear matrix contains a 33 kb globin full domain transcript as well as prosomes but no 26S proteasomes. *J Cell Biochem.* 94:529-539

Jagla, B. et al. (2005). Sequence characteristics of functional siRNAs. *RNA* 11:864-872

Jeon, Y., Lee, J.T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* 146:119-133

Jiang, C., Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. *Nat. Rev. Genet.* 10:161-172

Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., et al. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 296:916-919

Khalil, A. M. et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl Acad. Sci. USA* 106:11667-11672

Kertesz, M. et al. (2007). The role of site accessibility in microRNA target recognition, *Nat. Genet.* 39:1278-1284

Kertesz, M., Wan, Y., Mazor, E., Rinn, J.L., Nutter, R.C., et al. (2010). Genome-wide measurement of RNA secondary structure in yeast. *Nature* 467:103-107

Khvorova, A., Reynolds, A., Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell* 115:209-216

Koerner, M.V., Pauler, F.M., Huang, R., Barlow, D.P. (2009). The function of non-coding RNAs in genomic imprinting. *Development* 136:1771-1783

Kosak, S.T., Groudine, M. (2004). Form follows function: the genomic organization of cellular differentiation. *Genes Dev.* 18:1371-1384

Koziol, MJ., Rinn, J.L. (2010). RNA traffic control of chromatin complexes. *Curr. Opin. Genet. Dev.* 20:142-148

Kutter, C., Watt, S., Stefflova, K., Wilson, M.D., Goncalves, A., Ponting, C.P., Odom, D.T., Marques, A.C. (2012). Rapid turnover of long noncoding RNAs and the evolution of gene expression. *PLoS Genet.* 8:e1002841

de Laat, W., Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* 502:499-506

Lai, F., Shiekhattar, R. (2014). Enhancer RNAs: the new molecules of transcription. *Curr. Opin. Genet. Dev.* 25:38-42

Lawson, H.A., Cheverud, J.M., Wolf, J.B. (2013). Genomic imprinting and parent-of-origin effects on complex traits. *Nat Rev Genet.* 14:609-617

Lee, J.T. (2009). Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes Dev.* 23:1831-1842

Lee JT. (2010). The X as model for RNA's niche in epigenomic regulation *Cold Spring Harb Perspect Biol.* 2:a003749

Li, B., Carey, M., Workman, J.L. (2007). The role of chromatin during transcription. *Cell* 128:707-719

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K., Song, X., Oh, S., Kim, H.S., Glass, C.K., Rosenfeld, M.G. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 498:516-520

Lin, N., Chang, K.Y., Li, Z., Gates, K., Rana, Z.A., Dang, J., Zhang, D., Han, T., Yang, C.S., Cunningham, T.J., Head, S.R., Duester, G., Dong, P.D., Rana, T.M. (2014). An Evolutionarily Conserved Long Noncoding RNA TUNA Controls Pluripotency and Neural Lineage Commitment. *Mol. Cell* 53:1005-1019

Loewer, S., Cabili, M.N., Guttman, M., Loh, Y.H., Thomas, K., Park, I.H., Garber, M., Curran,

M., Onder, T., Agarwall, S., Manos, P.D., Datta, S., Lander, E.S., Schlaeger, T.M., Daley, G.Q., Rinn, J.L. (2010). Large non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat. Genet.* 42:1113-1117

Lu, D., Searles, M.A., Klug, A. (2003). Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature* 426: 96-100

Luo, K.Q., Chang, D.C. (2004). The gene-silencing efficiency of siRNA is strongly dependent on the local structure of mRNA at the targeted region. *Biochem. Biophys. Res. Commun.* 318:303-310

Lyle R, Watanabe D, te Vruchte D, Lerchner W, Smrzka OW, Wutz A, Schageman J, Hahner L, Davies C, Barlow DP. (2000). The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. *Nat Genet* 25: 19-21

Maison, C., Bailly, D., Peters, A.H., Quivy, J.P., Roche, D., Taddei, A., Lachner, M., Jenuwein, T., Almouzni, G. (2002). Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat. Genet.* 30:329-334

Martens, J.A., Laprade, L., Winston, F. (2004). Intergenic transcription is required to repress the *Saccharomyces cerevisiae* *SER3* gene. *Nature* 429:571-574

Martianov, I., Ramadass, A., Serra Barros, A., Chow, N., Akoulitchev, A. (2007). Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* 445:666-670

Mattick, J.S. (2007). A new paradigm for developmental biology. *J. Exp. Biol.* 210:1526-1547

Meister, G., Tuschl, T. (2004). Mechanisms of gene silencing by double-stranded RNA. *Nature* 431:343-349

Mercer, T.R., Dinger, M.E., Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* 10:155-159

Mercer, T.R., Mattick, J.S. (2013). Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct.Mol. Biol.* 20:300-307

Messerschmidt, D.M., Knowles, B.B., Solter, D. (2014). DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev.* 28:812-828

Mikkelsen, T. S. et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553-560

Misteli, T. (2005). Concepts in nuclear architecture. *Bioessays* 27:477-487

Moazed, D. (2009). Small RNAs in transcriptional gene silencing and genome defence. *Nature* 457, 413-420

Müller, J., Kassis, J.A. (2006). Polycomb response elements and targeting of Polycomb group proteins in *Drosophila*. *Curr Opin Genet Dev* 16:476-484

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., et al. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322:1717-1720

Nicolas, R.H., Partington, G., Major, G.N., Smith, B., Carne, A.F., Huskisson, N., Goodwin, G. (1991). Induction of differentiation of avian erythroblastosis virus-transformed erythroblasts by the protein kinase inhibitor H7: analysis of the transcription factor EF1. *Cell Growth Differ* 2:129-135

Ogawa, Y., Sun, B.K., Lee, J.T. (2008). Intersection of the RNA interference and X-inactivation pathways. *Science*. 320:1336-1341

Osborne, C.S., Chakalova, L., Brown, K.E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J.A., Lopes, S., Reik, W., Fraser, P. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet.* 36:1065-1071

Palis, J., Segel, G.B. (1998). Developmental biology of erythropoiesis. *Blood Rev.* 12:106-

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., Kanduri, C. (2008). *Kcnq1ot1* antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol. Cell* 32, 232-246

Pang, K.C., Frith, M.C., Mattick, J.S. (2006) Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet* 22:1-5

Paul J, Duerksen JD. (1975). Chromatin-associated RNA content of heterochromatin and euchromatin. *Mol Cell Biochem* 9:9-16

Pei, Y., Tuschl, T. (2006). On the art of identifying effective and specific siRNAs. *Nature Methods* 3:670-676

Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S., Brockdorff, N. (1996). Requirement for Xist in X chromosome inactivation. *Nature* 379:131-137

Phillips, J.E., Corces, V.G. (2009). CTCF: master weaver of the genome. *Cell* 137: 1194-1211

Pokholok, D.K. et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122:517-527

Ponjavic, J., Ponting, C.P., Lunter, G. (2007). Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.* 17:556-565

Ponjavic, J., Oliver, P. L., Lunter, G., Ponting, C. P. (2009). Genomic and transcriptional co-localization of protein-coding and long non-coding RNA pairs in the developing brain. *PLoS Genet.* 5:e1000617

Pontier, D.B., Gribnau, J. (2011). Xist regulation and function explored. *Hum. Genet.* 130:223-236

Ponting, C.P., Oliver, P.L., Reik, W. (2009). Evolution and functions of long noncoding RNAs. *Cell* 136:629-641

Prensner, J.R., Iyer, M.K., Sahu, A., Asangani, I.A., Cao, Q., Patel, L., Vergara, I.A., et al. (2013). The long noncoding RNA *SChLAP1* promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. *Nat Genet.* 45:1392-1398

Ramírez, J., Dege, C., Kutateladze, T. G., Hagman, J. (2012) MBD2 and multiple domains of CHD4 are required for transcriptional repression by Mi-2/NuRD complexes. *Mol. Cell. Biol.* 32:5078-5088

Ravasi, T. et al. (2006). Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res.* 16:11-19

Razin, S.V., Farrel, C.M., Recillas-Targa, F. (2003). Genomic domains and regulatory elements operating at the domain level. *Int Rev Cytol* 226:63-125

Razin, S.V., Rynditch, Alla., Borunova, V., loudinkova, E., Smalko, V., Scherrer, K. (2004). The 33 kb transcript of the chicken α -globin gene domain is part of the nuclear matrix. *J. Cell. Biochem.* 92:445-457

Reese, K. J., Lin, S., Verona, R. I., Schultz, R. M., Bartolomei, M. S.(2007) Maintenance of paternal methylation and repression of the imprinted H19 gene requires MBD3. *PLoS Genet.* 3, e137

Reik, W., Dean, W., Walter, J. (2001). Epigenetic reprogramming in mammalian development. *Science* 293:1089-1093

Reynolds, A. et al. (2004). Rational siRNA design for RNA interference. *Nat. Biotechnol.* 22:326-330

Rincón-Arano, H., Valadez-Graham, V., Guerrero, G., Escamilla-Del-Arenal, M., Recillas-Targa, F. (2005). YY1 and GATA-1 Interaction modulate the chicken 3'-side α -globin enhancer activity. *J. Mol. Biol.* 349:961-975

Ringrose, L., Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.* 38:413-443

Rinn, J.L., Euskirchen, G., Bertone, P., Martone, R., Luscombe, N.M., et al. (2003). The transcriptional activity of human Chromosome 22. *Genes Dev.* 17:529-540

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., Chang, H.Y. (2007). Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* 129:1311-1323

Rinn, J.L., Chang, H.Y. (2012). Genome Regulation by Long Noncoding RNAs. *Annu. Rev. Biochem.* 81:145-166

Saldaña-Meyer, R., González-Buendía, E., Guerrero, G., Narendra, V., Bonasio, R., Recillas-Targa, F., Reinberg, D. (2014). CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.* 28:723-734

Schmitt, S., Prestel, M., Paro, R. (2005). Intergenic transcription through a Polycomb group response element counteracts silencing. *Genes Dev.* 19:697-708

Schmitz, K.M., Mayer, C., Postepska, A., Grummt, I. (2010). Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. *Genes Dev.* 24:2264-2269

Schneider, R. et al. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. *Nat. Cell. Biol.* 6:73-77

Schubert, S., Grunweller, A., Erdmann, V.A., Kurreck, J. (2005). Local RNA target structure influences siRNA efficacy: systematic analysis of intentionally designed binding regions. *J. Mol. Biol.* 348:883-893

Schwarz, D.S. et al. (2003). Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115:199-208

Searles, M.A., Lu, D., Klug, A. (2000). The role of the central zinc fingers of transcription factor IIIA in binding to 5 S RNA. *J Mol Biol* 301:47-60

Seidl, C.I., Stricker, S.H., Barlow, D.P. (2006). The imprinted Air ncRNA is an atypical RNAPII transcript that evades splicing and escapes nuclear export. *EMBO J.* 25:3565-3575

Shabalina, S.A., Spiridonov, A.N., Ogurtsov, A.Y. (2006). Computational models with thermodynamic and composition features improve siRNA design. *BMC Bioinformatics* 7:65

Simon, M. D. et al. (2011). The genomic binding-sites of a non-coding RNA. *Proc. Natl. Acad. Sci. USA* 108:20497-20502

Siomi, M.C., Sato, K., Pezic, D., Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.* 12:246-258

Sleutels, F., Zwart, R., Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 415:810-813

Song, S.H., Hou, C., Dean, A. (2007). A positive role for NLI/Ldb1 in long-range beta-globin locus control region function. *Mol Cell.* 28:810-822

Sontheimer, E.J. (2005). Assembly and function of RNA silencing complexes. *Nat.Rev. Mol. Cell Biol.* 6:127-138

Spitale, R.C., Tsai, M.C., Chang, H.Y. (2011). RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. *Epigenetics* 6:539-543

Stamatoyannopoulos, G. (2005). Control of globin gene expression during development and erythroid differentiation. *Exp. Hematol.* 33:259-271

Sun, B.K., Deaton, A.M., Lee, J.T. (2006). A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol Cell.* 21:617-628

Takasaki, S., Kotani, S., Konagaya, A. (2004). An effective method for selecting siRNA target sequences in mammalian cells. *Cell Cycle* 3:790-795

Tripathi, V., Ellis, J.D., Shen, Z., Song, D.Y., Pan, Q., et al. (2010). The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Mol. Cell* 39:925-938

Tsai, M.C., Manor, O., Wan, Y., Mosammamarast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329:689-693

Ui-Tei, K. et al. (2004). Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res.* 32:936-948

Underwood, J.G., Uzilov, A.V., Katzman, S., Onodera, C.S., Mainzer, J.E., et al. (2010). FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods* 7:995-1001

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* 8:1-12

Vignali, M., Hassan, A.H., Neely, K.E., Workman, J.L.. (2000). ATP-dependent chromatin remodeling complexes. *Mol Cell Biol.* 20:1899-1910

Wan, Y., Kertesz, M., Spitale, R.C., Segal, E., Chang, H.Y. (2011). Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.* 12:641-655

Wang, J., Zhang, J., Zheng, H., Li, J., Liu, D., Li, H., Samudrala, R., Yu, J., Wong, G.K. (2004). Mouse transcriptome: neutral evolution of 'non-coding' complementary DNAs. *Nature* 431:1

Wang, K.C., Chang, H.Y. (2011). Molecular Mechanisms of Long Noncoding RNAs. *Mol. Cell* 43:904-914

Wang, K.C. et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 472:120-124

Wang, Y., Xu, Z., Jiang, J., Xu, C., Kang, J., Xiao, L., Wu, M., Xiong, J., Guo, X., Liu, H. (2013). Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Dev. Cell.* 25:69-80

Willingham, A.T. et al. (2005). A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science* 309:1570-1573

Willingham, A.T., Gingeras, T.R. (2006). TUF love for “junk” DNA. *Cell.* 125:1215-1220

Wilusz, J.E., Sunwoo, H., Spector, D.L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* 23:1494-1504

Wutz, A., Jaenisch, R.A. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol. Cell* 5:695-705

Wutz, A., Rasmussen, T.P., Jaenisch, R.A. (2002). Chromosomal silencing and localization are mediated by different domains of *Xist* RNA. *Nat. Genet.* 30:167-174

Wutz, A. (2013). Noncoding roX RNA remodeling triggers fly dosage compensation complex assembly. *Mol Cell.* 51:131-132

Yao, H., Brick, K., Evrard, Y., Xiao, T., Camerini-Otero, R.D., Felsenfeld, G. (2010). Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* 24:2543-2555

Zappulla, D.C., Cech, T. R. (2004). Yeast telomerase RNA: a flexible scaffold for protein subunits. *Proc. Natl Acad. Sci. USA* 101:10024-10029

Zhang, H., Zeitz, M.J., Wang, H., Niu, B., Ge, S., Li, W., Cui, J., Wang, G., Qian, G., Higgins, M.J., Fan, X., Hoffman, A.R., Hu, J. (2014). Long noncoding RNA-mediated

intrachromosomal interactions promote imprinting at the *Kcnq1* locus. *J Cell Biol.* 204:61-75

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322:750-756

Zhu, Y., Rowley, M.J., Böhmendorfer, G., Wierzbicki, A.T. (2013). A SWI/SNF Chromatin-Remodeling Complex Acts in Noncoding RNA-Mediated Transcriptional Silencing. *Mol Cell.* 49:298-309

11.ANEXOS

CTCF demarcates chicken embryonic α -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 α -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 π gene and the transcriptional activation of the adult α^D and α^A genes at the genomic domain level. In early stages, we describe a physical contact between the embryonic π 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 α^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 α^D gene. In avian-transformed erythroblast HD3 cells that are induced to differentiate, we found active DNA demethylation of the adult α^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 π 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 α -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.^{1,2} To address these aspects of gene regulation in more detail, we adopted the chicken α -globin gene domain as a paradigm to understand the influence of chromatin domain structure on gene expression (Fig. 1A).³ For years we have been interested in understanding genomic domain-level mechanisms of differential α -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 α - and β -globin gene expression (Fig. 1A). This process depends on a complex interplay among gene promoters, distal regulatory elements, nuclear dynamics, and the epigenetic milieu.^{4–6}

Unlike the chicken β -globin gene domain, the α -globin domain lies in a constitutive open chromatin context on

minichromosome 14 and contains the embryonic π gene and adult α^D and α^A genes (Fig. 1B). The α -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 α -major regulatory element (α -MRE), followed by the α EHS-1.4 element, which is a CTCF-dependent insulator.⁷ This insulator is located within an intron of the *C16orf35* gene antisense transcript,^{8,9} between the α -MRE element and the α -globin genes (Fig. 1B). Further downstream, 4 kb upstream of the π gene, there is a CpG island that corresponds to the promoter element of the *C16orf35* antisense transcript (Fig. 1B).

Besides three α -globin genes (π , α^D , and α^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 α^A gene.¹⁰ 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.¹⁰

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>

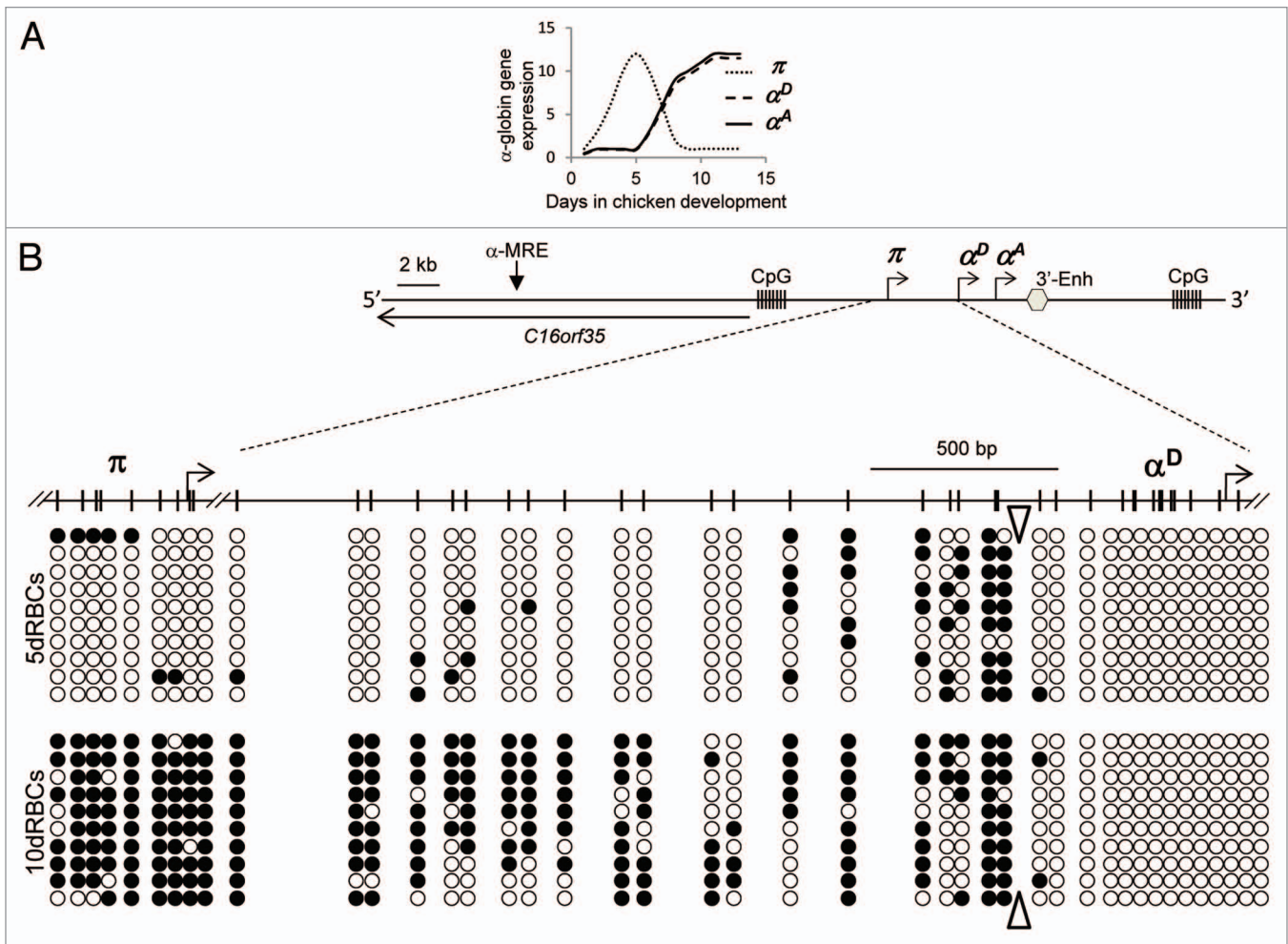


Figure 1. Different DNA methylation patterns of the chicken α -globin locus during development. **(A)** Differential α -globin gene expression profiles during chicken development.³ **(B)** Scheme of the α -globin domain showing the location of intergenic CpG-islands, the α -MRE locus control region and 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 π 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.^{3,4,11} 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 π gene.³ We have shown that the embryonic π 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 α -globin genes late in development.³ 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 α -globin genes.^{3,10,12} One conclusion of those studies is that the autonomous silencing of the embryonic π 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 π gene, its promoter and the 3' intergenic sequences up to the adult α^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 α^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 π gene, which subsequently allows activation of the adult α^D and α^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 π globin gene and surrounding genomic sequences during development.

Based on our previous observations that the embryonic π gene promoter is not DNA methylated in 5-days-old chicken embryos (5dRBCs) but is methylated in 8dRBCs and 10dRBCs,³ we asked whether such methylation is restricted to the promoter or if it can be expanded, in particular, toward the adult α^D gene (Fig. 1B). This aim is further supported by the fact that DNA methylation can be propagated along the genome.¹³⁻¹⁵ We performed sodium bisulfite transformation of genomic DNA and sequencing from 5dRBCs (where only the embryonic π gene is expressed) in comparison to 10dRBCs (where the adult α^D and α^A genes are expressed) (Fig. 1A and B). We found that the promoter regions of the π and α^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 α^D gene (Fig. 1B). In contrast, in 10dRBCs, the π promoter and the intergenic region between the π and α^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 π gene and the transcription of the adult α^D and α^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.¹⁶ Importantly, these cells do not express globin genes, even though they possess a permissive globin domain chromatin structure.^{10,17} Interestingly, these cells can specifically transcribe the adult α^D and α^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).¹⁷ Total RNA and genomic DNA were then isolated to verify the induction efficiency through the transcriptional activation of the adult α^D and α^A genes (Fig. 2B).^{3,17} 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 α^D and α^A gene expression was initiated, we observed a localized DNA demethylation of the α^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 α^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 α^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).³

Since there is growing evidence supporting active enzymatic reactions capable of erasing or modifying pre-existing DNA methylation patterns,¹⁸ we first investigated the possibility of active DNA demethylation over the α^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).¹⁹ 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 α^D gene (Fig. 2C). We do not believe that DNA demethylation of the α^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 π gene. This finding suggests an indirect correlation between DNA demethylation and transcriptional activation of the adult α^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 α -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.²⁰ In this report, we found more than 5,000 CTCF binding sites were identified in 5dRBCs, and more than 21,000 in 10dRBCs,²⁰ among them a novel CTCF-binding site that was found 190 bp upstream of the α^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

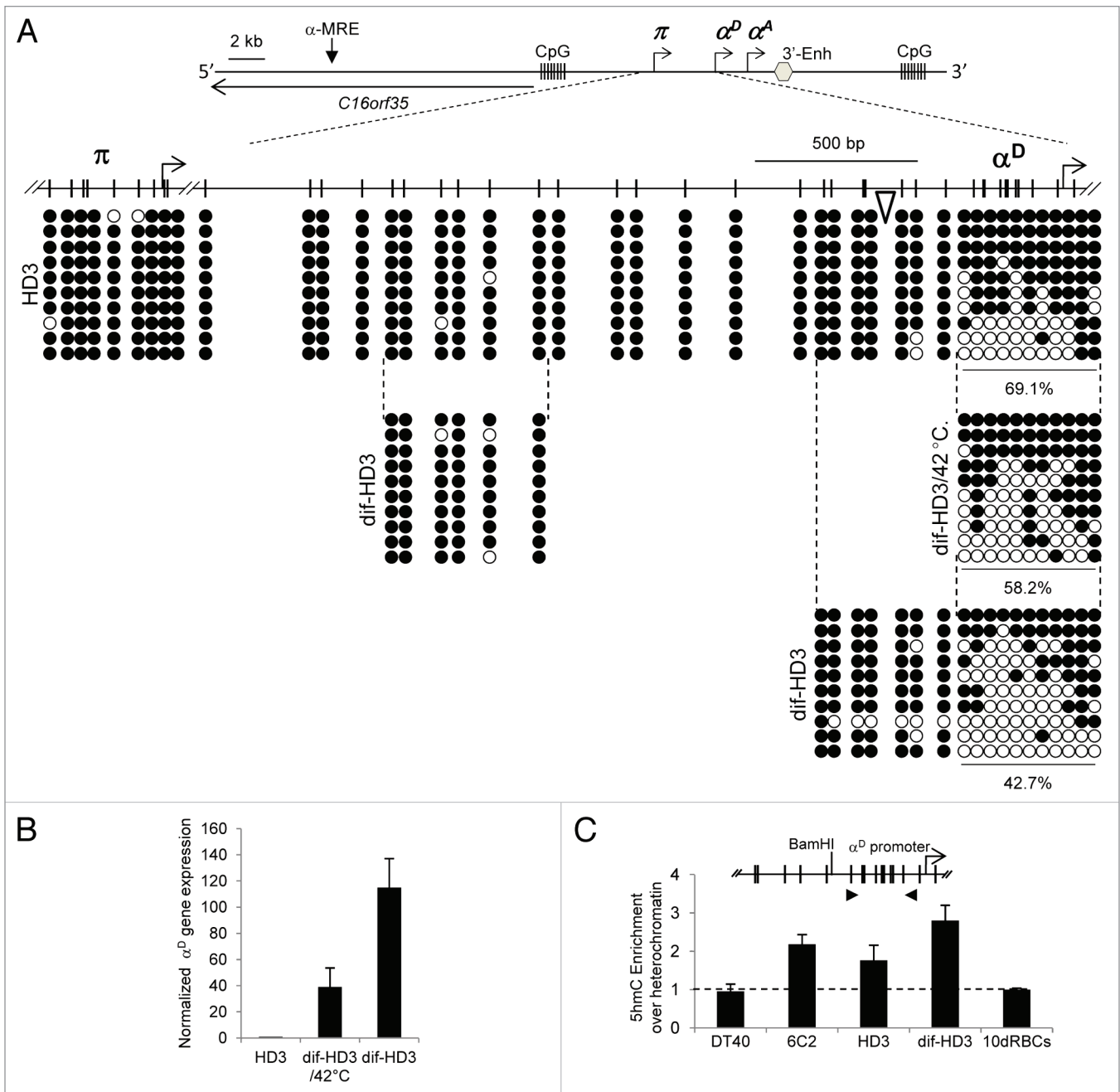


Figure 2. DNA demethylation and gain of the 5hmC modification correlates with adult α^D gene transcriptional activation. **(A)** Analysis of the DNA methylation status in the transformed avian HD3 cell line in non-induced (HD3) and differentiation-induced (dif-HD3; Iso-H-7 and 42°C) conditions. For comparative 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 α^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 α^D gene transcriptional activation. **(C)** Quantitative 5-hydroxymethylcytosine-DNA immunoprecipitation was assessed using an antibody against 5hmC and primers located in the α^D promoter. As controls we used bisulfite transformation to test the chicken B-lymphocyte DT40 cell line, where the α^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 α^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.²¹ Instead, we observed a constant incorporation of active histone marks like H3ac, H4ac, and H3K4me2 (Fig. 4B).

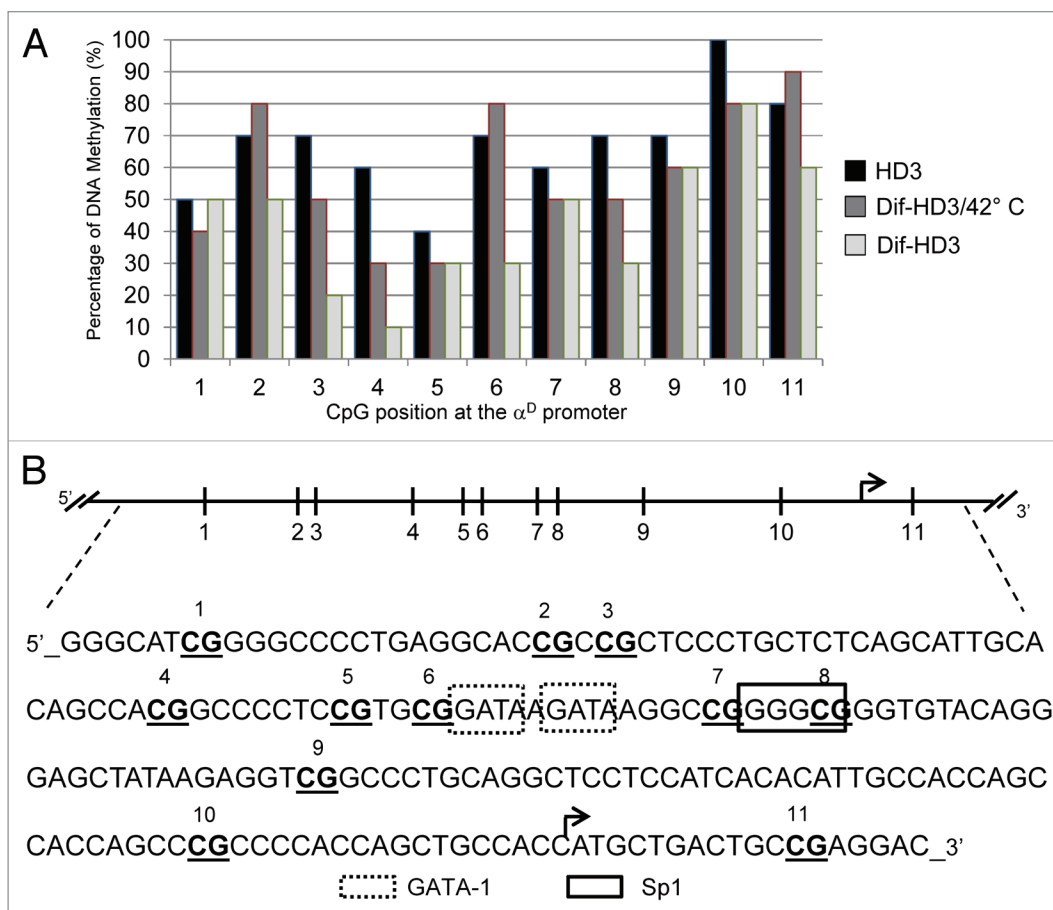


Figure 3. DNA methylation overview at individual CpG-dinucleotides in the α^D promoter. **(A)** A comparative percentage of DNA methylation is shown in the graph. The data are generated from the results in Figure 2A. HD3 cells represent undifferentiated erythrocytes, dif-HD3/42°C which corresponds to 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 α^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 α^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.³

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 α^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 π 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 α^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.^{7,20,22} 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,²⁰ we analyzed the genomic sequence in more detail and found two potential binding motifs about 190 bp upstream of the α^D gene TSS (Fig. 5A).²⁰ 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 α^D promoter activity were removed.^{3,12} These constructs were stably transfected in HD3 cells and pools were selected and maintained in continuous cell culture. *GFP*

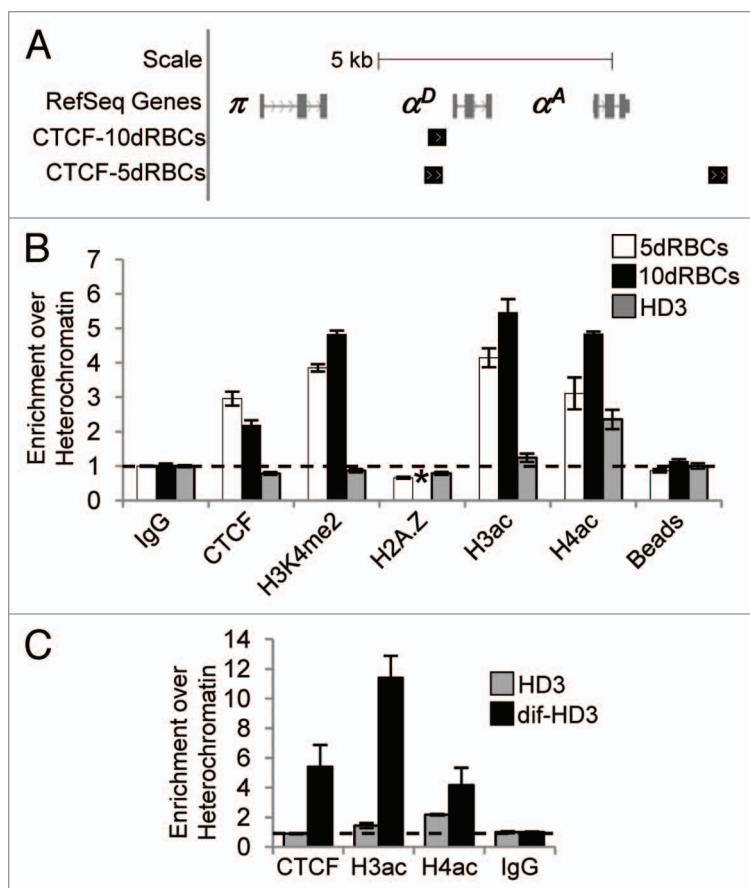


Figure 4. In vivo association of the CTCF nuclear factor and histone marks over the adult α^D gene promoter. (A) Location of the new CTCF binding site determined in a previous publication by chromatin immunoprecipitation and massive sequencing (ChIP-seq);²⁰ black squares represent CTCF in vivo occupancy (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 α^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 (αDIG) 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 α^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' α -globin enhancer in our transgenes.^{3,10} 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.^{23,24} 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,²⁰ we validated another CTCF site in the 3' enhancer (Fig. 7 and B). Thus, the presence of CTCF at both the α^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 α^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 α^D promoter region.

To further explore the contribution of CTCF in demarcating the DNA methylation sub-domain, we surveyed the DNA methylation of the α^D gene promoter in the transgenes both with and without the CTCF-binding sequence (αDIG 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.²⁵ In addition, even in the presence of the 3'

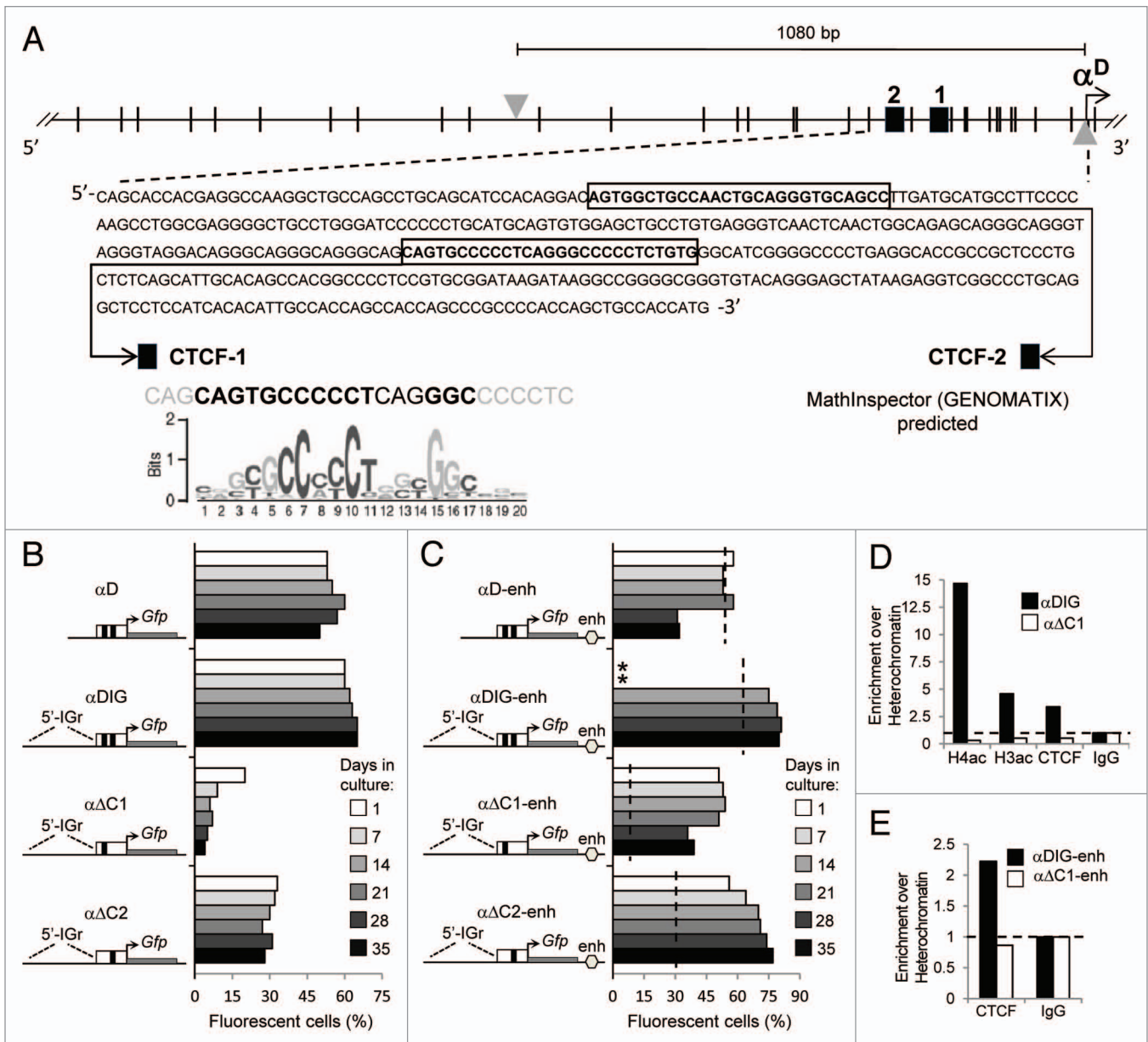


Figure 5. CTCF is a key component for α^D gene promoter function. **(A)** Schematic representation and nucleotide sequence of the putative CTFC binding motifs in the π - α^D intergenic region. Using the Logo generated from the chicken CTFC ChIP-seq experiments we identified a potential binding motif located 190 bp upstream of the α^D -TSS (CTCF-1, black square 1).²⁰ 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 α^D promoter (white box in schemes) containing the putative CTFC-binding sites (black boxes inside; α^D and α^{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 α^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 α^{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 α^D gene TSS represents a barrier element that counteracts the propagation of silencing chromatin, in particular, DNA methylation. In

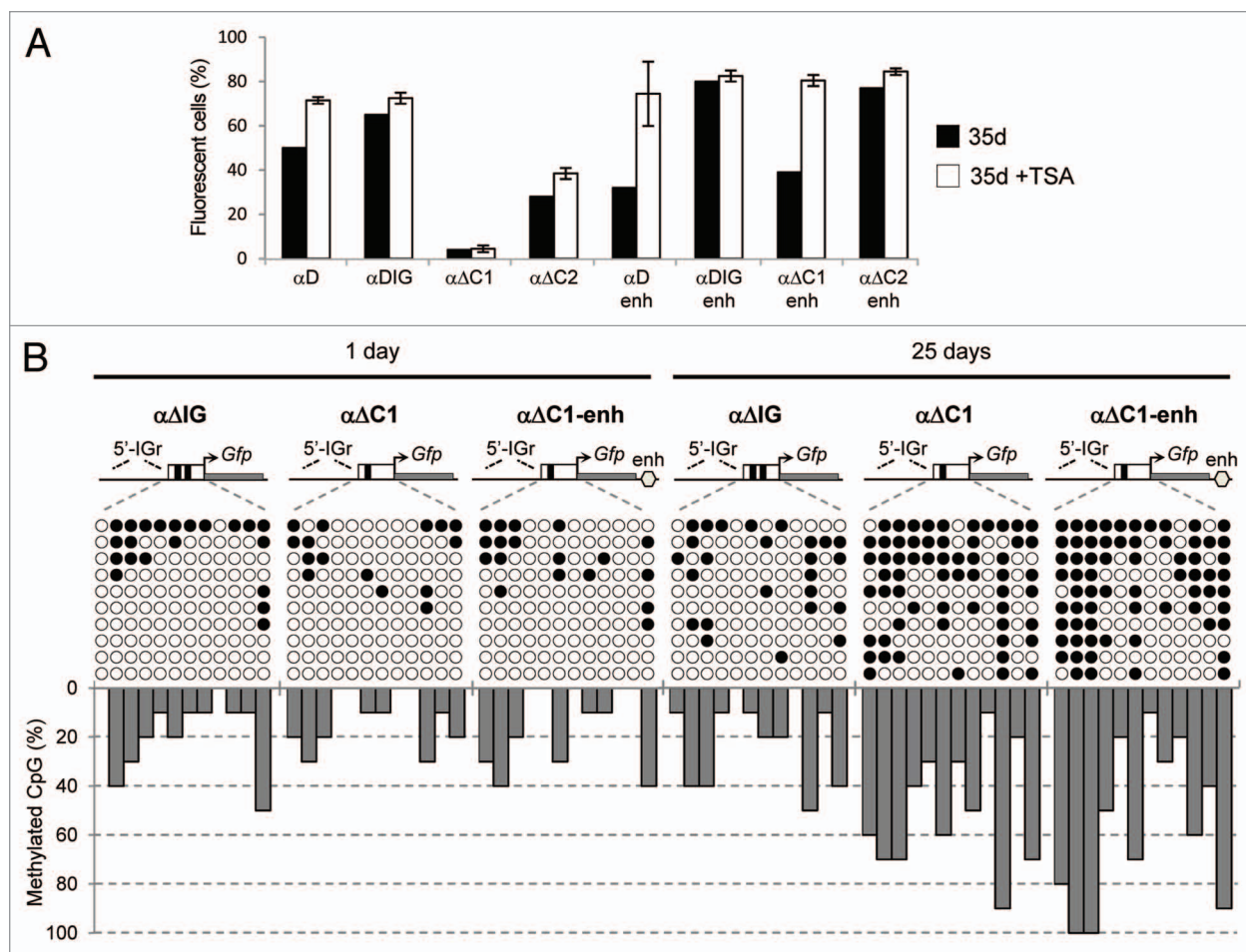


Figure 6. Transgene reactivation assays and DNA methylation analysis. (A) TSA-treatment at day 35 of culture. (B) DNA methylation analysis over the transgene. The percentage at individual CpG-dinucleotides of the transgene comparing day 1 and day 25 of continuous cell culture is shown. Interestingly, 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 α^D gene.

Stage-specific chromatin loop formation between the embryonic π 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.²⁶ The spatial organization of the chicken α -globin locus has already been established during erythroid differentiation (HD3 cells) and only during late development, revealing the differential formation of chromatin hubs.¹⁷

With all this in mind we asked whether there is a physical interaction between the 3' enhancer and the embryonic π 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).^{27,28} 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 π gene are DNA hypermethylated and the α^D gene is transcriptionally active. Of note, such interactions do not seem entirely dependent on CTCF, as suggested for other loci,²⁹ since we have not been able to demonstrate CTCF binding around the π 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 α^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 α^D promoter region as a constitutive anchor point for the α -globin domain chromatin hub.^{17,21} 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 π 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.³

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.^{14,30,31} Furthermore, the presence of CTCF and other nuclear factors upstream of the α^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).^{3,15,32,33} 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 β -globin *chs4* insulator elements.³³ Interestingly, mutation of the VEZF1 binding site in stably transfected constructs demonstrated the spreading of DNA methylation over the β^A promoter and transgene silencing.^{33,34} 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.⁷

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.³⁵ Furthermore, it has also been shown that broad histone H3 hyperacetylation occurs in CpG-islands.³⁵ 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.³⁶ 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.³⁶ 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

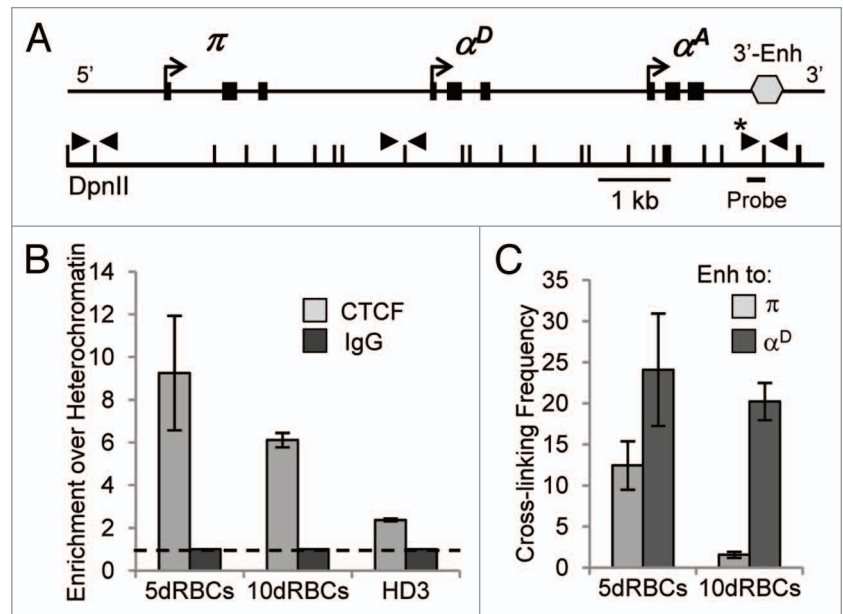


Figure 7. Physical contacts between the 3' enhancer and the π and α^D genes. (A) The scheme shows the distribution of the 4-base cutter DpnII restriction enzyme over the α -globin genes and the 3' enhancer (3'-Enh; black boxes represent gene exons). The arrowheads correspond to the test-primers and the 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.³⁷ 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.³⁸ Studies from our group in the chicken α -globin locus demonstrated that CTCF activity can be modulated by poly(ADP-ribosylation).⁷ Together, we conclude that the activity of CTCF located in proximity to the α^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 α -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 α^D gene is apparently a key contact point, which is coincident with the newly identified CTCF-binding site.^{17,20} In such a scenario, one possible model is that the embryonic π 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 β -globin locus.³⁹ Furthermore, and as suggested previously,⁷ 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.

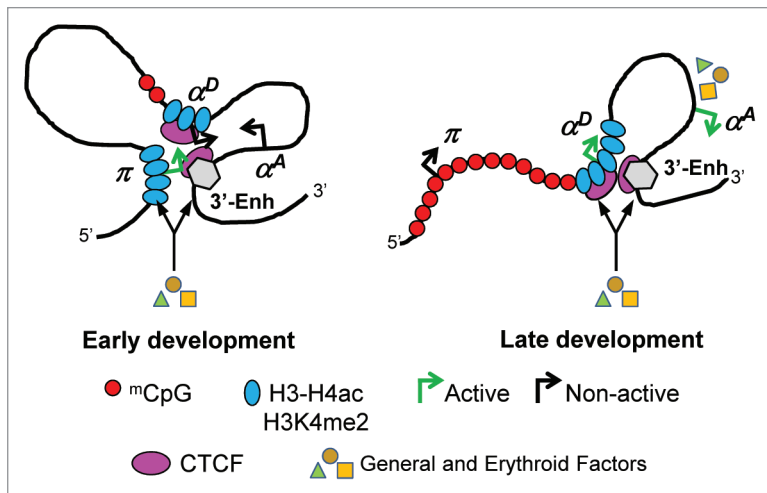


Figure 8. Autonomous silencing of the embryonic π gene and the derived topological conformation of the α -globin domain. In early development (5dRBCs), the model shows the formation of a hub composed by interactions between the π gene, the α^D promoter region and the 3' enhancer. In late development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

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 α -globin locus is found in a constitutively open chromatin conformation.⁴⁰ Tissue- and stage-specific transcription factors are then needed for the activation of the embryonic π gene in a context in which the rest of the domain remains in an open chromatin structure.³ 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 π gene and surrounding genomic regions through DNA methylation in an autonomous way (Fig. 8). As a consequence, the embryonic π 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 α^D and α^A genes in later stages of development. A controversy has been raised arguing that there is no silencing in any part of the chicken α -globin locus.²¹ 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.²¹ Our results support combinatorial regulatory events which could be similar to those observed for RAR β 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.^{41,42} 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.³ To induce erythroid differentiation, HD3 cell line were treated with Iso-H-7 (20 μ 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% CO₂ 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.³ For drug treatments, Trichostatin-A (TSA, Sigma) and 5-aza-2'-deoxycytidine (5-azadC, Sigma) were used at 2.5 ng/ml and 3 μ M, respectively.

Plasmids. The pG α^D , pG α^D DIG, pG Δ C1, and pG Δ C2 plasmids and their associated versions containing the 3'-enh were generated by PCR amplification of the α^D promoter (Fig. S4), the intergenic region between the π and α^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 μ 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.³ 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 α^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 α^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.⁴³ 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 α^D promoter amplification values over a 5hmC-depleted region (HI, repetitive region in a heterochromatic context),^{3,44} 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 α -globin locus.²⁰ The accession codes for the ChIP-seq data previously published are the following: GEO: GSM691016 for 5dRBCs and GSM691017 for 10dRBCs.²⁰ The identified CTCF binding sites were validated by quantitative PCR (see below).

Chromatin immunoprecipitation analysis. Chromatin was prepared for immunoprecipitation as previously described.³ The chromatin was immunoprecipitated with anti-CTCF antibody,⁴⁵ 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).^{3,46} In the inverse sense, for close chromatin marks, the enrichment was calculated over a constitutive open region (FII in Fig. S2A).^{3,46} 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.²⁷ In brief, 4×10^7 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 1×10^6 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 1×10^6 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 α -globin switching depends on autonomous silencing of the embryonic π 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 β -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 α -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 α -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 α 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 α -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 α , π , the chicken embryonic α -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. Gavrilo AA, Razin SV. Spatial configuration of the chicken α -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, Gavrilo AA, Razin SV. The inactivation of the π gene in chicken erythroblasts of adult lineage is not mediated by packaging of the embryonic part of the α -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, Gavrilo 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 β -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 β -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-Dnmt1 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 β -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, Roloff TC. 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
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 α -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>

A long non-coding RNA promotes full activation of adult gene expression in the chicken α -globin domain

Cristian Arriaga-Canon, Yael Fonseca-Guzmán[†], Christian Valdes-Quezada[†], Rodrigo Arzate-Mejía, Georgina Guerrero, and Félix Recillas-Targa*

Instituto de Fisiología Celular; Departamento de Genética Molecular; Universidad Nacional Autónoma de México; Distrito Federal, México

[†]These authors contributed equally to this work.

Keywords: long non-coding RNA, chromatin, globin genes, erythroid differentiation, development.

Long non-coding RNAs (lncRNAs) were recently shown to regulate chromatin remodelling activities. Their function in regulating gene expression switching during specific developmental stages is poorly understood. Here we describe a nuclear, non-coding transcript responsive for the stage-specific activation of the chicken adult α^D globin gene. This non-coding transcript, named α -globin transcript long non-coding RNA (lncRNA- α GT) is transcriptionally upregulated in late stages of chicken development, when active chromatin marks the adult α^D gene promoter. Accordingly, the lncRNA- α GT promoter drives erythroid-specific transcription. Furthermore, loss of function experiments showed that lncRNA- α GT is required for full activation of the α^D adult gene and maintenance of transcriptionally active chromatin. These findings uncovered lncRNA- α GT as an important part of the switching from embryonic to adult α -globin gene expression, and suggest a function of lncRNA- α GT in contributing to the maintenance of adult α -globin gene expression by promoting an active chromatin structure.

Introduction

Based on biochemical evidence more than 4 decades ago, Scherrer and Darnell proposed the existence of, what they called, “giant RNAs” as the major part of primary transcripts with high molecular weight.¹ Later, Scherrer, and collaborators described the presence of a long transcript, named the “Full Domain Transcript,” expanding 33 kilo bases (kb) and covering the entire α -globin locus.² More recently, genome-wide studies confirmed that the vertebrate genome encodes several thousands of non-coding short and long non-coding transcripts. Short non-coding RNAs (<200 nucleotides in length) include siRNAs, miRNAs, and piRNAs; and long non-coding RNAs (lncRNAs) (>200 nucleotides in length) include *Xist*, *HOTAIR*, and *HOTTIP* among many others.^{3,4}

lncRNAs affect many biological processes, such as genomic imprinting, gene dosage compensation, gene expression, and nuclear organization. In addition, they are associated with diseases like cancer.^{4,6} lncRNAs are expressed cell-specifically and in response to different stimuli, suggesting that their transcription is tightly regulated.^{4,7} Indeed, a genomic scale survey of the enrichment of different histone marks showed that, similar to protein-coding genes, non-coding RNAs have a characteristic signature with a short stretch of trimethylation of histone H3

at lysine 4 (H3K4me3), which marks transcriptionally active chromatin and is typically located in promoter regions.⁸ In addition, H3K36me3, associated to transcriptional elongation, covers the majority of lincRNAs.^{9–11} Four models have been proposed for the function of the lncRNAs: (1) as regulatory signals, (2) as a decoy system, (3) as guides of regulatory components to the genomic targets, and (4) as a ribonucleoprotein scaffold.^{3,4,11} However, the evidence indicates that lncRNAs function mainly through modulating chromatin-remodeling complexes.

Despite the evidence of the presence of non-coding RNAs in the α -globin domain,² little is known about their identity or function in the regulation of the locus. Therefore, we performed a detailed analysis of the intergenic transcripts distributed along the α -globin domain and describe a nuclear-restricted lncRNA, called α -globin transcript long non-coding RNA (lncRNA- α GT), which is transcribed in sense with the globin genes, and whose promoter drives erythroid-specific expression. The relative abundance of lncRNA- α GT is orders of magnitude less than the α -globin genes. This lncRNA is present in late stages of development and in terminally differentiated erythroid cells. Ablation of lncRNA- α GT caused down-regulation of the adult α^D gene, and its presence correlated with enrichment of H3K4me3 over the adult α^D gene promoter. Together, our results suggest that

*Correspondence to: Félix Recillas-Targa; Email: frecilla@ifc.unam.mx

Submitted: 09/18/2013; Revised: 10/25/2013; Accepted: 10/31/2013; Published Online: 11/06/2013
<http://dx.doi.org/10.4161/epi.27030>

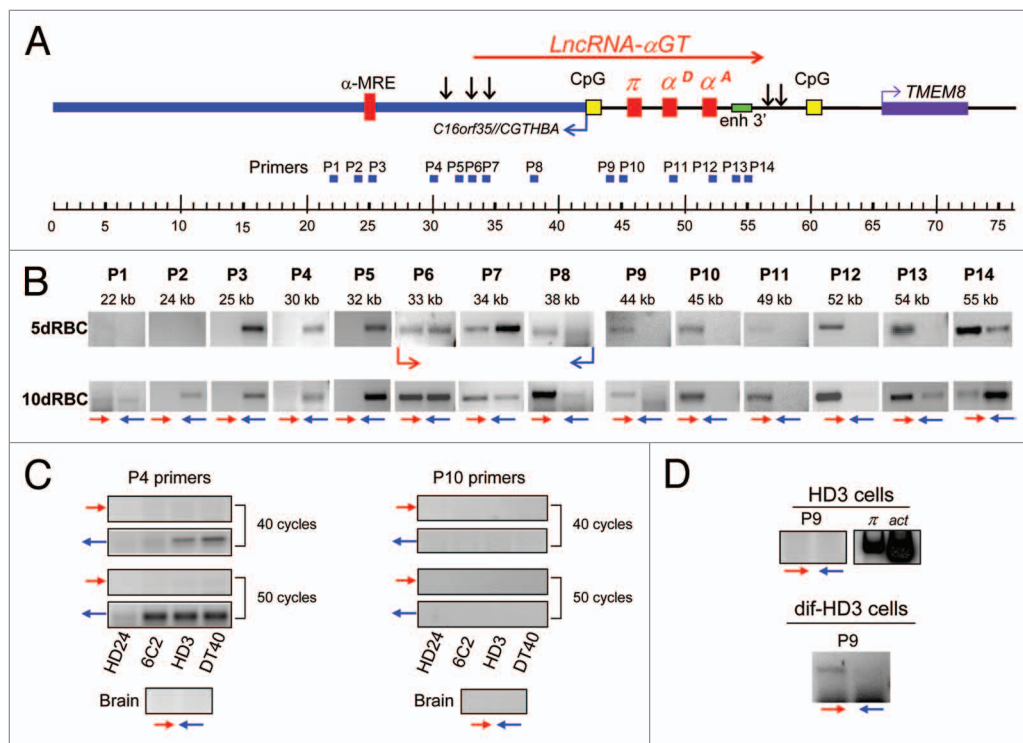


Figure 1. Systematic analysis of the non-coding sense and protein-coding transcripts along the chicken α -globin locus. **(A)** Scheme of the α -globin domain showing the location of intergenic CpG-islands (yellow boxes), the α -MRE locus control region, and the protein-coding $C16orf35/CGTHBA$ transcript. The black vertical arrows represent DNase I-hypersensitive sites. The blue squares symbolize the location of the primers used for the chain-specific RT-PCRs shown in B (P1 to P14). The $TMEM8$ gene is shown and the extension of the $LncRNA-\alpha GT$ is shown as a red arrow. **(B)** Reverse transcription and PCR-amplification of sense (red arrows) and protein-coding (blue arrows) transcripts from primary cells of 5 d and 10 d chicken embryo erythrocytes, 5dRBCs, and 10dRBCs, respectively. Estimated transcription initiation for the $LncRNA-\alpha GT$ (red arrow between P5 and P6) and $C16orf35/CGTHBA$ (blue arrow between P8 and P9) are shown. **(C)** RT-PCR using total RNA from erythroid (HD24, 6C2, and HD3 cell lines) and non-erythroid (DT40 cell line and chicken brain cells) are shown. PCR amplification was performed with primers P4 and P10. **(D)** HD3 cells induced to differentiation (dif-HD3) illustrate the presence of the $LncRNA-\alpha GT$ (primers P9). This set of experiments is representative of three independent assays and all the positive and negative controls are shown in Figure S1.

$LncRNA-\alpha GT$ is important for full activation of adult chicken α -globin genes, and favors maintenance of promoter structuring into transcriptionally permissive chromatin.

Results

A long non-coding transcript in the α -globin gene locus is expressed during chicken erythroid differentiation and development

The chicken α -globin locus, containing the embryonic π gene and adult α^D , and α^A genes, is organized in constitutive open chromatin on minichromosome 14 (Fig. 1A)^{12,13} and contains several regulatory elements in both the 5' and 3' ends of the locus. In the 5' non-coding end, there is a distal locus control region, the α -major regulatory element (α -MRE), a CTCF-dependent insulator (α EHS-1.4), and the $C16orf35/CGTHBA$ gene, which encodes a processed protein-coding transcript (Fig. 1 and not shown). In the 3' side of the locus, there is an erythroid enhancer upstream the adjacent domain, defined by the gene $TMEM8$.

The "Full Domain Transcript" of 33 kb covering the entire domain has not been described in detail,² and its function is unknown. As a first step to describe the "Full Domain Transcript," we performed reverse transcription followed by PCR (RT-PCR) using a set of 14 different pairs of primers (P1-P14) designed to amplify from the sense and antisense cDNA strains and targeting different segments along the locus (Fig. 1B–D). We used total RNA isolated from 5 and 10 d old chicken embryonic red blood cells (5dRBCs and 10dRBCs, respectively) (Fig. 1B). We detected a sense transcript starting between primers P5 and P6, around 12 kb upstream the embryonic π gene, overlapping three DNase I hypersensitive sites,¹³ and ending between probes P13 and P14, 3 kb downstream of the adult α^A gene (Fig. 1B). Our results suggest that this sense transcript covers a sub-region of the domain and is not transcribed from the α -MRE as previously suggested.² However, we are unable to determine if the $LncRNA-\alpha GT$ represent a continuous transcript. As expected, amplification from the antisense strand detected the $C16orf35/CGTHBA$ transcript starting from the CpG-island located around 4 kb upstream the embryonic π gene (Fig. 1A and B).¹⁴ Control experiments showed that that the sense and the $C16orf35/CGTHBA$ protein-coding transcripts are present in a very low abundance as compared with protein-coding genes in the domain (Fig. S1; see below).

Interestingly, a short genomic region of around 1 kb overlapping both the sense and antisense transcripts in the α -globin domain could potentially interact with the RNA double stranded RNA interference machinery due to its potential capacity to generate double-stranded RNA. Therefore, we addressed whether the corresponding genomic region could encode small RNAs. Purification of small molecular weight RNAs and northern-blot do not allow us to find any small RNAs (data not shown). Our results suggest that a long non-coding RNA of about 23 kb is transcribed in the α -globin domain, we called this transcript chicken long non-coding RNA α -globin transcript or $LncRNA-\alpha GT$. Next, we addressed the

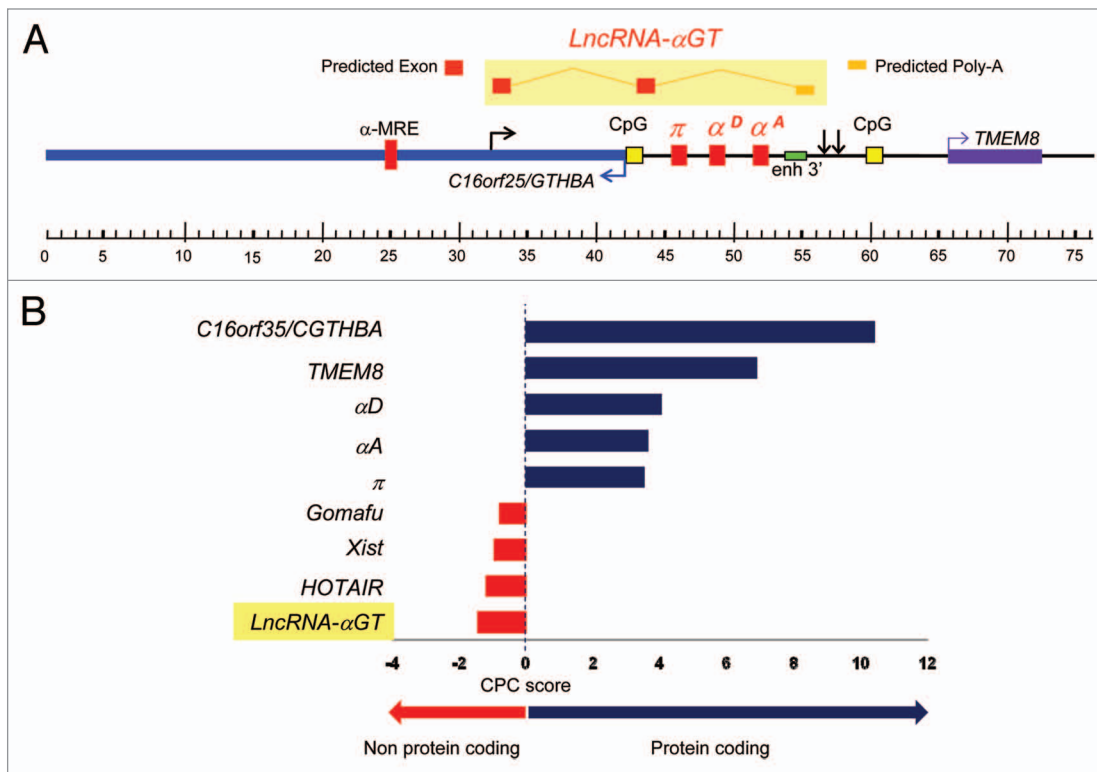


Figure 2. In silico prediction of the protein-coding capacity of the lncRNA- α GT and its intron-exon organization. **(A)** Bioinformatics annotation of lncRNA- α GT. A genomic sequence of ~25 kb from the α -globin gene cluster (NCBI AY016020.1) was used to computationally identify potential exons, introns, polyA signals, and regulatory regions, like promoters, of lncRNA- α GT. **(B)** Computational assessment of protein-coding capacity of lncRNA- α GT. In order to identify if lncRNA- α GT could potentially encode for a protein, we used a novel program, Coding Potential Calculator (CPC). CPC is a program that has been trained using high quality sequences of protein coding and non-protein coding RNAs. This program is able to identify those sequences that have a great probability of encoding a protein based on ORFs identification and homology searches. Sequences spanning the putative transcript of lncRNA- α GT were analyzed. Coding regions of π , α^D , α^A , *TMEM8*, and *C16orf35/CGTHBA* were used as controls for genes with validated protein coding capacity, and *Gomafu*, *Xist*, and *HOTAIR* sequences were used as control for validated lncRNA genes. All sequences have a score based on their estimated protein-coding capacity: <-1, non-coding; from -1 to 0, weakly non-coding; from 0 to 1, weakly coding; and >1, coding. Note that sequences spanning lncRNA- α GT have the highest value for non-coding transcripts.

expression of lncRNA- α GT in chicken erythroid cell lines that express α -globin adult genes when terminally differentiated HD3 cells (dif-HD3),^{15,16} non-erythroid DT40 lymphoid cells, and in primary chicken brain cells, and HD24 and 6C2 chicken erythroid cells (Fig. 1C). We used two primer pairs for RT-PCR, P4, which targets the *C16orf35/CGTHBA* protein-coding gene and P10, which targets lncRNA- α GT. We detected the *C16orf35/CGTHBA* protein-coding transcript in 6C2, HD3, and DT40, but not in HD24 or brain cells, after 40 and 50 cycles of PCR amplification with the P4 primer pair (Fig. 1C). When P10 primer pair were used to target the lncRNA- α GT no amplification products were detected (Fig. 1C). We further demonstrate the absence of the lncRNA- α GT transcript in undifferentiated HD3 cells, which do not express globin genes, by a more sensitive approach (a radioactive PCR) using the P10 primer pair (Fig. S1). Next we addressed the presence of lncRNA- α GT differentiated erythroid cells by RT-PCR using primer pair P9 on HD3 cells, which activate expression of the adult α^D and α^A genes upon differentiation induction.¹⁶ We detected lncRNA- α GT exclusively in the sense orientation in

differentiated HD3 cells. Thus, transcription of lncRNA- α GT correlates with full expression of α -globin adult genes (Fig. 1D).

Together, our analyses uncovered an erythroid-specific sense transcript, the lncRNA- α GT, spanning over a 23 kb sub-domain of the chicken α -globin locus, whose transcription correlates with α -globin adult gene expression.

High-molecular weight non-coding sense transcripts are produced from the α -globin domain during chicken development

We first analyzed the theoretical coding potential of lncRNA- α GT using the Coding Potential Calculator (CPC) program (Fig. 2; Table S1).¹⁷ This program is a support vector machine-based classifier that assesses the protein-coding potential of a transcript based on different sequence features, like high quality ORF identification. The high number and quality of data sets used to train the model enables CPC to discriminate coding from non-coding transcripts with high accuracy. Using the *Gomafu*, *Xist*, and *HOTAIR* long non-coding transcripts as controls, our analysis in silico showed that the lncRNA- α GT is a bona fide non-coding transcript (Fig. 2B).^{3,4} Furthermore, the analysis predicted the presence of two exons in lncRNA- α GT (Fig. 2A).

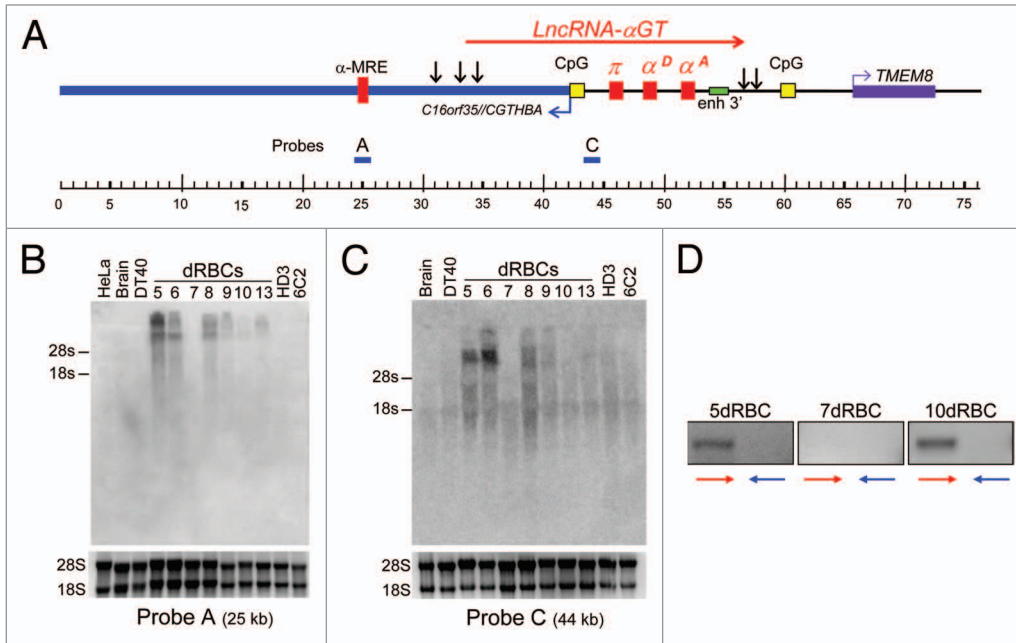


Figure 3. High-molecular weight transcripts in the 5' non-coding are of the chicken α -globin domain. (A) Scheme of the chicken α -globin domain showing the location of the DNA double-stranded probes A and C. (B) Northern-blot hybridized with probe A. No signal was detected for HeLa, brain, DT40 non-erythroid cells, as well as, for the erythroid-transformed HD3 and 6C2 cells. Notice that absence of hybridization signal in 7dRBCs. Additional blots, hybridized with different probes are shown in **Figure S2**. (C) Northern-blot as in (B) using the probe C that eventually only detects the sense transcript, i.e., the lncRNA- α GT. Ethidium bromide stained gel is shown to demonstrate the total RNA integrity and loading conditions. (D) To increment the sensitivity and confirm the absence of the lncRNA- α GT in 7dRBCs, RT-PCRs were performed forcing the amplification to 40 cycles.

Then, we addressed the size of this non-coding transcript and its relative abundance in distinct cell types and during different days of chicken development by northern blot. The first probe used was double stranded and corresponds to the α -MRE upstream regulatory element (**Fig. 3B**, Probe A). The second probe should only hybridize with the sense transcript and corresponds to a region between the CpG-island and the embryonic π gene (**Fig. 3C**, Probe C). We found hybridization signals corresponding to transcripts larger than 10 kb whose precise molecular weight was difficult to estimate. Unexpectedly, we consistently noticed absent hybridization signal in 7dRBCs (**Fig. 3B**; **Fig. S2**). Interestingly, hybridization with probes recognizing the adult α^D and α^A promoters, probes E and F, respectively, revealed several bands of lower size, suggesting the presence of processed RNA products (**Fig. S2**). Strand-specific RT-PCR on 7dRBCs using the P9 primer pair confirmed the specific absence of transcripts at day 7 of embryonic development (**Fig. 3D**), when the embryonic π gene is silenced and the adult α^D and α^A genes are transcribed.^{15,16} These results confirm the existence of non-coding RNAs of more than 10 kb long in the α -globin domain, and intriguingly suggest that both lncRNA- α GT and the protein-coding gene *C16orf35/CGTHBA* transcript are not expressed or degraded at the moment of embryonic to adult gene switching.

Identification of an erythroid-specific lncRNA- α GT promoter

The absence of non-coding RNAs in 7dRBCs prompted us to address whether lncRNA- α GT is negatively regulated or degraded. To this end, we aimed at identifying its promoter. We tested a DNA sequence of 1 kb long, located upstream of the P6 primer pair for promoter activity (**Fig. 1A** and **B**). In silico analysis predicted a promoter in a genomic region around the position 32 kb (**Fig S3**), located upstream the primer pair P6. Indeed, strain-specific RT-PCR estimated the transcription start site of lncRNA- α GT between 12 and 13 kb upstream the embryonic π gene, overlapping with the predicted promoter (**Fig. S4**). Surprisingly, no-erythroid specific transcription factors were predicted, but nuclear factors like GR and ER were (**Fig. S3**). Based on these results we tested the activity of a 1 kb DNA fragment overlapping the predicted promoter region.

Luciferase reporter plasmids were transiently transfected in HD3 cells and in chicken primary fibroblasts. The DNA fragment tested drove gene expression in erythroid cells, and in 5, 7, and 10dRBCs when cloned in the 5' to 3' orientation (**Fig. 4B** and **C**; **Fig. S5**). However, it was not active in fibroblasts (**Fig. 4C**), DT40 cells, or human HeLa cells (**Fig. S5**). To further support the presence of a regulatory element upstream lncRNA- α GT we addressed the local enrichment of the histone mark H3K4me3, which is characteristic of promoters. We designed primers around the estimated transcription start site of lncRNA- α GT to perform a chromatin immunoprecipitation (ChIP). We found that H3K4me3 occupied the promoter region in 5 and 7dRBCs and increased as development progresses, with a notorious increase in 10dRBCs (**Fig. 4D**), where adult α^D and α^A genes are actively transcribed. Furthermore, RNA polymerase II also occupied the promoter region in 10dRBCs (**Fig. 4E**).

Promoter activity and specific enrichment of H3K4me3 at later developmental stages raised the possibility that lncRNA- α GT is transcriptionally regulated during chicken development and that it might target the adult α^D and α^A genes.

The lncRNA- α GT transcript is upregulated in late stages of chicken development

To address the regulation of lncRNA- α GT during chicken development we performed quantitative RT-PCR (qRT-PCR) in 5, 6, 7, 8, and 10dRBCs. We found that lncRNA- α GT is a nuclear transcript expressed at significantly lower levels than the

α^D gene in 10dRBCs (Fig. S6). LncRNA- α GT expression increased modestly from 5 to 6dRBCs, and remained constant in 7 and 8dRBCs, before increasing significantly in 10dRBCs (Fig. 5A), coincident with silencing of the embryonic π gene and peaked expression of the adult α^D gene (Fig. 5B). These results opened the possibility that lncRNA- α GT may contribute to the transcriptional regulation of adult chicken α -globin genes.

To further corroborate the active chromatin configuration of the adult α^D gene promoter we performed a ChIP for H3K4me3.⁸ We found H3K4me3 enrichment in 5, 7, and 10dRBCs, with less enrichment in 7dRBCs (Fig. 5C). This agrees with previous findings showing that the entire locus is in an open chromatin configuration in 5dRBCs and that in 7dRBCs the embryonic π gene is silenced by DNA methylation, facilitating the switching to adult α^D gene transcriptional activation.^{15,16} Furthermore, decreased H3K4me3 on the α^D promoter correlate with very low levels of lncRNA- α GT in 7dRBCs (Fig. 5A), suggesting a possible role for 7dRBCs in the recruitment of this histone mark to the α^D gene promoter.

LncRNA- α GT is required for full activation of the adult α^D gene

To test the function of lncRNA- α GT in regulating transcription of the adult α^D and α^A genes, we knocked down 10dRBCs using a small hairpin RNA (shRNA) targeting the region between the CpG-island and the embryonic π gene promoter. This shRNA was cloned in the pSilencer vector under the control of the constitutive promoter of histone H1. The resulting vector was transfected transiently in 10dRBCs and stably in HD3 cells, which were induced to terminal differentiation to activate adult α^D and α^A gene transcription. qRT-PCR showed efficient lncRNA- α GT knockdown in differentiated HD3 cells (Fig. 6A). We found that lncRNA- α GT knockdown reduced mostly the expression of α^D , while α^A was only modestly down-regulated (Fig. 6B). LncRNA- α GT knockdown using three different siRNAs confirmed these results (Fig. S7).

To further confirm our results, similar experiments were performed in HD3 cells, which were now induced to terminally

differentiate using the ISO-H-7 compound, which favors more efficient differentiation (Fig. 6C).¹⁸ LncRNA- α GT was efficiently knocked down in these cells (Fig. 6D). We found that the lncRNA- α GT knockdown drastically affected the transcription levels of both α^D and α^A genes (Fig. 6C and D). Next, we addressed the effect of lncRNA- α GT knockdown on the chromatin structure of the α^D gene promoter. We performed ChIP for H3K4me3, acetylated histones H3 and H4, and the histone methyltransferase MLL1 (Fig. 6E). Interestingly, lncRNA- α GT knockdown reduced enrichment of H3K4me3 and acetylated histones H3 and H4, supporting a function of lncRNA- α GT in configuring active chromatin at the α -globin adult gene promoters. In contrast, lncRNA- α GT knockdown did not affect the chicken β -globin locus, which is located in a different chromosome (data not shown). Together, these results demonstrate that the lncRNA- α GT is required for full activation of adult α -globin gene expression in late developmental stages and in terminal erythroid differentiation. Furthermore,

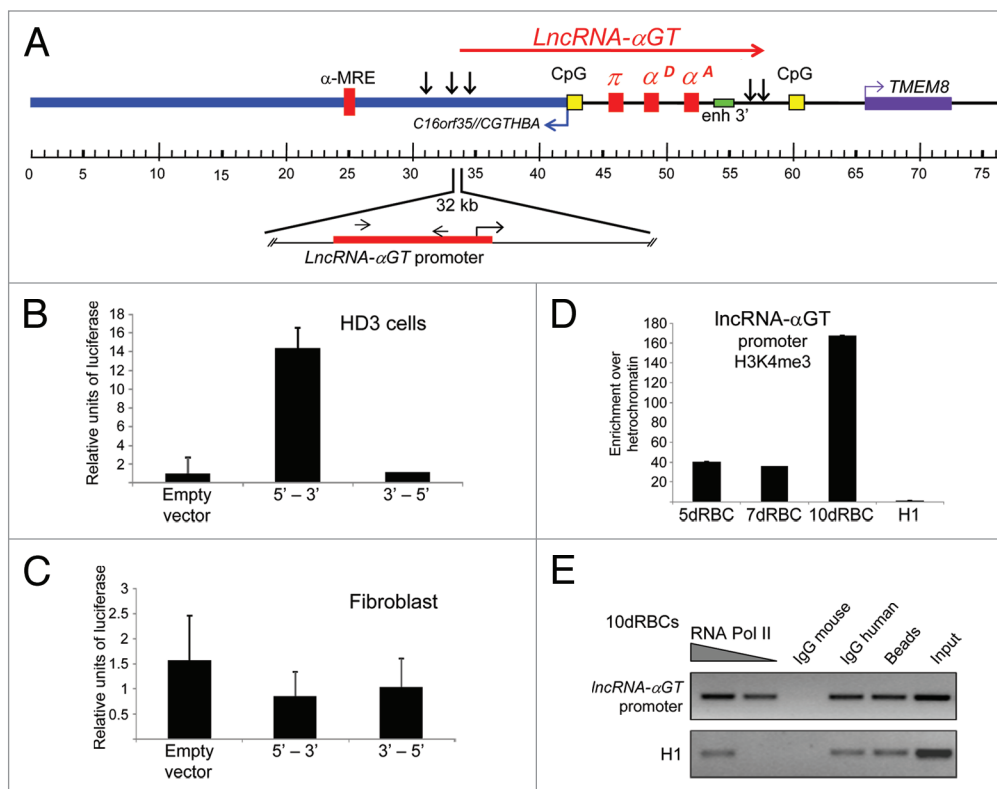


Figure 4. The lncRNA- α GT is transcribed from an erythroid-specific promoter. (A) Scheme of the chicken α -globin domain and the location of the in silico and experimentally defined transcription start region of the lncRNA- α GT (see Figs. S3 and S4). (B) Transient transfection of the lncRNA- α GT 1.1 kb promoter sequence driving the luciferase reporter gene in HD3 cells. The native 5'-3' and the opposite orientations were tested. The same experiment was performed in 5, 7, 10, and 13 d RBCs (see Fig. S5). (C) No promoter activity is found in transfected primary cultured chicken fibroblasts. (D) ChIP assay using an antibody against the promoter related histone mark H3K4me3 using primers located in the predicted lncRNA- α GT promoter elements. H1 primers are used as negative control (upstream β -globin heterochromatin region) (n = 3). (E) ChIP assay showing the enrichment of RNAP II in 10dRBCs over the lncRNA- α GT promoter region. Note, that the RNAP II enrichment should be compared with the mouse IgG lane, since both antibodies were generated in mouse. As control we used primers from a constitutive heterochromatin region (primers H1), located in the 16 kb region upstream of the chicken β -globin locus. This is a representative gel of two independent ChIPs.

lncRNA- α GT promotes structuring of the α^D gene promoter into open chromatin. Thus, lncRNA- α GT is an integral component of the regulatory and chromatin-associated machinery of the α -globin domain required for adult gene expression.

Discussion

The rigorous coordination between genetic and epigenetic regulation is required for proper stage-specific gene expression. Non-coding RNAs recently got into the scene as sensors and amplifiers of cellular signals targeting tens or even thousands of genes. However, their functions are still poorly understood. Here we characterized a ~23 kb long non-coding RNA (lncRNA- α GT) required for full activation of the adult α^D , and to a lesser extent α^A transcription, and for establishment of transcriptionally active chromatin.

We found that lncRNA- α GT promotes activation specifically of the α -globin adult genes. Accordingly, it is activated after the embryonic day 8 (Fig. 5A) concomitantly with down-regulation of the embryonic π gene and activation of the adult α^D and α^A genes (Fig. 5B).¹⁵ Our results support a model in which the lncRNA- α GT is not required in early developmental stages, when that the embryonic π gene is mainly regulated by the 3'-side enhancer (Fig. 1A).^{15,16} Later on, the embryonic π gene is autonomously silenced by DNA methylation¹⁶ and the expression of the adult genes α^D and α^A are activated. At this stage, lncRNA- α GT might be required to maintain adult gene expression.

The identification of the lncRNA- α GT and processed RNAs raise several new questions, starting by which are the mechanisms of such non-coding RNAs in terms of the regulatory signals, in particular, if such signals are restricted to later developmental and terminal differentiation stages. For example, our finding of RNAs of multiple sizes produced from the α^D and α^A gene promoters was intriguing (Fig. S2, Probes E and F). It is known that divergent transcription of short RNAs exist around the transcription start sites of many genes and that such RNAs derive from large RNA precursors with the ability to recruit chromatin remodeling activities.¹⁹⁻²¹ More recently, a large study on transcription of lncRNAs within promoters demonstrated that a significant amount of those promoters are transcribed according to cell cycle timing and that their disruption causes alterations in the expression of human genes affecting cell differentiation, DNA repair, or even in cancer.²² One specific example is PANDA, an lncRNA regulated by p53 who interacts with the transcription factor NF- κ B, resulting in miss-regulation of a set of pro-apoptotic genes.²² In addition, recent evidence suggest that active enhancer elements are also transcribed to produce ncRNAs interacting with nuclear factors, co-factors, cohesins, and even mediating long-distance chromatin interactions.²³⁻²⁶ This opens the possibility that the 3'-side enhancer, which is transcribed, is able to contact physically the promoter regions of the embryonic π and adult α^D genes in different moments, relevant for full activation of chicken α -globin genes, and participates in the maintenance of promoter permissive chromatin structure.^{16,18}

Another interesting aspect is to understand which are the mechanisms regulating the expression and the low nuclear levels

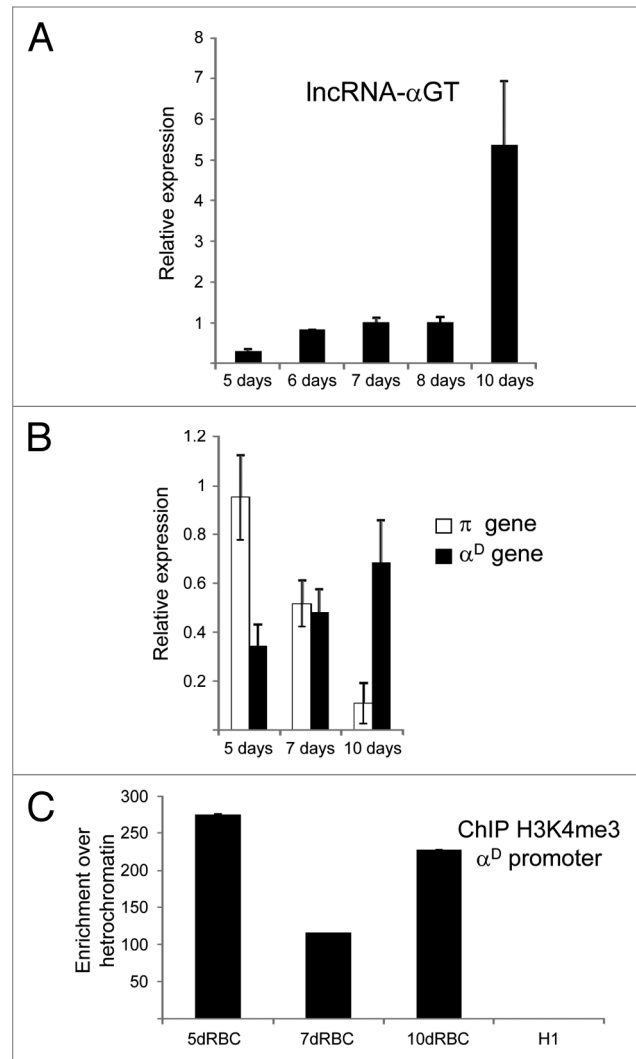


Figure 5. The long non-coding RNA, lncRNA- α GT, is transcribed in later stages of development. (A) Quantitative RT-PCR of the lncRNA- α GT in 5, 6, 7, 8, and 10 d of chicken embryo erythrocytes (5, 6, 7, 8, and 10dRBCs). (B) Expression pattern in the same days of the embryonic π gene and the adult α^D gene (n = 3). (C) ChIP assay using primers expanding the α^D gene promoter and an antibody recognizing the histone mark H3K4me3. H1 primers are used as negative control (upstream β -globin heterochromatin region). The constant enrichment of the histone H3K4me3 is in agreement with the constitutive open chromatin configuration of the α^D promoter during development.¹⁶

of the lncRNA- α GT. Our *in silico* analysis and the transcription factors associated the lncRNA- α GT promoter do not clarify this aspect since, unexpectedly, we have not identify classical erythroid transcription factors suggesting a more complex regulation, perhaps through an independent signal transduction pathway (Fig. S3). Otherwise, proteins like CTCF, distributed along the domain and apparently participating in the three-dimensional organization of the locus, can have a partnership with the lncRNA- α GT to guide its function.²⁷

Based on different examples of lncRNAs and their functional role, one can assume that their function is to regulate negatively gene expression and/or chromatin structure.¹¹ A growing list of

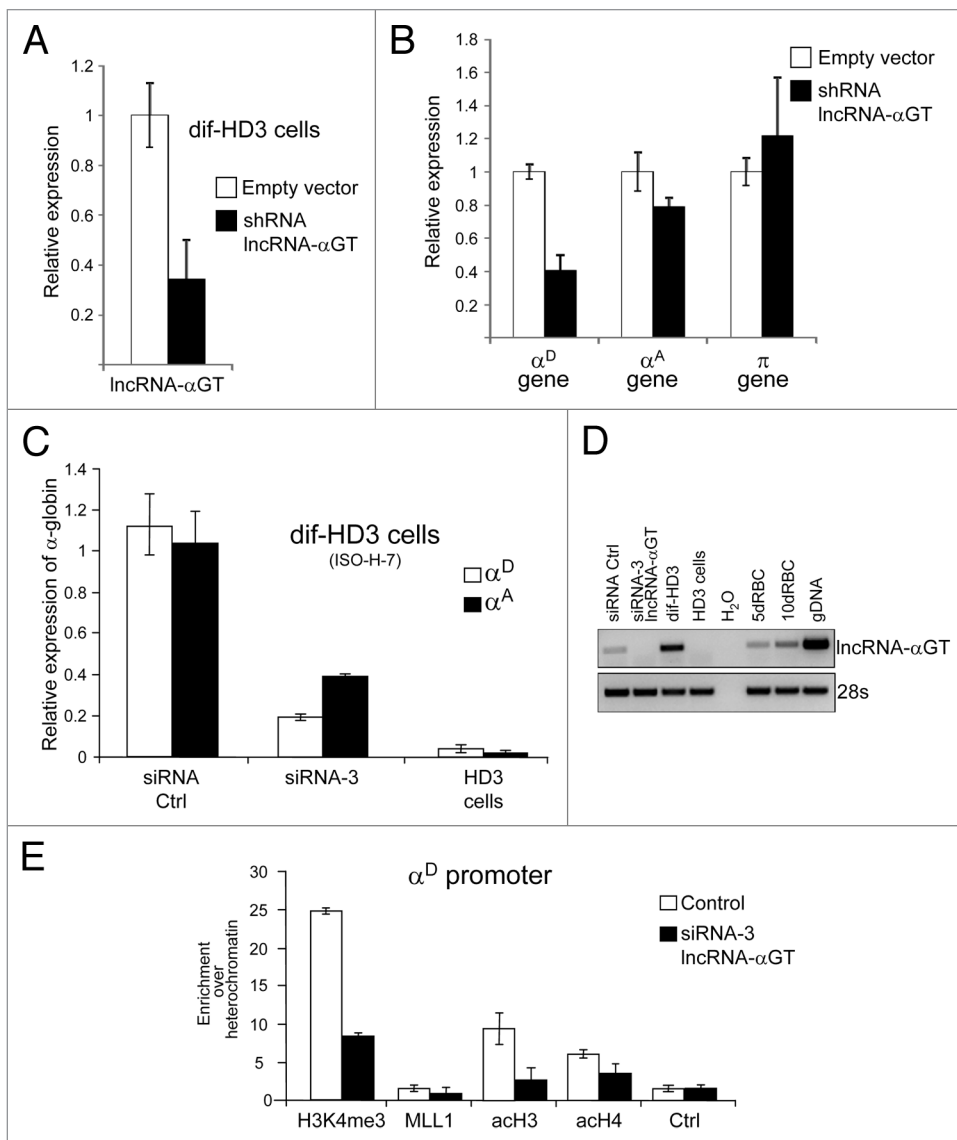


Figure 6. LncRNA- α GT depletion affects the adult α -globin genes expression. **(A)** Validation of the LncRNA- α GT shRNA knockdown in stably transfected HD3 cells that were induced to differentiation (dif-HD3). Mean expression patterns from three independent experiments are shown. **(B)** Effect of the LncRNA- α GT knockdown in the expression of the chicken α -globin genes in differentiated HD3 cells. **(C)** To confirm the effect of the LncRNA- α GT knockdown we transiently transfected the siRNA-3 into differentiated HD3 cells (dif-HD3; lane 2) and confirmed its effect. The down-regulation of α -globin adult genes was also corroborated and shown in **Figure S7**. **(D)** With optimal differentiation conditions (ISO-H-7) we confirmed one more time the down-regulation of α^D and α^A adult globin genes. **(E)** The LncRNA- α GT loss of function affects histone open marks at the adult α^D gene promoter. ChIP assay in differentiated HD3 cells (dif-HD3) using antibodies against different open histone marks and the histone methyltransferase MLL1. This ChIP assay was performed in dif-HD3 cells with primers expanding the adult α^D gene promoter. Error bars were calculated from 6 independent PCR reactions from two independent ChIP assays.

examples is revealing that this is not the only case. One of the seminal examples is *HOTTIP* lincRNA which is transcribed from the 5' end of the *HOXA* locus, acting positively over the proximal *HOXA* genes through the formation of chromosomal contacts, and the *WDR5* mediated recruitment of the MLL1 complexes to incorporate the active transcription associated to the histone mark H3K4me3.²⁸ This is in agreement with our data showing the

its mechanisms of action.

Material and Methods

Cell culture

The avian arrested cell lines were cultured as previously described.³¹ To induce erythroid differentiation, HD3 cells

enrichment of H3K4me3, not only at the α^D gene promoter but also at the LncRNA- α GT promoter (**Figs. 4D** and **5C**). Similar features are described for the lincRNA *Mistral*, also known as *Mira*, which is responsible for the activation of the transcription of the homeotic genes *Hoxa6* and *Hoxa7* in mouse embryonic stem cells by recruiting the MLL1 chromatin complex through an lincRNA-DNA hybrid interaction.²⁹ A more recent case is the non-coding transcription of a distal enhancer, named *utNgn1*, acting positively and in *cis* over the *neurogenin 1* gene expression. *Neurogenin 1* gene product is critical for neuronal differentiation during mammalian development.³⁰ These examples illustrate a direct or indirect interplay between the lincRNAs action which is mediated by the PcG and TrxG group of proteins. In the α -globin locus, the LncRNA- α GT way of action seems to be partially different since its upregulation coincides with a period of time where the adult α^D and α^A genes are already expressed, and not in their initial steps of activation (**Fig. 5A**). Furthermore, this is also in agreement with an enforced wave of active chromatin modifications starting upstream from the LncRNA- α GT promoter.¹³ In summary, the LncRNA- α GT helps to maintain histone marks like the H3K4me3 to adult α^D and α^A promoters in a stage-specific manner.

The overall organization of the chicken α -globin gene locus and the amount of different regulatory elements within the locus and the strategies for differential gene expression of globin genes demonstrate the complexity of such processes. Adding to such complexity the LncRNA- α GT represents a novel and sophisticated level of regulation that raises new challenges in terms of

were incubated in medium containing 10 mM HEPES (pH 8.0) (Sigma) and 20 μ M ISO-H-7 (1-(5-Isoquinoliny)sulfonyl)-3-methylpiperazine, Sigma) at 42 °C in 1% of CO₂ atmosphere for 48 h. RBCs were collected from 5 to 13 d old embryos as previously described.³¹

RT-PCR analysis strand specific

Total RNA was purified with TRIZOL according to manufacturer's instructions (Invitrogen) and was treated with DNase I (RQ1, Promega). To prepare cDNA, 5 μ g of RNA were reversely transcribed (Reverse Transcription System, Promega) using specific primers (Table S2). To initiate cDNA synthesis from the negative-strand a forward primer was used, while a reverse primer was used to initiate synthesis from the positive-strand. The products of PCR reactions were analyzed by agarose gel electrophoresis with ethidium bromide staining.

Quantitative PCR

RNA levels were quantified by Real-Time PCR. The cDNA was synthesized using AMV reverse transcriptase and random primers (Reverse Transcription System, Promega). Samples were treated with DNase I (RQ1, Promega). Quantitative RT-PCR measurements (qRT-PCR) were done by using SYBR green PCR mix (Sigma). The PCR reaction was performed with StepOne detection System (Applied Biosystems) at 94 °C for 2 min, followed by 40 two-step cycles of 94 °C for 15 s and 60 °C for 20 s, triplicates were made for each sample. Relative mRNA levels of α -globin genes were calculated using the comparative $\Delta\Delta$ Ct method. The transcript levels were normalized against 28S rRNA abundance when primary erythrocytes were used, in contrast, for transformed chicken erythroid cell lines β -actin was used for normalization.

Northern blot analysis

Total RNA (20 μ g) was used for blot analysis. After electrophoretic separation on denaturing MOPS formaldehyde agarose gel, RNA was transferred to a Hybond N⁺ nylon membrane (Amersham Hybond, GE Healthcare) and hybridized with dCTP[P32]-labeled DNA probes specific to intergenic regions in the α -globin domain (Table S2). The probes were prepared by Random Primer DNA labeling System (Invitrogen) of PCR amplified DNA fragments of chicken genomic DNA.

Transient transfection and luciferase reporter assays

Luciferase assays were performed as in.³² In brief, the lncRNA- α GT promoter was amplified from chicken genomic DNA with the primers: Forward 5'-CCTGAAGCTT GCACCCTTCT GAGAGCTGC-3' and Reverse 5'-CCTGAAGCTT CCTATTTCAG ATGTGATCAC C-3' to give an amplicon of about 1.1 kb that was cloned into plasmid pGL3-Basic in orientation 5'-3' and 3'-5'. RBCs were transfected at several stages of chicken development (5dRBC-13dRBC) and cell lines: HD3, HD3 induced, chicken fibroblast, DT40, and HeLa were also used in these experiments. Cells were cultured for 48 h and cell extracts were subsequently prepared and assayed for luciferase activity (Promega). Firefly luciferase activity was normalized relative to the activity of Renilla luciferase.

Knockdown assay

siRNA and shRNA sequences were designed with Invitrogen's Block-It RNAi designer (<https://rnaidesigner.invitrogen.com/>

RNAiexpress/). We designed one siRNA for *C16orf35/CGTHBA* gene; siRNA-3: 5'-CAGAACATCA CTACATAAA-3', and two different siRNAs, siRNA-1: 5'- GCAGGTGATA ACAACTGAA-3' and siRNA-2: 5'-CTGGGAGAAA GCTACAACA-3' and a shRNAf: 5'- GATCCGGACA TCTCACAGCT TCATTTCAAG AGAATGAAGC TGTGAGATGT CCTTTTTTGG AAA- 3' and shRNAr: 5'-AGCTTTTCCA AAAAAGGACA TCTCACAGCT TCATTCTCTT GAAATGAAGC TGTGAGATGTCCG-3' against lncRNA- α GT. Additionally, we used a siRNA scramble¹³ and empty vector as a control, respectively. shRNA was cloned into pSilencer 3.1-H1 Puro plasmid (Ambion) by using the BamHI and HindIII restriction sites. All transfection assays were performed by using Lipofectamine 2000 (Invitrogen). shRNA constructs were stably transfected in HD3 cells as described in,¹³ integrants were selected with 1 μ g/ml puromycin (Sigma) for 7–15 d. The levels of lncRNA- α GT and α -globin genes were quantified by conventional PCR and quantitative PCR.

Chromatin immunoprecipitation

The chromatin immunoprecipitation (ChIP) assay was done as previously reported.^{13,15} The chromatin fragments were quantified by qPCR (primers in Table S2) using Platinum Taq DNA polymerase (Invitrogen) and SYBR Green I (Sigma). We performed the experiment in the StepOne detection System (Applied Biosystems). Enrichments were calculated relative to negative region corresponding to the heterochromatic region located upstream of the chicken β -globin locus (heterochromatin primers H1) and values were normalized with IgG measurements.¹⁶ Antibodies used: anti-acetyl-Histone H3 (#06-599, Millipore); anti-acetyl-Histone H4 (#06-866, Millipore); anti-HistoneK4 trimethyl H3 (#07-473, Millipore); anti-MLL1 (#A300-086A, Bethyl); anti-RNAP II (MMS-126R, COVANCE); rabbit IgG (#12-370, Millipore) and mouse IgG (#12-371, Millipore).

Bioinformatic analysis

In order to identify potential promoter regions of lncRNA- α GT we used the Neural Network Promoter Prediction program with default parameters.^{33,34} To identify potential polyadenylation sites we used the PolyAdq program with default parameters and for identification of splicing sites we used the Splice Site Prediction by Neural Network program. For the coding potential analysis of lncRNA- α GT we used the Coding Potential Calculator (CPC) program.¹⁷ Coding regions of *H π* , *H α^D* , *H α^A* , and *TMEM8* were used as controls for genes with validated protein coding capacity and that are in the same cluster as lncRNA- α GT. Also, we used the sequences of *Gomafu* (*Gallus gallus*), *C16orf35/CGTHBA* (*Gallus gallus*), *Xist* (*Homo sapiens*), and *HOTAIR* (*Homo sapiens*) as controls for validated lncRNA genes.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank Paul Delgado-Olguín and Ricardo Saldaña-Meyer for critical reading of the manuscript and suggestions. 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). Ph.D. fellowship from CONACyT and Dirección General de Estudios de Posgrado-Universidad Nacional Autónoma de México (DGEP) (Arriaga-Canon C, Valdes-Quezada C, Arzate-Mejía R and Fonseca-Guzmán Y). Additional support was provided by the PhD Graduate Program, “Doctorado en Ciencias Biomédicas,” to

the Instituto de Fisiología Celular and the Universidad Nacional Autónoma de México. We thank the technical assistance of Fernando Suaste Olmos.

Supplemental Materials

Supplemental materials may be found here: www.landesbioscience.com/journals/epigenetics/article/27030/

References

- Scherrer K. Regulation of gene expression and the transcription factor cycle hypothesis. *Biochimie* 2012; 94:1057-68; PMID:22234303; <http://dx.doi.org/10.1016/j.biochi.2011.12.010>
- Razin SV, Rynditch A, Borunova V, Ioudinkova E, Smalko V, Scherrer K. The 33 kb transcript of the chicken α -globin gene domain is part of the nuclear matrix. *J Cell Biochem* 2004; 92:445-57; PMID:15156557; <http://dx.doi.org/10.1002/jcb.20066>
- Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cell* 2011; 43:904-14; PMID:21925379; <http://dx.doi.org/10.1016/j.molcel.2011.08.018>
- Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 2012; 81:145-66; PMID:22663078; <http://dx.doi.org/10.1146/annurev-biochem-051410-092902>
- Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 2009; 10:155-9; PMID:19188922; <http://dx.doi.org/10.1038/nrg2521>
- Wapinski O, Chang HY. Long noncoding RNAs and human disease. *Trends Cell Biol* 2011; 21:354-61; PMID:21550244; <http://dx.doi.org/10.1016/j.tcb.2011.04.001>
- Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 2012; 22:1775-89; PMID:22955988; <http://dx.doi.org/10.1101/gr.132159.111>
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. *Cell* 2007; 129:823-37; PMID:17512414; <http://dx.doi.org/10.1016/j.cell.2007.05.009>
- Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 2007; 448:553-60; PMID:17603471; <http://dx.doi.org/10.1038/nature06008>
- Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 2009; 458:223-7; PMID:19182780; <http://dx.doi.org/10.1038/nature07672>
- Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. *Nature* 2012; 482:339-46; PMID:22337053; <http://dx.doi.org/10.1038/nature10887>
- 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)
- Furlan-Magaril M, Rebollar E, Guerrero G, Fernández A, Moltó E, González-Buendía E, Cantero M, Montoliu L, Recillas-Targa F. An insulator embedded in the chicken α -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>
- Klochov D, Rincón-Arango H, Ioudinkova ES, Valadez-Graham V, Gavrillov A, Recillas-Targa F, Razin SV. 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>
- Rincón-Arango H, Guerrero G, Valdes-Quezada C, Recillas-Targa F. Chicken α -globin switching depends on autonomous silencing of the embryonic π globin gene by epigenetics mechanisms. *J Cell Biochem* 2009; 108:675-87; PMID:19693775; <http://dx.doi.org/10.1002/jcb.22304>
- Valdes-Quezada C, Arriaga-Canon C, Fonseca-Guzmán Y, Guerrero G, Recillas-Targa F. CTCF demarcates chicken embryonic α -globin gene autonomous silencing and contributes to adult stage-specific gene expression. *Epigenetics* 2013; 8:827-38; PMID:23880533; <http://dx.doi.org/10.4161/epi.25472>
- Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res* 2007; 35(Web Server issue):W345-9; PMID:17631615; <http://dx.doi.org/10.1093/nar/gkm391>
- Gavrillov AA, Razin SV. Spatial configuration of the chicken α -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>
- Core LJ, Waterfall JJ, Lis JT. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 2008; 322:1845-8; PMID:19056941; <http://dx.doi.org/10.1126/science.1162228>
- Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA. Divergent transcription from active promoters. *Science* 2008; 322:1849-51; PMID:19056940; <http://dx.doi.org/10.1126/science.1162253>
- Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, Lassmann T, Forrest AR, Grimmond SM, Schroder K, et al. Tiny RNAs associated with transcription start sites in animals. *Nat Genet* 2009; 41:572-8; PMID:19377478; <http://dx.doi.org/10.1038/ng.312>
- Hung T, Wang Y, Lin MF, Koegel AK, Kotake Y, Grant GD, Horlings HM, Shah N, Umbricht C, Wang P, et al. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet* 2011; 43:621-9; PMID:21642992; <http://dx.doi.org/10.1038/ng.848>
- Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Lapetavic M, Barbara-Haley K, Kuersten S, et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* 2010; 465:182-7; PMID:20393465; <http://dx.doi.org/10.1038/nature09033>
- De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* 2010; 8:e1000384; PMID:20485488; <http://dx.doi.org/10.1371/journal.pbio.1000384>
- Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X, et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 2013; 498:516-20; PMID:23728302; <http://dx.doi.org/10.1038/nature12210>
- Melo CA, Drost J, Wijchers PJ, van de Werken H, de Wit E, Oude Vrielink JA, Elkon R, Melo SA, Léveillé N, Kalluri R, et al. eRNAs are required for p53-dependent enhancer activity and gene transcription. *Mol Cell* 2013; 49:524-35; PMID:23273978; <http://dx.doi.org/10.1016/j.molcel.2012.11.021>
- Martin D, Pantoja C, Fernández Miñán A, Valdes-Quezada C, Moltó E, Matesanz F, Bogdanovic O, de la Calle-Mustienes E, Domínguez O, Taher L, 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>
- Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* 2011; 472:120-4; PMID:21423168; <http://dx.doi.org/10.1038/nature09819>
- Bertani S, Sauer S, Bolotin E, Sauer F. The noncoding RNA *Mistral* activates *Hoxa6* and *Hoxa7* expression and stem cell differentiation by recruiting MLL1 to chromatin. *Mol Cell* 2011; 43:1040-6; PMID:21925392; <http://dx.doi.org/10.1016/j.molcel.2011.08.019>
- Onoguchi M, Hirabayashi Y, Koseki H, Gotoh Y. A noncoding RNA regulates the neurogenin1 gene locus during mouse neocortical development. *Proc Natl Acad Sci U S A* 2012; 109:16939-44; PMID:23027973; <http://dx.doi.org/10.1073/pnas.1202956109>
- Escamilla-Del-Arenal M, Recillas-Targa F. GATA-1 modulates the chromatin structure and activity of the chicken α -globin 3' enhancer. *Mol Cell Biol* 2008; 28:575-86; PMID:17984219; <http://dx.doi.org/10.1128/MCB.00943-07>
- Rincón-Arango H, Valadez-Graham V, Guerrero G, Escamilla-Del-Arenal M, Recillas-Targa F. YY1 and GATA-1 interaction modulate the chicken 3'-side α -globin enhancer activity. *J Mol Biol* 2005; 349:961-75; PMID:15913647; <http://dx.doi.org/10.1016/j.jmb.2005.04.040>
- Knudsen S. Promoter2.0: for the recognition of PolII promoter sequences. *Bioinformatics* 1999; 15:356-61; PMID:10366655; <http://dx.doi.org/10.1093/bioinformatics/15.5.356>
- Reese MG. Application of a time-delay neural network to promoter annotation in the *Drosophila* melanogaster genome. *Comput Chem* 2001; 26:51-6; PMID:11765852; [http://dx.doi.org/10.1016/S0097-8485\(01\)00099-7](http://dx.doi.org/10.1016/S0097-8485(01)00099-7)