UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

EL PAPEL DE CTCF Y EL ARN ANTISENTIDO WRAP53 EN LA REGULACIÓN DEL GEN P53 HUMANO

TESIS

QUE PARA OPTAR POR EL GRADO DE:

DOCTOR EN CIENCIAS

PRESENTA:
M. EN C. RICARDO SALDAÑA MEYER

TUTOR PRINCIPAL

DR. FÉLIX RECILLAS TARGA
INSITUTO DE FISIOLOGÍA CELULAR

MIEMBROS DEL COMITÉ TUTOR

ROBERTO CORIA ORTEGA INSTITUTO DE FISIOLOGÍA CELULAR
FELIPE CRUZ GARCÍA FACULTAD DE QUÍMICA

MÉXICO, D. F. MAYO, 2014

RECONOCIMIENTOS

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

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

Dr. Félix Recillas Targa	Instituto de Fisiología Celular, UNAM
Dr. Roberto Coria Ortega	Instituto de Fisiología Celular, UNAM
Dr. Felipe Cruz García	Facultad de Química, UNAM

Se reconoce la colaboración del Dr. Danny Reinberg y los miembros de su laboratorio del Departamento de Bioquímica de la Universidad de Nueva York, en cuyo laboratorio se llevaron a cabo parte de los experimentos realizados durante el desarrollo de la Tesis.

Se reconoce la asesoría técnica de Georgina Guerrero Avendaño en técnicas de biología molecular.

Se reconoce la asesoría de Leticia García Gutiérrez.

Se reconoce la asesoría de Ana María Escalante.

El proyecto fue apoyado parcialmente por CONACYT (42653-Q, 128464) y DGAPA-UNAM (IN209403 and IN203811). Durante los estudios de doctorado goce de una beca otorgada por CONACYT (213029) para la realización de la presente tesis y una estancia de investigación en la Universidad de Nueva York en el Laboratorio del Dr. Danny Reinberg.

El Jurado de Examen Doctoral estuvo constituido por:

Presidente	Dr. Federico Sánchez Rodríguez	Instituto de Biotecnología, UNAM
Vocal	Dr. Luis Alfonso Vaca Domínguez	Instituto de Fisiología Celular, UNAM
Vocal	Dr. Luis Alonso Herrera Montalvo	Instituto de Investigaciones Biomédicas, UNAM
Vocal	Dr. José Luis Reyes Taboada	Instituto de Biotecnología, UNAM
Secretario	Dr. Emilio Rojas del Castillo	Instituto de Investigaciones Biomédicas, UNAM

INDICE

RESUMEN

En esta tesis doctoral describimos que la disminución de los niveles de CTCF en el núcleo tuvo un efecto directo en la actividad transcripcional del gen *p53* sorprendentemente también en la expresión de su ARN antisentido *Wrap53*. Basándonos en esta observación, abordamos la relación entre CTCF y Wrap53 en función de la regulación del gen *p53* humano. Describimos que CTCF y el ARN antisentido Wrap53 forman contactos directos y que estos son importantes durante la respuesta al daño al ADN. Definimos una nueva función para CTCF, su habilidad intrínseca para unir ARN. Logramos identificar una región dentro de esta proteína que es necesaria y suficiente para unir ARN, además pudimos separar su capacidad de unir ARN con su bien establecida unión al ADN. Finalmente, las herramientas que generamos en el desarrolló de este proyecto nos permitieron identificar una posible función a escala del genoma completo para la interacción de CTCF con ARNs al interior del núcleo. Basándonos en los resultados obtenidos, proponemos que la múltimerización de CTCF es dependiente de ARN y así contribuir a la organización tri-dimensional del genoma.

ABSTRACT

In this thesis we describe that depletion of CTCF levels in the nucleus have a direct effect on the transcriptional activity of the *p53* gene, surprisingly also in the expression of its antisense RNA *Wrap53*. Based on this observation, we addressed the relationship between CTCF and *Wrap53* in the regulation of the human p53 gene. We described that CTCF makes direct contacts with Wrap53 and these are important for the DNA damage response. Importantly, we describe a new role for CTCF, its intrinsic ability to bind RNA. We identified a region within this protein that is necessary and sufficient to bind RNA, we could also separate its ability to bind RNA with its well-established DNA binding activity. Finally, the tools developed in this project allowed us to identify a possible general role for the interaction of CTCF with RNA in the nucleus. Based on these results, we propose that CTCF multimerization is dependent on RNA and contribute to the three-dimensional organization of the genome.

ABREVIATURAS

ENCODE - La Enciclopedia de elementos del ADN

CTCF - CCCTC-binding protein

ChIP - inmunoprecipitación de la cromatina

ChIP-seq - ChIP seguido de secuenciación masiva

DZ – dedos de zinc

FISH - hibridación *in situ* con fluorescencia

3C - captura conformacional de cromosomas

CTCFc - sitios de unión a CTCF constitutivos

Wrap53$_{total}$ – se refiere a todas las isoformas de Wrap53 detectadas por qPCR

RIP - inmunoprecipitaciónes de ARN

qPCR - PCR cuantitativa

PAR-CLIP - *Photoactivatable Ribonucleoside enhanced Crosslinking and Immunoprecipitation*

4-SU - 4-tiouridina

PAR-CLIP seq – PAR-CLIP seguido de secuenciación masiva

lincARNs - transcritos largos no-codificantes intergénicos

RUR - región de unión a ARN de CTCF que incluye los aminoácidos 520-727

CME - células madre embrionarias

iPSCs - células humanas pluripotenciales inducidas

INTRODUCCION

Cada una de nuestras células contiene alrededor de 2 metros de ADN que tienen que compactarse hasta una escala menor para poder ser contenido al interior del núcleo de la célula. Para lograr esto el ADN se envuelve alrededor de nucleosomas para crear lo que conocemos como cromatina. Al momento de escribir esta tesis, habían pasado más de 10 años desde la publicación inicial del primer borrador del genoma humano que fue un parte aguas en la biología molecular y en la genética humana (Lander et al. 2001). Diez años después le toco el turno a la regulación epigenética, lo cual representa otro parte aguas en la evolución de la biología moderna. La Enciclopedia de elementos del ADN (ENCODE) ha publicado 30 artículos de acceso público, todos conectados con el análisis y procesamiento de datos obtenidos por un consorcio internacional donde participaron más de 30 instituciones en diferentes partes del mundo (ENCODE Project Consortium, 2011). El articulo principal cuenta con cerca de 450 autores e incluye la anotación de genes, pseudogenes y ARNs no-codificantes, el análisis del transcriptoma, el análisis de modificaciones post-traduccionales de histonas, la unión de factores de transcripción, además de otros tipos de análisis que incluyen desde el estado de la metilación del ADN hasta el análisis del genoma en tres dimensiones. Esta impresionante empresa científica representa un gran paso para mejorar nuestro entendimiento de la biología humana y servirá como base para poder abordar problemas tan complejos como la biología del cáncer entre otras enfermedades.

En esta tesis doctoral deseamos seguir entendiendo la función de una proteína esencial en la regulación epigenética, la formación de asas de cromatina y de dominios topológicos en el núcleo: la proteína múltifuncional CTCF. Determinaremos la relevancia de CTCF en la regulación del gen supresor de tumores *p53* y de *Wrap53*, un ARN antisentido asociado a la regulación transcripcional del gen *p53* humano.

Antecedentes

El factor nuclear “CCCTC-binding protein” o CTCF es una de las proteínas reguladoras más dinámicas y multifuncionales que se conocen; tiene una distribución ubicua y además esta altamente conservada entre especies (Phillips y Corces 2009). Esta proteína se identificó originalmente a partir de un análisis de factores de transcripción que podían unirse al promotor de *c-myc* de pollo (Lobanenkov et al. 1986). Lo que resultó en su purificación (Lobanenkov et al. 1990), clonación (Klenova et al. 1993) y posteriormente su descripción como un factor de transcripción (Filippova et al. 1996). Desde entonces, se han descrito para CTCF una gran variedad de funciones ligadas a diversos procesos celulares que incluyen su función como factor de transcripción, como *insulator*, su papel en el procesamiento alternativo de mensajeros, en la impronta genómica y en la inactivación del cromosoma X, así como en la organización del genoma al interior del núcleo mediante la formación de asas cromatínicas (Phillips y Corces 2009; Holwerda y de Laat 2013). En un inicio, para identificar sitios de unión a CTCF y estudiar su relevancia era necesario hacer un análisis de hipersensibilidades al corte de la DNasa I. Estas hipersensibilidades hacían evidente a los investigadores las regiones del genoma donde la cromatina estaba en una conformación abierta y esto se correlacionó con transcripción activa (Gross y Garrard, 1988). Una vez que se determinaba una hipersensibilidad se corroboraba la presencia de CTCF mediante geles de retardo e inmunoprecipitación de la cromatina (ChIP) (Bell et al. 1999; Recillas-Targa et al. 2002). Actualmente se ha facilitado inmensamente la identificación de sitios de unión a factores de transcripción, como CTCF, y del estatus de modificaciones de histonas mediante ensayos de ChIP seguidos de secuenciación masiva (ChIP-seq). Esta nueva estrategia experimental permite tener un panorama epigenético global del contexto celular de interés y facilita su estudio (Bhinge et al. 2007).

Estructura y Relevancia Biológica

La proteína CTCF humana esta conformada por 727 aminoácidos que se organizan en un dominio central conformado por 11 dedos de zinc (DZ) enmarcados por segmentos terminales sin estructura secundaria definida, que a la fecha tienen una función desconocida (Martínez y Miranda 2010). Actualmente no ha sido posible cristalizar la proteína para resolver su estructura. No se ha observado que los segmentos del amino y el carboxilo terminal puedan, *in vitro*, interactuar entre ellos (Martínez y Miranda 2010). El dominio de DZ de CTCF comprende diez DZ de tipo C_2H_2 y un onceavo de tipo C_2HC (Figura 1), además este dominio está conservado casi por completo entre especies como el ratón, el pollo y el humano (Ohlsson y Renkawitz, 2001). Asimismo, la unión de CTCF al ADN puede ser mediante diferentes combinaciones de sus 11 DZ y puede verse afectada por la metilación del ADN (Kanduri et al. 2000; Pant et al. 2003).

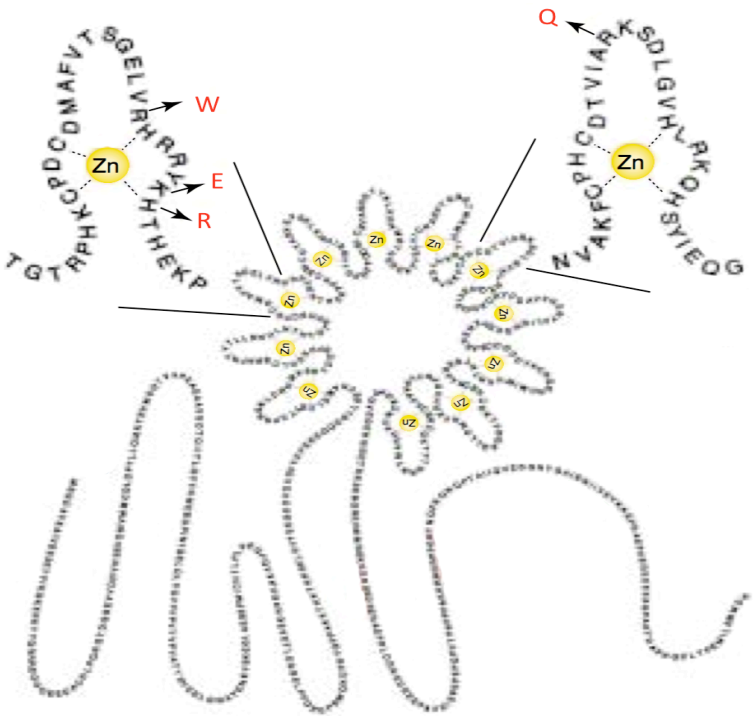


Figura 1. Representación gráfica de la proteína CTCF. Se muestran las substituciones de aminoácido especificas de tumores en rojo (Filippova et al. 2002).

Recientemente en un análisis de todo el genoma, se describieron los patrones de unión de CTCF al ADN después de mutar sistemáticamente cada uno de sus DZ. Con este estudio se pudo concluir que CTCF se une principalmente a su secuencia de reconocimiento al ADN con sus DZ centrales (4-7) y que sus DZ restantes tienen más bien un papel estabilizador (Figura 2A; Nakahashi et al. 2013). Este grupo propone un modelo para la unión de CTCF al ADN comparándola a una silla de montar (Figura 2B).

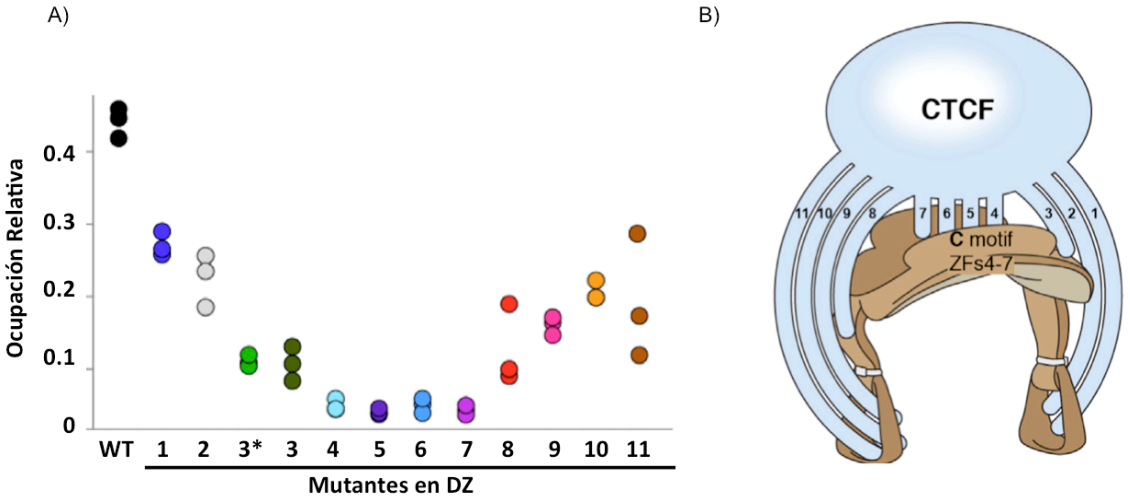


Figura 2. La relevancia de los dedos de zinc de CTCF en la unión al ADN
Se muestran como las mutaciones en cada DZ afectan la unión de CTCF. Graficado como la ocupación relativa (fracción de sitios de unión obtenidos por ChIP-seq) para la proteína silvestre o cada mutante de CTCF. Se muestra el modelo propuesto utilizando como analogía una silla de montar (Tomado de Nakahashi et al. 2013)

Es lógico pensar que debido a la gran cantidad de procesos celulares en los que esta involucrada CTCF debería ser una proteína esencial para la célula. Efectivamente el *knockout* de CTCF en ratón en 2004 confirmó su relevancia para la viabilidad celular, ya que este causa letalidad embrionaria temprana (Fedoriw et al. 2004). Un estudio posterior mas detallado demostró que el ARNm materno está presente en oocitos $CTCF^{-/-}$ hasta la etapa E3.5 y una vez que se pierde la herencia materna el desarrollo se detiene (Moore et al. 2012). A la fecha, no se ha encontrado ninguna enfermedad o síndrome relacionado con la deleción de *CTCF* o alguna mutación que produzca una versión truncada, lo que apoya todavía más su relevancia biológica esencial. Un estudio que analizó más

de 100 muestras de tumores de Wilms, de cáncer de mama y de próstata, solamente pudo encontrar 4 mutaciones que causan substituciones de un aminoácido en el dominio de DZ, haciendo evidente que el papel regulatorio de CTCF se afecta al mutar su dominio de unión al ADN (Figura 1; Filippova et al. 2002). CTCF también está altamente regulada post-traduccionalmente, puede ser fosforilada por la cinasa CK2 preferencialmente en 4 serinas localizadas en la región carboxilo terminal (Klenova et al 2001; El-Kady et al. 2005), poli-(ADP)-ribosilada por PARP-1 en su amino terminal (Yu et al. 2004; Furlan-Magaril et al. 2011) e incluso puede ser SUMOilada por SUMO 1, 2 y 3 en por lo menos dos aminoácidos, las lisinas 74 y 698 (MacPherson et al. 2009). Estas modificaciones post-traduccionales se han visto que pueden afectar su función como *insulator*, su asociación con otras proteínas, su papel como factor de transcripción e incluso se han relacionado con diferentes tipos de cáncer de mama (Docquier et al. 2009).

Proteínas Asociadas

¿Cómo puede una proteína ubicua como CTCF realizar tantas funciones, que a menudo no parecieran tener relación entre si? La respuesta más favorecida hasta el momento es que las funciones de CTCF son dependientes de su asociación con diversos cofactores. A pesar que existe una gran variedad de proteínas reportadas a la fecha que tienen la habilidad de asociarse con CTCF (Zlatanova y Caiafa, 2009), se sabe muy poco sobre los mecanismos que determinan estas asociaciones. Por el momento no existen respuestas claras a esta pregunta, pero existen algunas posibilidades que vale la pena considerar. Por ejemplo, el uso específico de un subconjunto de sus DZ para la unión al ADN podría facilitar que sus segmentos terminales sirvan como plataformas específicas para la interacción con otras proteínas. También una hipótesis muy atractiva es que la elección de cofactores se regule mediante las diferentes modificaciones post-traduccionales de CTCF y/o de su cofactores.

CTCF se ha visto asociada con proteínas de unión al ADN y proteínas asociadas a la cromatina entre muchas otras (Zlatanova y Caifa 2009). Por ejemplo, con

estudios proteómicos se han identificado proteínas que se asocian con CTCF como lamina A/C, importinas, topoisomerasa II, nucleofosmina (Yusufzai et al. 2004). Pero su identificación proteómica no es suficiente para determinar su asociación directa o su relevancia funcional. Por ejemplo, el grupo de investigación de Gary Felsenfeld encontró que el complejo de cohesinas, integrado por Smc1, Smc3, Rad21, SA1 y SA2, co-purifican con CTCF (Xiao et al. 2011). Pero mediante una caracterización bioquímica mas detallada determinaron que de todas las subunidades estudiadas solamente SA1 y SA2 interactúan directamente con el carboxilo terminal de CTCF (Xiao et al. 2011). La asociación del complejo de cohesinas con CTCF ha ganado gran relevancia en la regulación de la estructura del genoma en tres dimensiones en los últimos años, por lo tanto será discutido con mayor detalle en la siguiente sección (Phillips-Cremins et al. 2013).

Distribución de CTCF en el genoma y en tres dimensiones

A la fecha existen un gran número de bases de datos que describe la distribución a escala de genomas completos de CTCF obtenidas por ChIP-seq que incluyen diversos tejidos y líneas celulares en diferentes especies (Martin et al. 2011; Schmidt et al. 2012; Wang et al. 2012; Shen et al. 2012; Holwerda y de Laat, 2013). Estos estudios a gran escala han identificado una gran cantidad de sitios de unión a CTCF con una distribución particular en: regiones promotoras, enhancers, regiones intergénicas, así como a lo largo del cuerpo de los genes (Chen et al. 2012). El análisis de las bases de datos obtenidas por ChIP-seq han proporcionado una idea más clara de la distribución de CTCF en una variedad de contextos celulares y esto nos permitirá validar su posible relevancia en loci específicos y generar nuevas hipótesis (Martin et al. 2011).

A pesar de los avances técnicos seguimos pensando en el genoma de manera lineal. Sin embargo, el contexto dentro de la célula es mucho más complicado y tiene que considerarse en tres dimensiones. El uso de técnicas clásicas como la hibridación *in situ* con fluorescencia (FISH) nos ayudaron a aclarar como se organiza el ADN dentro del núcleo. Se pudo determinar que la cromatina no se

organiza al azar sino de manera tejido-especifico, que forma territorios cromosómicos y que la organización nuclear se puede correlacionar con la actividad transcripcional (revisado en Cremer y Cremer, 2001). Recientemente se ha mostrado gran interés en abordar este problema tan complicado utilizando técnicas basadas en la captura conformacional de cromosomas (3C) que nos permite analizar genes y regiones regulatorias en tres dimensiones (Dekker et al. 2002). Estas técnicas se basan en fijar las interacciones de asas de cromatina que se encuentran próximas entre ellas utilizando formaldehido. Después la cromatina se fragmenta con una enzima de restricción y se ligan las moléculas hibridas de ADN que se forman, se revierte el fijado y finalmente se analizan por PCR. Esta técnica se ha desarrollado continuamente desde el protocolo original de 3C, donde solo se podía analizar una posible interacción a la vez. De hecho, CTCF fue una de las primeras proteínas para la que se demostró su participación en la formación de asas de cromatina entre distintos sitios de unión mediante 3C (Splinter et al. 2006; Handoko et al. 2011). Ahora la herramientas disponibles para la comunidad científica van desde métodos como el 4C, que analiza un locus contra todo el genoma, el 5C que analiza una gran cantidad de loci entre ellos mismos, el ChIA-PET que analiza todas las interacciones que se forman con una proteína de interés, hasta el Hi-C donde se pueden analizar todas las interacciones en el núcleo a la vez (Dekker et al. 2013). Mediante el uso de técnicas como ChIP-seq y variantes del 3C se ha observado que CTCF se une con frecuencia a límites entre regiones cromosómicas con diferentes perfiles de marcas de cromatina y/o diferentes actividades transcripcionales, y en límites entre dominios topológicos (Guelen et al 2008; Dixon et al. 2012). A pesar de los avances técnicos el estudio de CTCF parece complicarse cada vez más ya que es difícil explicar sus funciones que frecuentemente pueden ser contrastantes. Se ha propuesto que las diferentes asas de cromatina dependientes de CTCF pueden crear contactos con regiones regulatorias y genes que podrían causar diferentes efectos sobre la transcripción de los mismos (Martin et al. 2011).

CTCF no esta sólo en lo que parece ser su función más importante, la organización nuclear, sino también se deben tomar en cuenta al complejo de cohesinas y al Mediador (Phillips-Cremins et al. 2013). Que a pesar de tener funciones bien definidas en replicación y transcripción, respectivamente, recientemente se han visto involucradas, junto a CTCF, con el establecimiento de subclases de regiones genómicas con una conformación topológica específica (Figura 3).

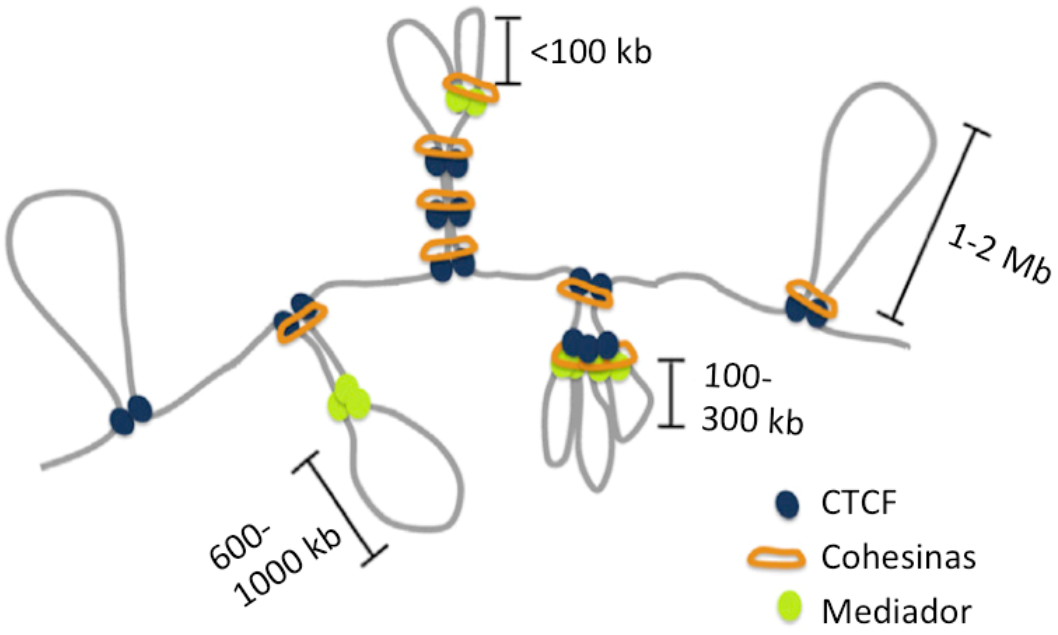


Figura 3. Modelo de subclases topológicas reguladas por proteínas arquitectónicas. La imagen representa gráficamente las diferentes escalas de asas de cromatinas que se relacionaron con la presencia o ausencia de las proteínas arquitectónicas. (Imagen tomada de Phillips-Cremins et al 2013).

Una subclase topológica la determinan los sitios de unión con el complejo de cohesinas y con el Mediador favoreciendo las interacciones más cercanas <100 kb, mientras que Mediador por si mismo, es decir independiente de las cohesinas, favorece interacciones a escalas intermedias que van de 600 a 1000 kb. La subclase de las tres proteínas juntas (CTCF, Mediador y cohesinas) muestra una distribución de asas de escala pequeña a intermedia (<300 kb). Por el contrario, las interacciones mediadas por CTCF y Cohesinas o CTCF solo

fueron significativamente enriquecidas hacia las interacciones a larga distancia (>1 Mb) (Figura 3). Estos resultados apoyan la hipótesis en la cual diferentes subclases de proteínas que contribuyen a la organización del genoma al interior del núcleo funcionan a distintas escalas para cumplir diversas funciones (Phillips-Cremins et al. 2013).

Recientemente un estudio bioinformático clarificó un poco la cantidad de información que esta disponible para CTCF. Determinó que alrededor de 24,000 sitios de unión a CTCF están conservados en 51 o más de las 56 bases de datos analizadas (Li et al. 2013). Este análisis obtuvo resultados muy interesantes e informativos especialmente cuando se compararon con sitios de unión a cohesinas. Por ejemplo, encontraron que en promedio los sitios CTCF constitutivos (CTCFc) están dos veces más propensos a estar conservados entre especies que los sitios CTCF no-constitutivos y constituyen alrededor del 50% de los sitios de unión del total de los sitios para CTCF (Li et al. 2013). La mayoría de los sitios CTCFc que también se comparten con las cohesinas se localizan en regiones intergénicas o intrónicas, lejos de sitios de inicio de transcripción e islas CpG. El hecho que estos sitios se encuentran lejos de islas CpG es consistente con la hipótesis que una gran parte de los sitios CTCF que son tejido específicos pueden ser regulados por metilación del ADN (Chen et al. 2012). En contraste, los sitios CTCFc que no se comparten con cohesinas tienden a estar ubicados cerca de sitios de inicio de la transcripcion y de islas CpG. Juntos estos resultados sugieren que el papel de los sitios constitutivos de CTCF y cohesinas puedan ser más bien estructurales y los demás puedan estar directamente relacionados con regulación transcripcional (Li et al. 2013).
En conclusión, cualquier estudio describiendo las funciones de CTCF en la regulación transcripcional también debe tomar en cuenta su capacidad de formar asas de cromatina y como éstas puedan ser reguladas.

ANTECEDENTES PARTICULARES

CTCF regula la expresión de genes supresores de tumores como *Rb* (De la Rosa-Velázquez et al. 2007), *BRCA1* (Butcher et al. 2007) y *p16* (Witcher et al. 2009) al unirse a sus regiones promotoras. En estos tres casos, se correlacionó la metilación del ADN y el silenciamiento epigenético del gen con la perdida de la unión de CTCF en sus respectivos promotores. En nuestro laboratorio se demostró que este no es el caso para el gen *p53* humano donde la perdida de CTCF se correlacionó con el enriquecimiento de marcas en la cromatina asociadas a represión transcripcional (Soto-Reyes y Recillas-Targa, 2010). A pesar de ser indispensable para evitar la transformación oncogénica, la regulación transcripcional y epigenética del gen *p53* ha sido un tema muy poco estudiado (Saldaña-Meyer y Recillas-Targa, 2011).

Esta tesis se basa principalmente en observaciones de nuestro laboratorio donde se propone un modelo en el cual CTCF contrarresta el silenciamiento epigenético anormal en promotores de genes supresores de tumores (Soto-Reyes y Recillas-Targa, 2010). Igualmente en evidencia reciente donde se reportó la existencia de un transcrito anti-sentido del gen *p53*, denominado *Wrap53*, que regula directa y positivamente la expresión del gen *p53*. Se observó que la sobre-expresión de *Wrap53* incrementa los niveles de transcritos y de la proteína de p53, en contraste su *knockdown* resulta en la disminución de los mismos (Mahmoudi, 2009). Este descubrimiento reveló una vía regulatoria alternativa y/o complementaria para modular la expresión de *p53* y además propone un mecanismo general para la regulación génica mediada por un transcrito antisentido en células humanas. Pensamos que este nuevo aspecto en la regulación de p53 puede ser muy relevante a nivel epigenético. Nuestra intensión es la de profundizar en el entendimiento de los mecanismos epigenéticos a través de la acción del factor de transcripción CTCF y del ARN anti-sentido *Wrap53* (Mahmoudi et al. 2009; Soto-Reyes y Recillas-Targa 2010). Por lo cual decidimos incorporar el estudio de Wrap53 a este proyecto de tesis

doctoral en un esfuerzo para obtener una visión más amplia e integral de la regulación del gen *p53* humano en relación con las funciones de CTCF.

PLANTEAMIENTO DEL PROBLEMA

CTCF contrarresta el silenciamiento epigenético anormal del gen *p53* humano. Por otro lado, su transcrito anti-sentido, *Wrap53*, regula directa y positivamente la expresión del gen *p53.* Sin embargo estas son observaciones independientes y se desconocen los detalles del mecanismo(s) mediante el cual llevan a cabo su función. Tomando en cuenta estos antecedentes buscaremos determinar si existe una inter-dependencia entre CTCF y el ARN antisentido *Wrap53* y, si existe, describir su mecanismo molecular.

HIPÓTESIS

El factor de transcripción CTCF y la presencia del ARN anti-sentido *Wrap53* coordinan la regulación de la expresión del gen *p53* humano a nivel genético y epigenético.

OBJETIVO

Determinar si existe interdependencia entre CTCF y *Wrap53* en la regulación del gen *p53* humano, así como caracterizar los mecanismos moleculares que determinan su expresión en un contexto de células humanas en cultivo.

MATERIALES Y MÉTODOS

Construcciones

ADN que codifica para CTCF o con la mutación Δ576-614 fue clonado en el sistema de pINTA-N3 (Kaneko et al. 2014) donde la expresión del transgen se activa de una manera dependiente de doxiciclina por el *trans*-activador rtTA contenido en el vector pTRIPZ que a su vez también expresa un shARN contra el extremo 3'UTR de CTCF (Thermo Scientific V3THS_409881).
Para producir proteínas truncadas GST-CTCF en *E. coli*, se amplificó por PCR una serie de cADN de CTCF humano y se clonaron en el vector pGEX6 -P1.

Inmunoprecipitación de ARN nativo (RIP)

Los extractos nucleares se obtuvieron usando un protocolo establecido (Dignam et al. 1983) con modificaciones menores para minimizar la actividad de RNasas. Brevemente, las células se lavaron con PBS y después con Buffer A (10 mM Tris pH 7.9_{4^oC} , 1.5 mM $MgCl_2$, 10 mM KCl, inhibidores de proteasas, inhibidores de fosfatasas) y se lisaron en Buffer A con 0.2 % de Igepal CA-630 durante 5 min en hielo. Los núcleos se aislaron por centrifugación a 2500 g durante 5 min y se lisaron en Buffer C (20 mM Tris, pH 7.9_{4^oC}, 25 % glicerol, 400 mM NaCl, 1.5 mM $MgCl_2$, 10 mM EDTA, 0.4 U de inhibidor de RNasas, inhibidores de proteasas, inhibidores de fosfatasas) por 30 min a 4ºC. El lisado fue centrifugado a 20,000 g durante 30 minutos.
Para la inmunoprecipitación, los extractos nucleares se diluyeron a 1 mg/ml en Buffer RIP (20 mM Tris pH 7.9_{4^oC}, 200 mM KCl, 0.05 % Igepal CA-630, EDTA 10 mM) luego centrifugados a 20,000g durante 10 minutos y se incubaron con una cantidad predeterminada de anticuerpo durante 3 hrs a 4ºC. Los inmunocomplejos se recuperaron de la siguiente manera: se añadió 5 µl de Dynabeads acopladas a proteína G (Invitrogen) por µg de anticuerpo utilizado y se incubaron durante 1 hr a 4ºC. Las perlas se lavaron en buffer RIP-W (20 mM Tris pH 7.9_{4^oC}, 200 mM KCl, 0.05% IGEPAL CA- 630, 1mM $MgCl_2$) dos veces. Para eliminar posibles efectos de interacciones de proteína-ADN y ADN-ARN se

incubaron con 2 U de TURBO DNasa (Ambion) en buffer RIP-W durante 10 min a temperatura ambiente. Después de dos lavados adicionales en buffer RIP-W, el ARN se purificó con Trizol (Invitrogen) y el ADN residual se eliminó con un tratamiento adicional de TURBO DNasa.

PAR-CLIP y PAR-CLIP seq

El PAR-CLIP se realizó como se ha descrito anteriormente con algunas modificaciones (Kaneko et al. 2014) Brevemente, las células U2OS se cultivaron bajo condiciones estándar y se incubaron con 400 µM de 4-SU (Sigma) en el medio de cultivo durante 16-24 hrs. Después de lavar las placas con PBS, las células fueron fijadas con 400 mJ/cm2 de UVA (365 nm) usando un UV Stratalinker (Stratagene, CA). Se utilizaron lisados nucleares (LN) obtenidos por fraccionamiento de citoplasma y núcleo por un método estándar (Dignam et al. 1983). Los núcleos se incubaron durante 10 min a 37ºC en un volumen apropiado de buffer de CLIP (20 mM HEPES, pH 7.4, 5 mM EDTA, 150 mM NaCl, 2% lauryldimethylbetaine) suplementado con inhibidores de proteasas , 20 U/µl de Turbo DNasa (Life Technologies) y 200 U/µl de inhibidor de RNasas (New England Biolabs). Las inmunoprecipitaciones se llevaron a cabo utilizando 200 µg de LN, el anticuerpo y las Dynabeads acopladas a proteína G (Life Technologies) se añadieron en el mismo buffer durante toda la noche a 4ºC. Después, cuando fue necesario, se incubaron con varias concentraciones de un coctel de RNasa A + T1 (Ambion) por 5 minutos a 37ºC. La posible contaminación de ADN se eliminó mediante el tratamiento de las perlas con Turbo DNasa (2U en 20 µl). El ARN obtenido se incubó con 5 U de fosfatasa Antártica (New England Biolabs) y luego se marcó radioactivamente con 5U de T4 PNK (New England Biolabs) en presencia de 10Ci [γ - ^{32}P] ATP (PerkinElmer, MA). El material final se resolvió en un gel de 8 % Poliacrilamida-Bis-Tris y luego se transfirió a membranas de nitrocelulosa para finalmente visualizar el ARN marcado mediante autoradiografía.

Para el PAR-CLIP seq se aumentó la cantidad de material utilizado en los experimentos, se utilizó 1 mg de LN. Se ligó un adaptador de ADN al extremo 3' bloqueado (100 pmol/µl) al ARN mediante la incubación de las perlas con T4 ARN ligasa 1 (New England Biolabs) durante 1 hr a 25°C después de la desfosforilación y antes del marcaje terminal. El material se resolvió en un gel de 8 % Poliacrilamida-Bis-Tris y luego se transfirió a membranas de nitrocelulosa para visualizar el ARN marcado mediante autoradiografía. La banda de interés se escindió y el ARN se eluyó de la membrana mediante tratamiento con proteinasa K (Roche, 4 mg/ml) durante 30 minutos a 37°C y después con proteinasa K en presencia de 3.5 M UREA durante 30 minutos a 55°C. Después de la extracción con fenol/cloroformo, se ligaron adaptadores de ARN en el extremo 5', los productos se seleccionaron por tamaño en geles de agarosa y las librerias se amplificaron y secuenciaron en un sistema de secuenciación de Illumina HiSeq 2000.

Para el PAR-CLIP seq en el mapeo inicial se descartaron las secuencias provenientes de los adaptadores y mantuvimos todas las secuencias mayores a 17 nucleótidos. Las secuencias resultantes se analizaron para eliminar las secuencias duplicadas con el kit de herramientas FASTX, luego mapeados con el paquete bowtie-v2-m40-best-strata a la versión hg19 del genoma humano. Las dos replicas biológicas se unificaron y se consideraron como positivos, de manera arbitriaria, los genes con por lo menos 5 secuencias únicas. Los elementos repetitivos que figuran en la base de datos genemasker (la versión descargada la página web de UCSC el 1 de noviembre de 2013) se descartaron dejando 17.201 transcritos como el interactoma de CTCF-ARN. Se realizó un análisis basado en la anotación de la base de datos Ensemble.

Ensayos de unión in vitro

Se sintetizaron sondas de ARN utilizando amplificación por PCR a partir de cADN añadiendo el promotor de la polimerasa de ARN T7. Los ensayos de unión se llevaron a cabo en la presencia de GST-CTCF (100 pmol) y el Wrap53

ARNc en 100 µl de buffer ARN (50 mM de Tris-HCl, pH 7.9_{TA}; 100 mM KCl; 0,1 % de Nonidet P-40; 1,5 mM de $MgSO_4$). La reacción se incubó durante 10 min a 4°C. Perlas de GST se añaden a la reacción y se incubaron durante 30 min a 4°C. Los complejos CTCF-ARN se precipitaron por centrifugación, se lavaron dos veces con buffer de ARN y se dividió por la mitad para el análisis de proteínas y ARN. La proteína se separó por SDS-PAGE y se detectó por tinción con SYPRO Red. El ARN se purificó usando Trizol (Invitrogen), y se separó por UREA 8M-PAGE y detectó por tinción con SYBR Gold.

Cultivo Celular

Células U2OS inducibles que expresan HA-$CTCF_{wt}$ y HA-$CTCF_{\Delta 576\text{-}614}$ se generaron mediante la transfección de pTRIPZ-shCTCF y las construcciones N3-pINTA pertinentes usando lipofectamina 2000 (Invitrogen), seguido de selección con 50 mg/ml de zeocina (Invitrogen). Las células fueron tratadas con doxiciclina durante 72 hrs a una concentración calibrada para permitir la reducción de los niveles de proteínas endógenas y dar lugar a niveles de CTCF y $CTCF_{\Delta 576\text{-}614}$ similares.

Cromatografía de filtración en gel

CTCF completa recombinante o la RUR sola se cargaron en un columna Superose 6 PC 3.2/30 2.4 ml (GE Lifesciences) con o sin la presencia de Wrap53 y se colectaron fracciones cada 100 µl.

Ensayos de cambio de movilidad electroforética (EMSA)

La proteína recombinante se incubó a temperatura ambiente con concentraciones crecientes de los competidores antes de añadir la sonda de ARN. Después, se añadió sonda de ARN y se incubó durante 30 min a TA en 15 µl de buffer que contiene 50 mM de Tris-HCl pH $7.5_{4°C}$,150 mM de NaCl, 5 mM de $MgCl_2$, 0,1 mM de $ZnSO_4$, 1 mM de DTT, 10% de glicerol, 0,1 % de Tween-20, 8 U de inhibidores de RNasas (NEB) y 1 µg de ARN de transferencia de

levadura (ARNt). Las muestras fueron resueltas a 4ºC por 0.5X TGE-4% PAGE y se visualizaron mediante autoradiografía.

Inmunoprecipitación de la cromatina (ChIP)

Las células fueron fijadas con formaldehído al 1% durante 10 min, después se lavó tres veces con PBS y se resuspendieron en ChIP Buffer (50 mM de Tris-HCl, pH 7.9, 150 mM de NaCl, 1 % de Triton X-100, 0.5 % de NP-40, 5 mM de EDTA pH 8.0, 1 mM de PMSF e inhibidores de proteasas).

Las muestras se sometieron a ultrasonido con un Diagenode Bioruptor para generar fragmentos de ADN de ~500 pb pares de bases. Se añadió el anticuerpo (1 µg) a la cromatina fragmentada (200 µg), y se incubó a 4ºC durante toda la noche. Después, se incubaron con Dynabeads acopladas a proteína G (Invitrogen) durante 45 min a 4ºC. Se lavaron dos veces con ChIP Buffer, se trató con RNasa A a 37°C durante 15 min, y se lavó tres veces más con ChIP Buffer. Para revertir la fijación las muestras se calentaron a 65°C durante ~16 horas y se incubaron con Proteinasa K durante 15 min a 55°C. Finalmente, el ADN fue extraído mediante fenol-cloroformo y se precipitó con etanol.

RESULTADOS

CTCF participa en la regulación del locus p53/Wrap53

Estudios previos de nuestro laboratorio caracterizaron un sitio de unión para CTCF en el promotor del gen *p53* humano que contribuye a su expresión transcripcional y tiene la capacidad de mantener este elemento regulador en una configuración local de cromatina abierta (Soto-Reyes y Recillas-Targa, 2010). Por otro lado, un transcrito antisentido natural de p53, llamado *Wrap53*, fue descrito por otro grupo donde observaron que regulaba positivamente los niveles de ARNm endógenos de *p53* (Mahmoudi et al. 2009). Dado que el sitio de unión a CTCF en el promotor de p53 corresponde con el primer intrón del gen *Wrap53* en la cadena opuesta (Figura 4A), hipotetizamos que CTCF regula tanto a *p53* como a *Wrap53* a través de la unión a su promotor y al mismo tiempo al transcrito antisentido.

Para probar esta hipótesis, realizamos un *knockdown* estable de CTCF en células U2OS. Observamos que en el *knockdown* de CTCF (Figura 4B), los niveles de ARNm tanto de *p53* como de todas la isoformas de *Wrap53 (Wrap53$_{total}$)* se redujeron al compararlos con el gen constitutivo *Gapdh* (Figura 4C). Lo mismo también se observó con un *knockdown* inducible después de 72 hrs de inducción con doxiciclina. Los niveles de expresión de p53 también disminuyeron en un *knockdown* de Wrap53$_{total}$ utilizando siRNAs (Figura 4C-D) (Mahmoudi et al. 2009) y el *knockdown* de ambos, CTCF y Wrap53$_{total}$, tuvo efectos similares (Figure 4C). Estos resultados siguieren que CTCF regula la transcripción de *p53* y *Wrap53* y confirma que el ARN de Wrap53 regula la transcripción de *p53*.

CTCF contacta *in vivo* una gran variedad de transcritos incluyendo al transcrito antisentido Wrap53

El *knockdown* de CTCF o de Wrap53 da lugar a efectos similares en los niveles de ARNm de *p53* y dado que el *knockdown* doble no genera un efecto mayor, esto sugiere que se encuentran en la misma via regulatoria. Esto nos llevó a formular la hipótesis que *Wrap53* y CTCF pueden estar relacionados no sólo

funcionalmente, sino también físicamente. Para determinar si el ARN de *Wrap53* se asocia con CTCF *in vivo*, realizamos inmunoprecipitaciónes de ARN (RIP) seguidas por PCR cuantitativa (qPCR).

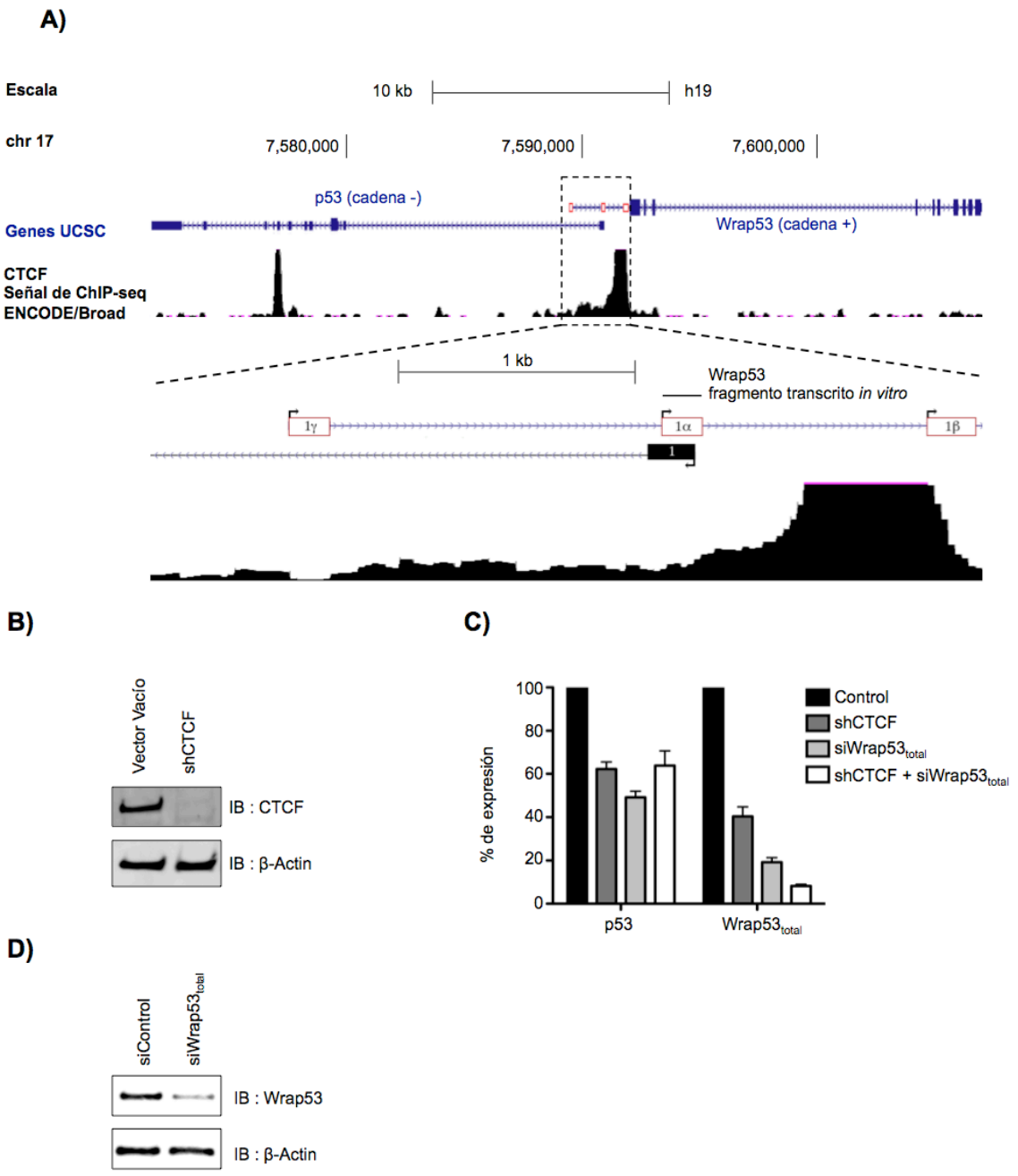


Figura 4. CTCF participa en la regulación del locus p53/Wrap53.
(A) Organización genómica del locus p53/Wrap53. (B) Inmunoblot en células U2OS que expresan un shRNA contra CTCF o como control el mismo vector vacío. β-actina se muestra como control de carga (C) Los niveles de ARN se cuantificaron por RT-qPCR y fueron normalizados contra los niveles de *Gapdh*. Las barras indican la media de 3 repeticiones biológicas y el error estándar. (D) Inmunoblot igual que en (B) pero usando un siARN contra *Wrap53*$_{total}$ o un siARN control.

De acuerdo con nuestra hipótesis, tanto el transcrito de *p53* y como el de *Wrap53*$_{total}$ se ven enriquecidos en RIP usando un anticuerpo contra CTCF cuando se compararon con la RIP usando solamente IgG (Figura 5A). Por otra parte, el ARN de *Wrap53*$_{total}$ se ve enriquecido específicamente con CTCF, cuando se comparó con un ensayo de RIP usando un anticuerpo contra SCML2 (Figura 5A-B), una proteína no relacionada que también une ARN en el núcleo (observaciones no publicadas). Aunque observamos enriquecimiento especifico para el ARN de *Wrap53*$_{total}$ usando SCML2 como control, las RIPs frecuentemente pueden producir falsos positivos debido a la re-asociación de ARN-proteína después de la lisis celular (Mili y Steitz 2004; Riley et al. 2012).

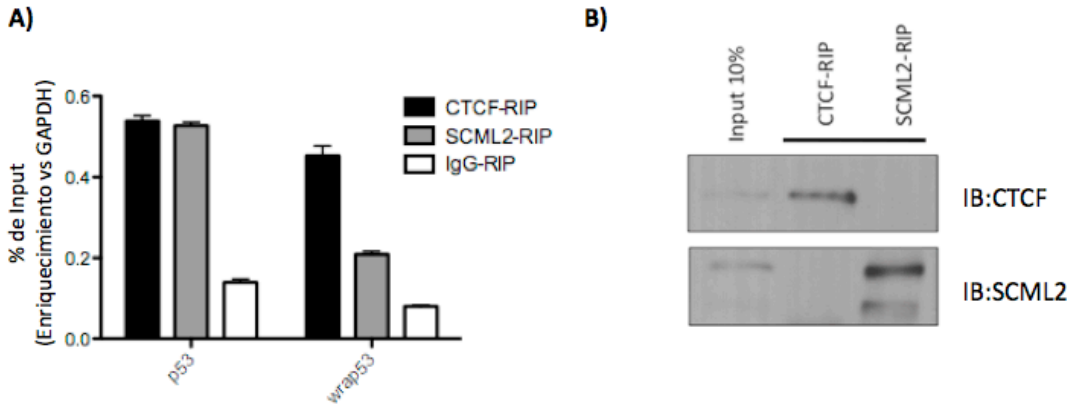


Figura 5. CTCF co-precipita con el transcrito anti-sentido *Wrap53*
(A) RIP utilizando anticuerpos contra CTCF o SCML2. RT-qPCR a partir de ARN recuperado se muestra como % de enriquecimiento normalizado a los niveles de GAPDH. Las barras indican la media de 3 repeticiones biológicas y error estándar (B) Inmunonlot para CTCF y SCML2. (C) Membrana de PAR-CLIP con cantidades crecientes de RNasa A después de 24 horas de exposición. (D) Inmunoblot de la misma membrana que en (C) utilizando anticuerpos contra el CTCF.

Tomando en cuenta las limitaciones de esta técnica, reevaluamos la unión directa y específica de *Wrap53* a CTCF utilizando una estrategia imparcial. Esto implicó estandarizar el ensayo PAR-CLIP, por sus siglas en inglés: *Photoactivatable Ribonucleoside enhanced Crosslinking and Immunoprecipitation*, ya que esta técnica se desarrolló para analizar proteínas del spliceosoma que unen la mayoría del ARN celular. Esta técnica hace uso de un análogo de un nucleósido foto-activable, llamado 4-tiouridina (4-SU), que de

forma selectiva fija irreversiblemente proteínas con ARN en contacto directo cuando se exponen a luz ultravioleta a 365 nm (Hafner et al. 2008).

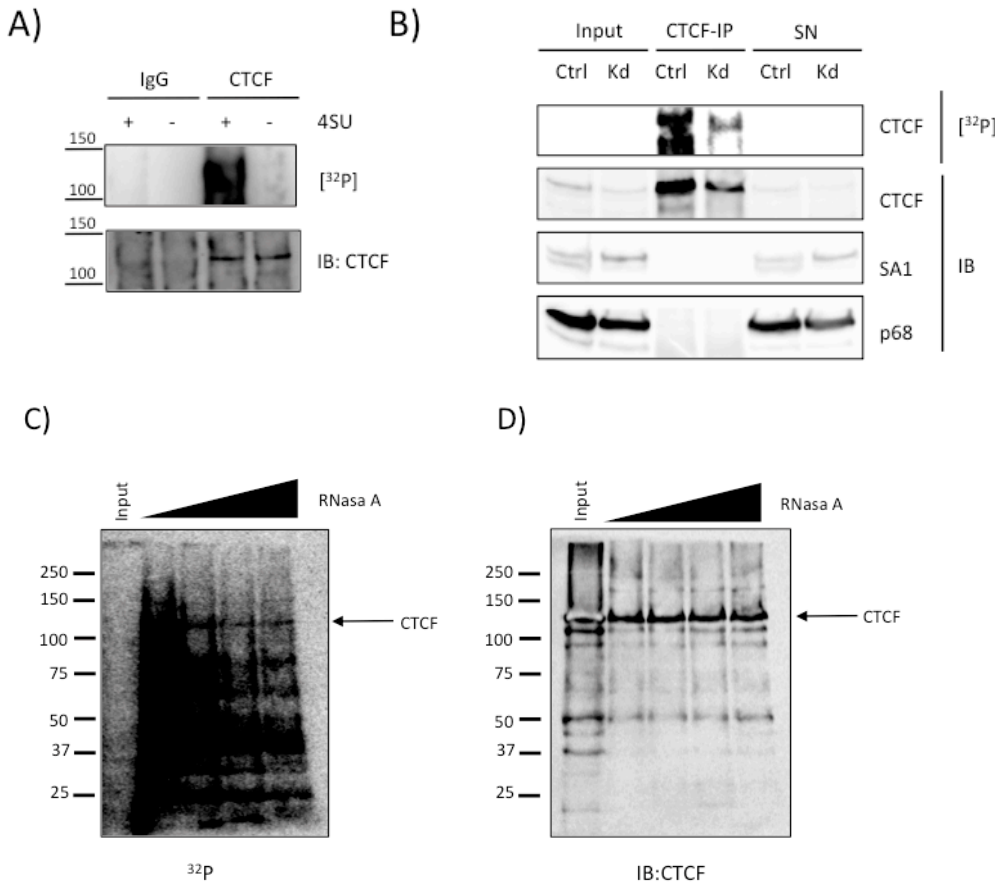


Figura 6. CTCF contacta ARN *in vivo*
(A) PAR-CLIP (arriba) e inmunoblot para CTCF (abajo) en células U2OS con (+) o sin (-) incorporación de 4-SU. Extractos nucleares enteros se inmunoprecipitaron con un anticuerpo contra CTCF o IgG. (B) PAR -CLIP (arriba) e inmunoblots para CTCF, DDX5 o SA1 utilizando la misma membrana. La inmunoprecipitación se realizó utilizando un anticuerpo contra CTCF en células transfectadas de manera estable con un vector vacío (Ctrl) o con un shARN contra CTCF (Kd). (C) Membrana de PAR-CLIP con cantidades crecientes de RNasa A después de 24 horas de exposición. (D) Inmunoblot de la misma membrana que en (C) utilizando anticuerpos contra el CTCF.

Después de inmunoprecipitar el ARN con un anticuerpo contra CTCF y marcar radioactivamente el ARN, pudimos concluir que la señal radiactiva obtenida por autoradiografía era debida al ARN específicamente fijado a CTCF. La señal era dependiente de la presencia de 4-SU en el medio de cultivo (Figura 6A), disminuyó con el *knockdown* de CTCF (Figura 6B) y era sensible al tratamiento con RNasa A (Figura 6C-D). Para evitar la posible contaminación de otras

proteínas de unión a ARN con un peso molecular similar a CTCF, las RIPs se realizaron con un buffer que contiene 2% lauryldimethylbetaine, un detergente bipolar que conserva la reactividad del anticuerpo mientras que disminuye significativamente la co-precipitación de proteínas asociadas a CTCF, por ejemplo la helicasa de ARN p68 (DDX5) o la subunidad SA1 del complejo de cohesinas (Yao et al. 2010; Xiao et al. 2011) (Figura 6B). Es importante mencionar, que la preparación del material para PAR-CLIP conlleva la degradación parcial de CTCF dada la presencia de especies de peso molecular más bajo que se detectan usando un anticuerpo contra CTCF, y disminuyen después del *knockdown* de CTCF (Figura 6 B y D).

Para identificar el conjunto de ARNs unidos a CTCF *in vivo*, cortamos la banda del peso molecular de CTCF de la membrana del PAR-CLIP. Después ese ARN se preparó para la construcción de librerías para su posterior secuenciación masiva. Se obtuvieron aproximadamente 1,200,000 secuencias únicas para cada réplica biológica. Aunque *p53* se expresa en niveles bajos y $Wrap53_{total}$ es mucho menos abundante comparado con el ARNm de *p53*, logramos obtener secuencias para las isoformas, α y β, que se expresan en las células U2OS (Figura 7A). Este resultado confirma nuestras observaciones iniciales de la interacción CTCF con el ARN de *Wrap53* (Figura 5A-B). Cuando comparamos el enriquecimiento relativo promedio de *p53* y $Wrap53_{total}$ obtenido por qPCR, con la cantidad de secuencias únicas obtenidas mediante PAR-CLIP, se observó un claro enriquecimiento para $Wrap53_{total}$ sobre *p53* y *Gapdh* (Figura 7B).
Nuestros datos no muestran enriquecimiento para ARN ribosomal, las especies de ARN más abundantes en la célula, pero en su lugar vemos una preferencia por transcritos originados a partir de la ARN polimerasa II. A continuación, se inspeccionó la distribución genómica de las secuencias obtenidas por PAR-CLIP y encontramos que alrededor del 34 % localizan en secuencias repetitivas, el 30% en los exones, 28% en los intrones y el 8% en regiones intergénicas (Figura 7C).

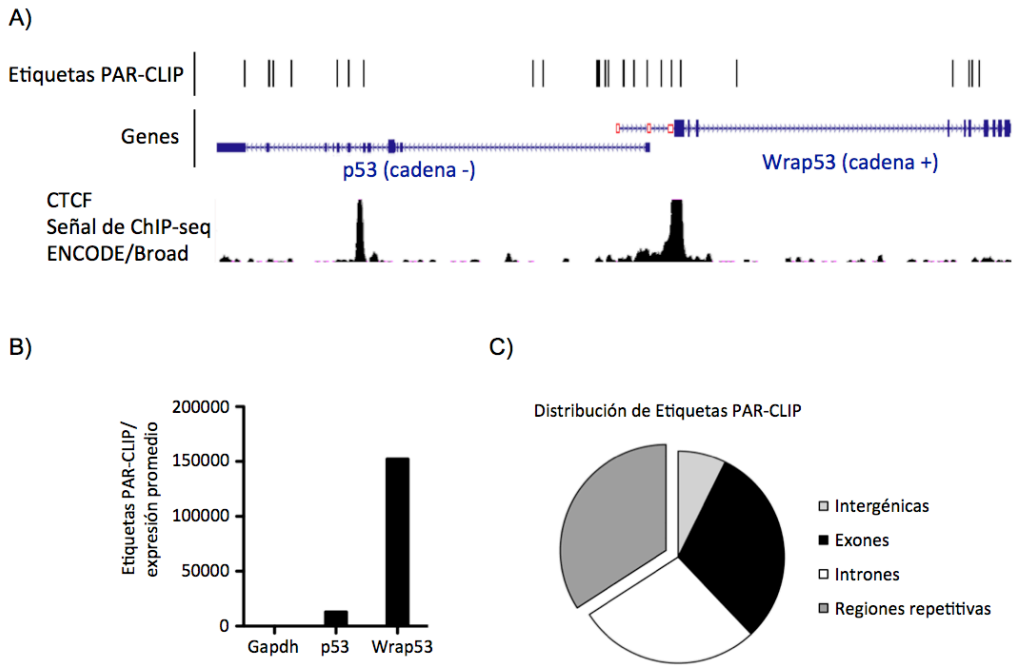


Figura 7. CTCF contacta *in vivo* una gran variedad de transcritos incluyendo al transcrito anti-sentido *Wrap53*
(A) Diagrama que muestra las etiquetas obtenidas por PAR-CLIP después de secuenciación masiva en los loci p53/Wrap53. También se muestra la señal de ChIP contra CTCF del proyecto ENCODE. (B) Representación gráfica de la normalización de las etiquetas obtenidas por PAR-CLIP contra el enriquecimiento relativo promedio obtenido por qPCR. (C) Gráfica circular que muestra la distribución de las etiquetas obtenidas por PAR-CLIP en todo el genoma.

No observamos enriquecimiento preferencial para secuencias provenientes de exones o de intrones, lo que indica que CTCF se une tanto a transcritos nacientes como maduros. Utilizando un corte arbitrario de 5 secuencias únicas, nuestra base de datos esta compuesta de transcritos provenientes de 17,201 genes de los cuales 12,220 son genes que codifican para proteínas, 2,056 son pseudogenes, 1,423 son transcritos antisentido, 866 son transcritos largos no-codificantes intergénicos (lincARNs) y 636 representan otros tipos de ARNs (GEO SuperSeries: GSE53554). Podemos concluir que el ARN de *Wrap53*$_{total}$ se puede recuperar tanto en ensayos de RIP en condiciones nativas como por PAR-CLIP utilizando un anticuerpo contra CTCF. Nuestra base de datos obtenida por PAR-CLIP es la primera evidencia a gran escala que muestra que CTCF une directamente una gran variedad de transcritos en todo el genoma. (transcritos mas representados? Y wrap en relación)

Identificación y mapeo de la región de unión al ARN de CTCF

Teniendo en cuenta los datos obtenidos *in vivo*, nuestro siguiente paso fue tratar de identificar el dominio de CTCF requerido para dicha interacción con el ARN. Para este objetivo expresamos y purificamos en la bacteria *E. coli* la proteína de CTCF completa o una serie de mutantes fusionadas con el dominio humano de la glutatión S-transferasa (GST; Figura 8A y C). Después se incubaron con ARN que se obtuvo por transcripción *in vitro* a partir de ADNc que abarca los nucleótidos 1-167 del exón 1α de *Wrap53* (Figura 4A). Observamos que la unión al ARN fue evidente en el dominio de DZ, pero no en los segmentos terminales (Figura 8A-B). Análisis más detallados demostraron que los DZ 9-11 muestran una mayor afinidad por el ARN comparados con los DZ restantes (Figura 8C-D). La eliminación sistemática de los DZ 9 al 11 reveló que la región óptima para la unión de ARN abarca los aminoácidos 520-727 que incluyen los DZ 10-11 y el carboxilo terminal. Finalmente, el carboxilo terminal por sí solo (residuos 576 a 727) fue incapaz de unir ARN, la mutante que comprende a los residuos 520-615, que carece de la mayor parte del carboxilo terminal pero que contiene los ZF 10-11, también fue incapaz de unir ARN (Figura 8C-D). Estos resultados en conjunto sugieren que, aunque no es suficiente, se requiere el carboxilo terminal para la unión eficaz de ARN. De ahora en adelante nos referimos a esta región optima (residuos 520-727) como la región de unión a ARN (RUR).

Independientemente de los experimentos anteriores decidimos diseñar una proteína mutante con una deleción interna que pudiera abolir la unión al ARN para poder usar como control. Para esto utilizamos el software RNABindR que está diseñado específicamente para predecir regiones putativas de unión de ARN (Terribilini et al. 2007). El software predijo para CTCF con alta sensibilidad y especificidad que la región del carboxilo terminal justo río arriba del último DZ podría unir ARN. Basándonos en esta predicción removimos los residuos 576 a

614 de la RUR, esto causó que la unión al ARN se redujera drásticamente como se puede observar mediante ensayos de unión al ARN (Figura 8E).

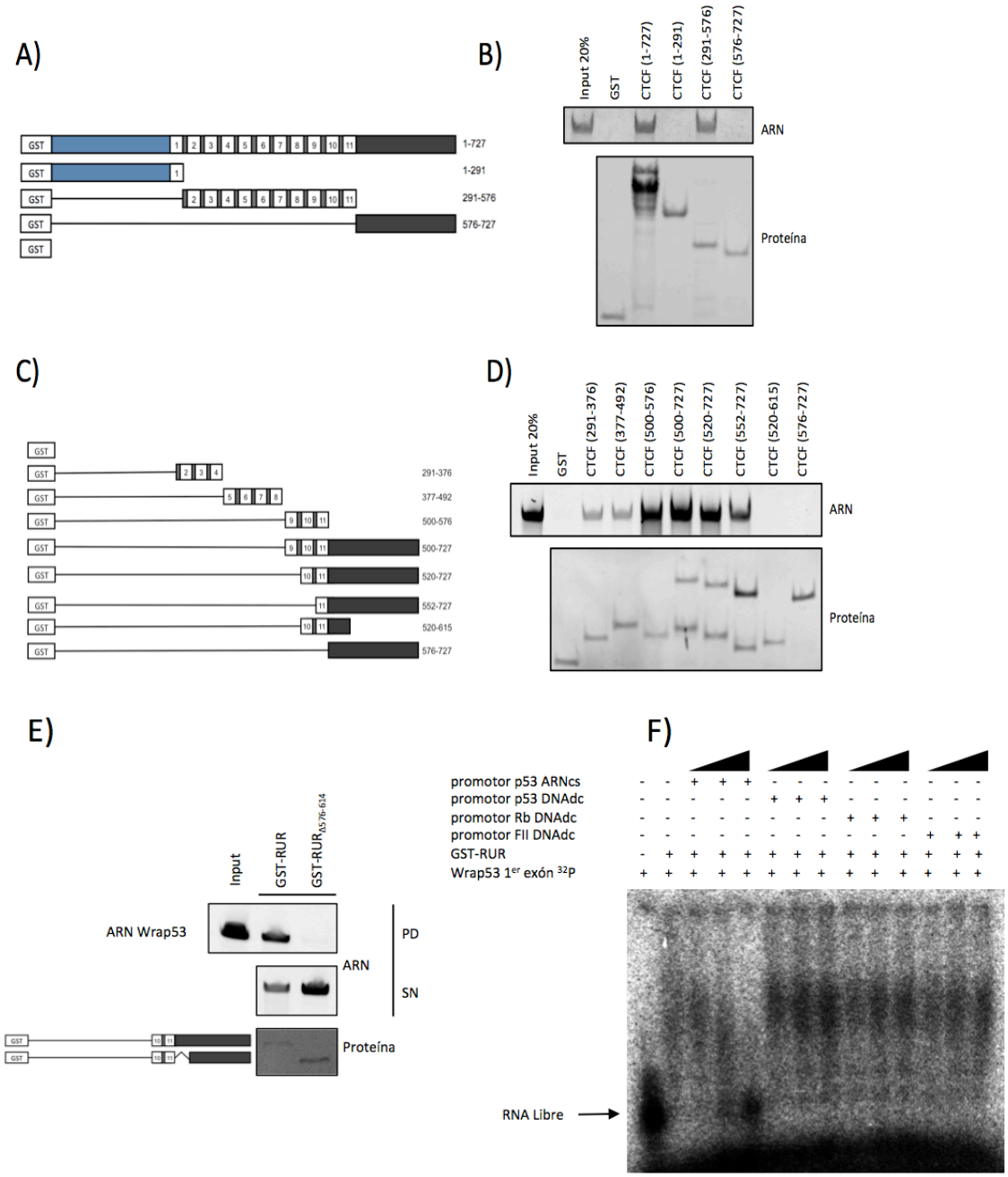


Figura 8. Identificación y caracterización de la región de unión a ARN de CTCF.
(A y C) Representaciones esquemáticas de las construcciones fusionadas a GST, con CTCF completo o sus mutantes (B , D y E) Ensayos de unión a ARN, se muestra el ARN recuperado teñido con SYBR-Gold (panel superior) y las proteínas recuperadas se tiñeron con SYPRO Red (panel inferior). PD= precipitación; SN= sobrenadante. (F) EMSA utilizando GST-RBR y una

sonda radiactiva de los nucleótidos 1-167 de *Wrap53* ARNm, se incubaron con cantidades crecientes de ADN o ARN no marcado.

A continuación probamos la capacidad de la RUR para diferenciar entre ADN y ARN utilizando ensayos de cambio de movilidad electroforética (EMSA). La unión de GST-RUR a una sonda que contiene el primer exón de Wrap53 se comparó en la presencia de cantidades crecientes de ARN o ADN (Figura 8F). Cada una de las sondas de ADN que compiten incluyen un sitio de unión a CTCF que se ha demostrado previamente mediante EMSA. El sitio de unión de CTCF situado en el promotor del gen *p53* (Soto-Reyes y Recillas-Targa 2010), el sitio de unión de CTCF en el promotor del gen *Rb* humano (De la Rosa-Velázquez et al. 2007) y el *insulator* FII del gen *β-globina* de pollo (Valadez-Graham et al. 2004). La sonda de ARN fue obtenida por transcripción *in vitro* a partir de la región promotora del gen *p53* descrita anteriormente.

Observamos que el ARN derivado de la región promotora del gen *p53* compitió con éxito la sonda que contiene el exón Wrap53 (Figura 8F). Por otro lado, ninguno de los competidores de ADN que contienen un sitio de unión a CTCF fueron capaces de competir con la sonda radioactiva de ARN (Figura 8F). Estos datos sugieren que CTCF contiene un región de unión a ARN, que no se había reportado previamente y que además muestra preferencia de unión para el ARN de p53 y *Wrap53 in vitro*.

La fosforilación de CTCF por CK2 interfiere in vitro con la unión al ARN

Vale la pena destacar que la RUR descrita anteriormente y la deleción (Δ576-614), contienen un motivo de 4 serinas altamente conservado que puede ser fosforilado preferencialmente por la cinasa CK2 tanto *in vivo* como *in vitro* (Klenova et al. 2001) y posiblemente por otras cinasas. Además, la mutación de estos residuos de serina no afecta la localización nuclear de CTCF o su unión al promotor de *c-myc* (Klenova et al. 2001). Para determinar si la fosforilación de estos residuos de CTCF afecta su unión al ARN o al ADN, incubamos *in vitro* a la GS-$CTCF_{wt}$ o a la proteína mutante GST-RUR con CK2 recombinante. La

GST-CTCF$_{wt}$ tuvo significativamente menor afinidad por el ARN después de ser fosforilada (Figura 9A). Del mismo modo, la GST-RUR exhibió una reducción drástica en la unión a la sonda de ARN (Figura 9B). Este efecto inhibidor era dependiente de las condiciones adecuadas para la actividad cinasa ya que es dependiente de ATP (Figura 9B). En ninguno de los casos la RUR o la RUR$_{\Delta 576-614}$ pudieron unir una sonda de ADN del promotor del gen *p53* (Figura 9B).

Por lo tanto, estos resultados demuestran que la unión al ARN de CTCF completo y de la RUR por si misma esta sujeta *in vitro* a regulación a través de la fosforilación por CK2.

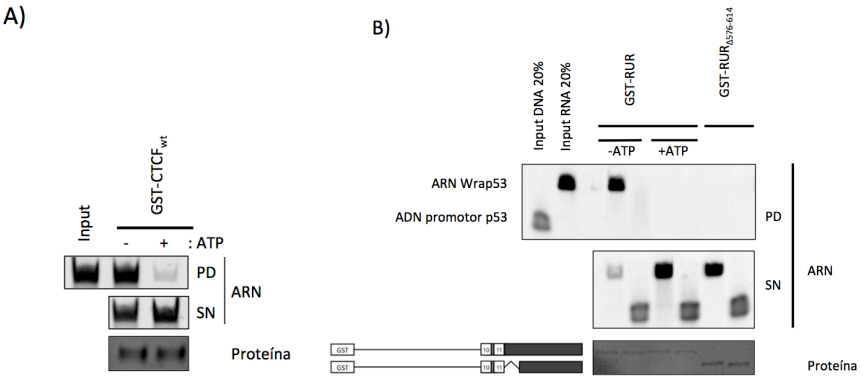


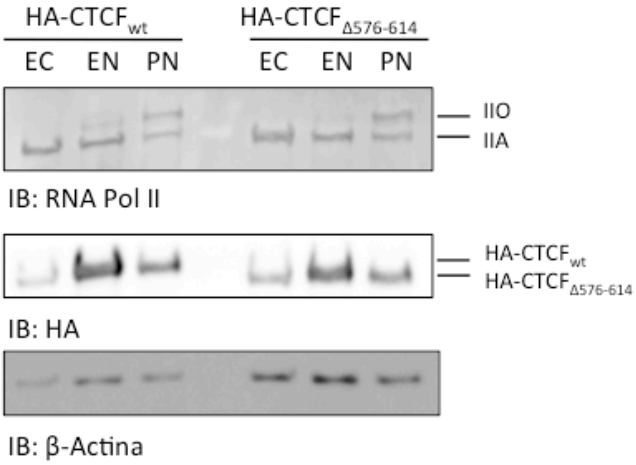
Figura 9. La fosforilación por CK2 interfiere *in vitro* con la unión al ARN.
Ensayos de unión al ARN utilizando (A) CTCF completo o (B) la GST-RUR. Se muestra el ARN recuperado teñido con SYBR-Gold (panel superior y central) y la proteína teñida con SYPRO Red (panel inferior). PD = precipitación; SN = Sobrenadante.

Una deleción interna en el RUR perturba la respuesta el daño al ADN.

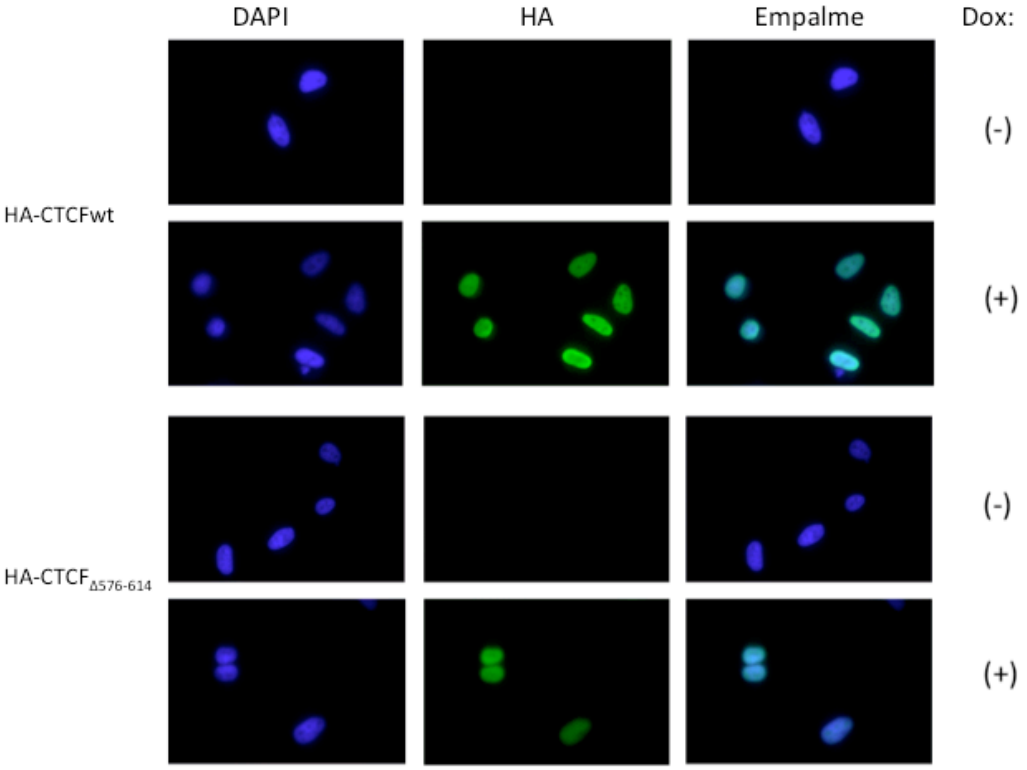
Para abordar la relevancia funcional de las interacciones Wrap53-CTCF *in vivo*, generamos un sistema celular inducible. Este consta de un shRNA, diseñado en el extremo 3'UTR, que disminuye los niveles de CTCF endógeno mientras al mismo tiempo se rescatan los niveles de proteína con versiones recombinantes de CTCF etiquetadas con el péptido pequeño HA, ya sea silvestre (HA-CTCF$_{wt}$) o mutante en la unión al ARN (HA-CTCF$_{\Delta 576-614}$) (Figura 10). Es necesario destacar que existe una señal de localización nuclear putativa dentro de la región mutada (576-614), específicamente en los aminoácidos 590-603 (Klenova

et al. 2001). No obstante, la mutante HA-CTCF$_{\Delta 576\text{-}614}$ tiene la misma capacidad que la HA-CTCF$_{wt}$ para localizarse *in vivo* en el núcleo, para asociarse con la cromatina o con las proteínas DDX5 o la subunidad SA1 del complejo de cohesinas (Yao et al. 2010; Xiao et al. 2011) (Figura 10A-C). Además, las proteínas recombinantes de ambas HA-CTCF$_{wt}$ y HA-CTCF$_{\Delta 576\text{-}614}$ fueron capaces de desalojar al CTCF endógeno residual de la fracción de la cromatina después de inducir su expresión durante 72 horas con doxiciclina (Figura 10D).

A)


HA-CTCF$_{wt}$
HA-CTCF$_{\Delta 576-614}$
EC EN PN
EC EN PN
IIO
IIA
IB: RNA Pol II
HA-CTCF$_{wt}$
HA-CTCF$_{\Delta 576-614}$
IB: HA
IB: β-Actina

B)


DAPI
HA
Empalme
Dox:
HA-CTCFwt
(-)
(+)
HA-CTCF$_{\Delta 576-614}$
(-)
(+)

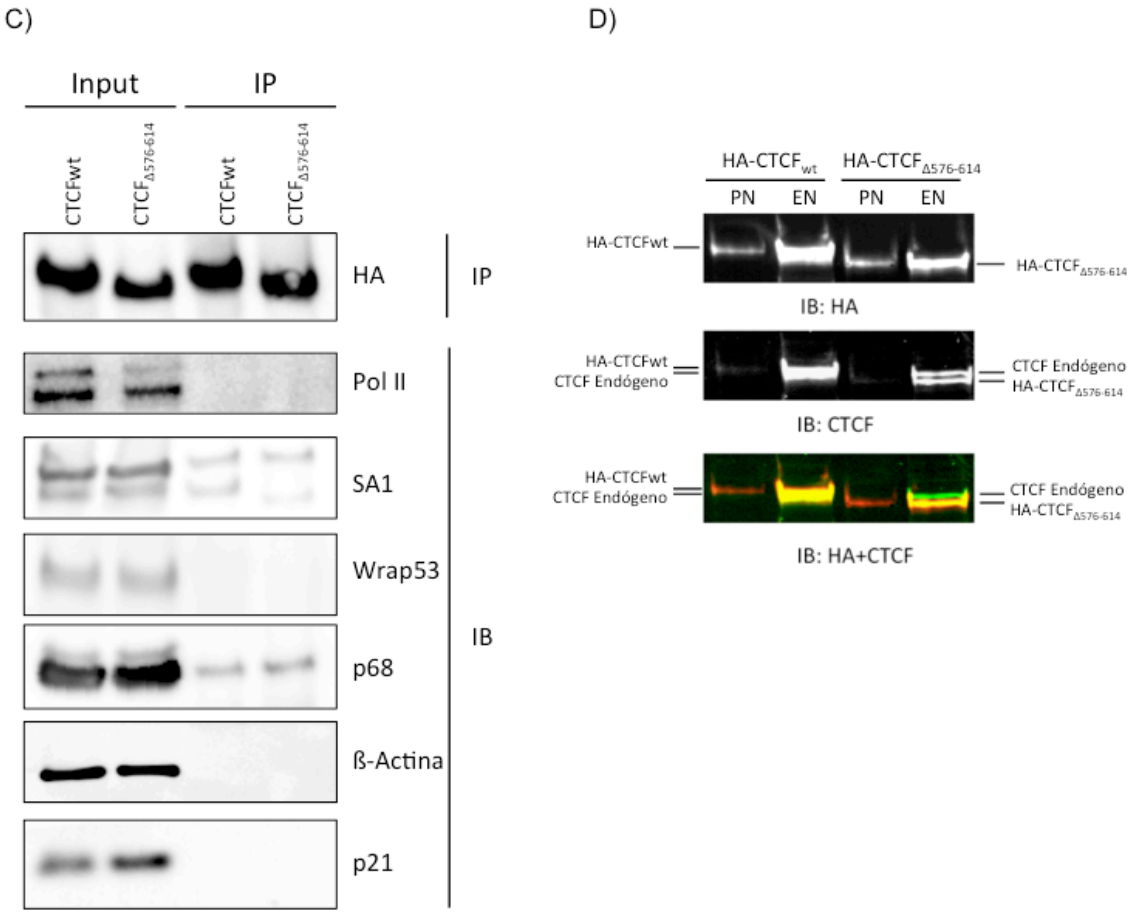


Figura 10. Localización subnuclear de las lineas celulares HA-CTCFwt o HA-CTCF$_{\Delta 576-614}$
Todos los experimentos se realizaron con líneas de células U2OS que expresan HA-CTCFwt o HA-CTCFΔ576-614 tras el tratamiento con doxiciclina durante 72 h. (A) Inmunoblots usando anticuerpos contra HA. La distribución de RNAPIIA y RNAPIIO se muestran como controles de fraccionamiento y β-actina como control de carga. (B) Inmunofluorescencia utilizando un anticuerpo contra HA con o sin tratamiento doxiciclina. DAPI se utilizó para teñir el núcleo. (C) Inmunoblot para las proteínas indicadas después de la IP con anticuerpo contra HA. (D) Inmunoblot usando anticuerpos contra HA y CTCF. CE = extracto citoplasmático; NE = extracto nuclear; NP = pellet nuclear.

Mahmoudi y colaboradores reportaron que los niveles de los transcritos de *Wrap53* y *p53* aumentan en respuesta al daño al ADN. También sugirieron que *Wrap53* no sólo mantiene los niveles de mARN de *p53* basales sino que también desempeña un papel en la estabilización de mARN de *p53* durante la respuesta al daño del ADN (Mahmoudi et al. 2009). Para determinar si un complejo CTCF-*Wrap53* regula la expresión de *p53*, las líneas celulares que expresan HA-CTCF$_{wt}$ o HA-CTCF$_{\Delta 576-614}$ fueron inducidas con doxiciclina durante 56 horas y después se trataron durante 16 horas con doxiciclina junto con metanosulfonato

de metilo (MMS), una droga que daña el ADN. De acuerdo con el reporte anterior (Mahmoudi et al. 2009), después de inducir daño al ADN las células rescatadas con HA-CTCF$_{wt}$ mostraron niveles mas altos de ARNm de *p53* y de las isoformas tanto α como β de *Wrap53* (Figura 11A), así como de sus proteínas (Figuras 11B). Por otra parte, los genes implicados en el arresto del ciclo celular y la apoptosis a través de la vía de *p53*, *p21* y *PUMA* respectivamente, también mostraron niveles más altos de sus transcritos (Figura 11A).

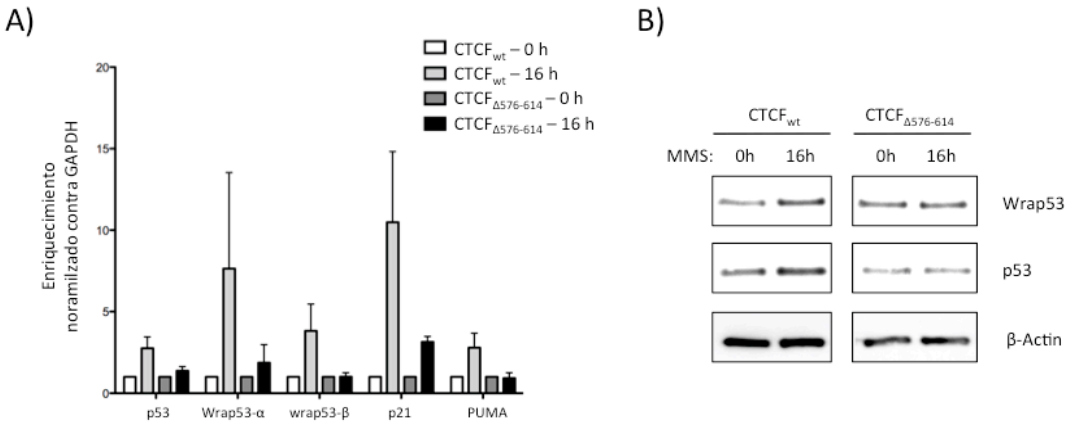


Figura 11. La baja en el RBR dentro CTCF perturba la respuesta al daño del ADN.
Las líneas celulares que contienen HA-CTCFwt o HA-CTCF$_{\Delta 576-614}$ fueron inducidas con doxiciclina durante 56 h y después se trataron durante 16 h con o sin MMS. (A) RT-qPCR se muestra como enriquecimiento normalizado a los niveles de GAPDH. Las barras indican la media de 3 repeticiones biológicas y el error estándar (B) Inmunoblot de p53 y Wrap53, con β-actina como control de carga.

En contraste con este resultado, después del daño al ADN las células con la expresión de la mutante con un defecto en la unión a ARN, HA-CTCF$_{\Delta 576-614}$, no pudo rescatar ni los niveles de ARNm ni los de proteína (Figura 11). Cabe destacar, que el *knockdown* de *Wrap53* no afectó el reclutamiento de CTCF endógeno al promotor del gen *p53* (Figura 12A).

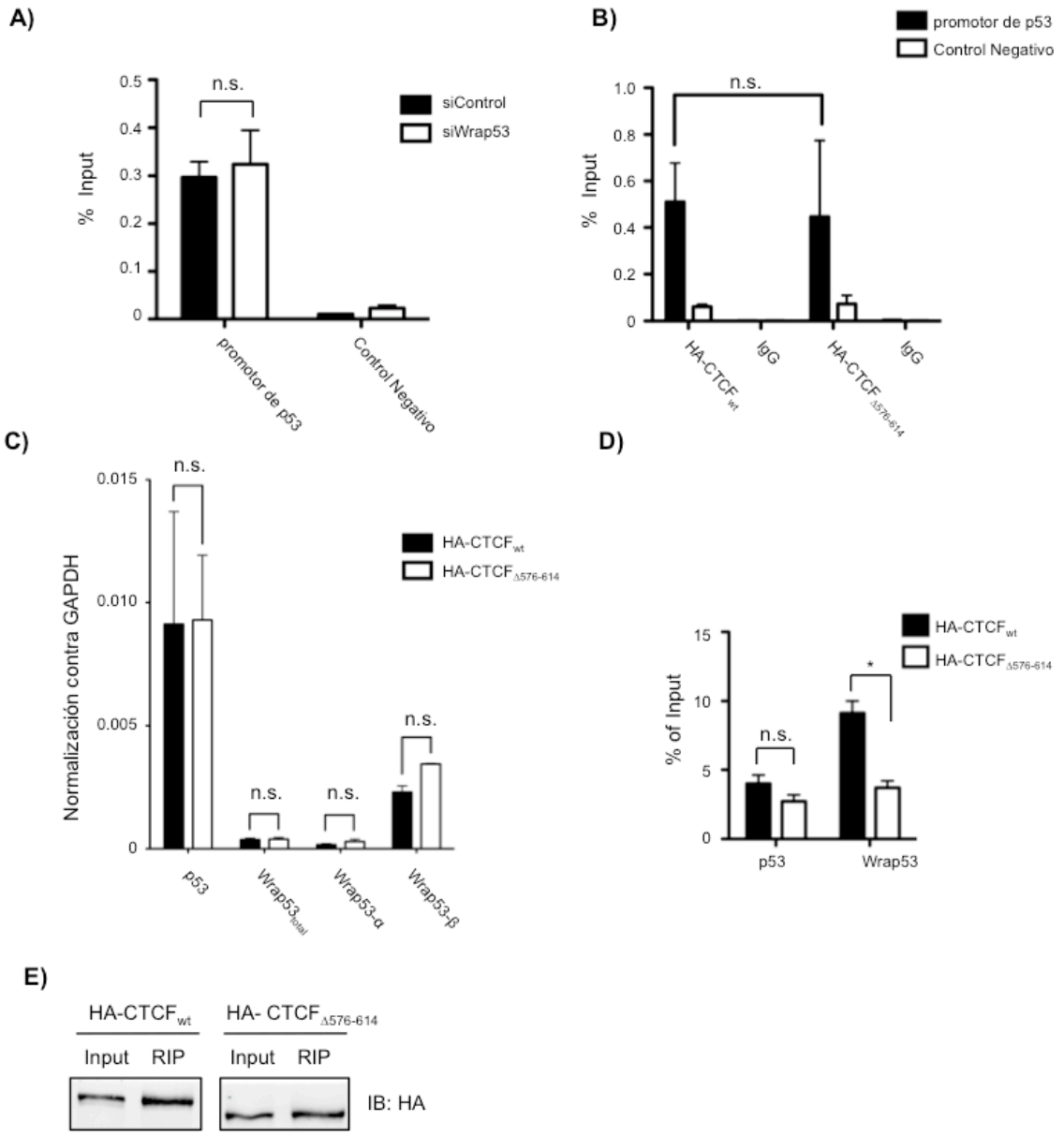


Figura 12. El papel de Wrap53 en el reclutamiento de CTCF endógeno y recombinante.
(A) ChIP qPCR utilizando un anticuerpo contra CTCF en células U2OS después de 24 hrs de transfección con siWrap53 o siControl. (B) ChIP qPCR utilizando un anticuerpo contra IgG o HA en líneas celulares que expresan HA-CTCF$_{\Delta 576\text{-}614}$ o HA-CTCF$_{wt}$ después de tratamiento de 72 hrs con doxiciclina. Los primers están situados en el promotor del gen *p53* o en el exón 27 del gen *Rb* como control negativo. (C) Niveles de ARNm obtenidos por RT-qPCR y normalizados contra GAPDH. (D) RIP utilizando anticuerpos contra HA en el peso establecido y Δ576-614 líneas. RT-qPCR a partir de ARN recuperado se muestra como % de *Input*, normalizado contra los niveles de GAPDH. (E) Inmunoblot del mismo material que en (D). n.s. = no significativo. * p <0,05 obtenido mediante la prueba de Mann-Whitney. Las barras indican la media de al menos 3 repeticiones biológicos y el error estándar.

Además, en el experimento de rescate, HA-CTCF$_{\Delta 576\text{-}614}$ no mostró defectos en la unión al promotor del gen *p53* (Figura 12B), o en los niveles de ARNm endógenos de *p53* o de *Wrap53* (Figura 12C), pero sí mostró una disminución en la co-precipitación con Wrap53 (Figura 12D-E) destacando la relevancia de su defecto en condiciones de daño en el ADN.

En su conjunto estos resultados muestran que una mutante de CTCF con un defecto en la unión al ARN no puede restaurar las características de la respuesta al daño al ADN en células con un *knockdown* de CTCF endógeno, posiblemente debido a su incapacidad para unirse al ARN de *Wrap53* impidiendo así la inducción transcripcional del locus de p53.

El ARN ayuda en la formación de dímeros/multímeros de CTCF

CTCF puede interactuar con una gran cantidad de proteínas, incluyendo otras moléculas de CTCF (Valadez-Graham et al. 2004; Yusufzai et al. 2004). Sin embargo, las bases moleculares para su dimerización/multimerización no se han reportado. Si la dimerización/multimerización de CTCF esta modulada mediante ARN, esto podría tener importantes implicaciones en la formación de asas de cromatina. Especulamos que las interacciones ARN-CTCF podrían participar en la formación de multímeros de CTCF así facilitando la interacción entre los monómeros.

Con este fin, aprovechamos las versiones de la proteína CTCF con diferentes etiquetas que teníamos a nuestra disposición. Primero probamos la capacidad de las proteínas completas para interactuar entre sí. La proteína GST-CTCF$_{wt}$ purificada en *E. coli* se añadió a extractos nucleares derivados de la línea celular que expresa HA-CTCF$_{wt}$ y después se analizó en ensayos de pull-down para GST seguida de un inmunoblot contra HA. Consistente con los reportes en la literatura, la proteína GST-CTCF$_{wt}$ es capaz de co-precipitar HA-CTCF$_{wt}$ (Figura 13A-B). Sin embargo, cuando se añadieron RNasas durante el periodo de incubación observamos una disminución de ~50% en la cantidad de HA-CTCF$_{wt}$ que co-precipitaba, comparado con el caso sin tratamiento; esta disminución no se observó con la adición de DNasas (Figura 13A-B). Esto lo podemos atribuir a

una de dos posibles explicaciones: que otras proteínas que interactúan con CTCF también pueden contribuir a su oligomerización o que solo se logró una digestión parcial del ARN. También es posible que una sub-población de CTCF forme díameros para realizar funciones independientes de RNA. Adicionalmente, también logramos concluir que la RUR es suficiente para lograr la dimerización ya que la incubación con GST-RUR también dió lugar a una eficiente co-precipitación de HA-CTCF$_{wt}$, mientras que la adición de RNasas evitó por completo la recuperación de HA-CTCF$_{wt}$ y el tratamiento con DNasas de nuevo no tuvo ningún efecto (Figura 13C). Es importante destacar que la capacidad de CTCF para interactuar con el ARN se requiere para la dimerización/multimerización, concluimos esto ya que no pudimos observar la co-precipitación de HA-CTCF$_{wt}$ o la mutante HA-CTCF$_{\Delta 576-614}$ cuando se incubó con GST-RUR$_{\Delta 576-614}$ (Figura 13C y D, respectivamente).

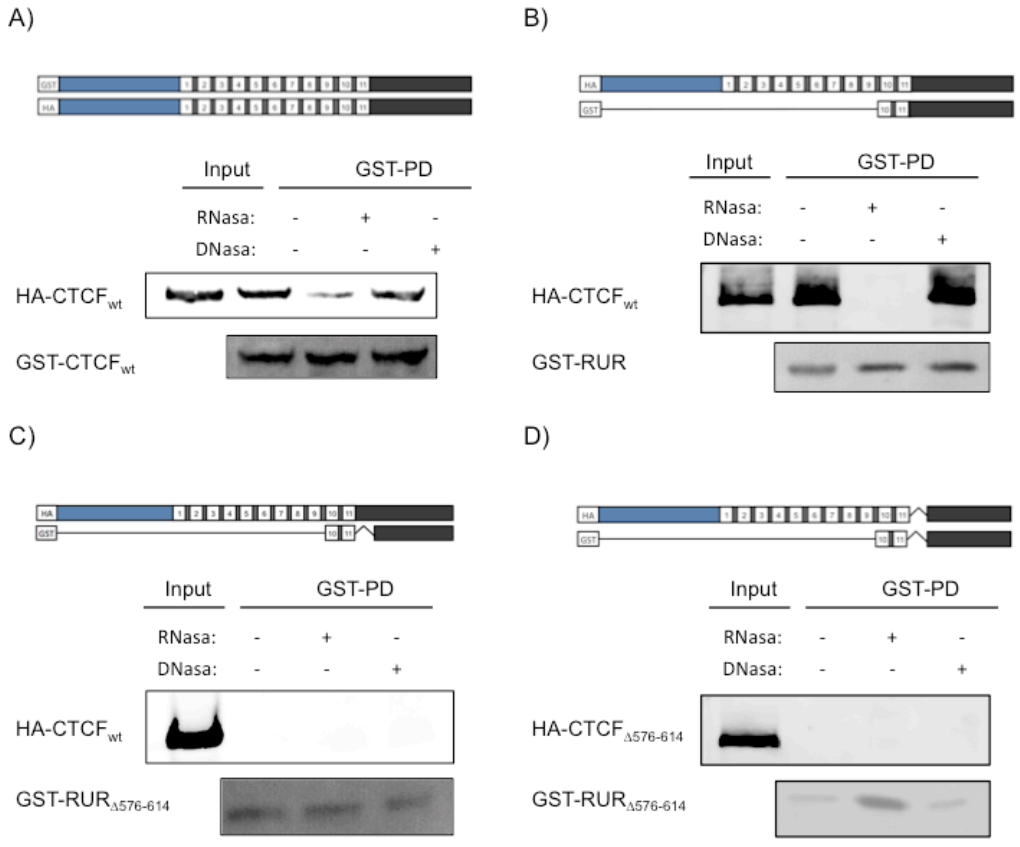


Figura 13. El ARN facilita la dimerización/multimerización CTCF .
Los pull-down con GST se realizaron con extractos nucleares de células U2OS que expresan ya sea (A - C) HA-CTCF$_{wt}$ o (D) HA-CTCF$_{\Delta 576-614}$ y se incubaron con la versión marcada con GST de

CTCF indicada, en presencia o ausencia de nucleasas. Las proteínas GST se tiñeron con Rojo de Ponceau y las proteínas HA-CTCF se detectaron con anticuerpos contra HA. Un diagrama esquemático de las proteínas utilizadas en cada experimento se presenta por encima de cada panel.

Para corroborar esta observación mediante otra técnica, eliminamos la etiqueta GST del $CTCF_{wt}$ y realizamos cromatografía de exclusión en presencia o ausencia de 167 pb del ARN de *Wrap53*. El perfil de elución de la muestra de CTCF alcanzó una masa aparente de 265 kDa, lo que indicaría la formación de un trímero. Curiosamente, el perfil de elución en la presencia del ARN de *Wrap53* sugirió la formación de un complejo más grande que 2 MDa (Figura 14A). Como evidencia adicional para sugerir que el ARN ayuda a la dimerización/multimerización de CTCF a través de su RUR, el perfil de elución del fragmento RUR (~22 kDa) fue de alrededor de 60 kDa, pero fue incrementado hasta 1 MDa en presencia del ARN de *Wrap53* (Figura 14B).

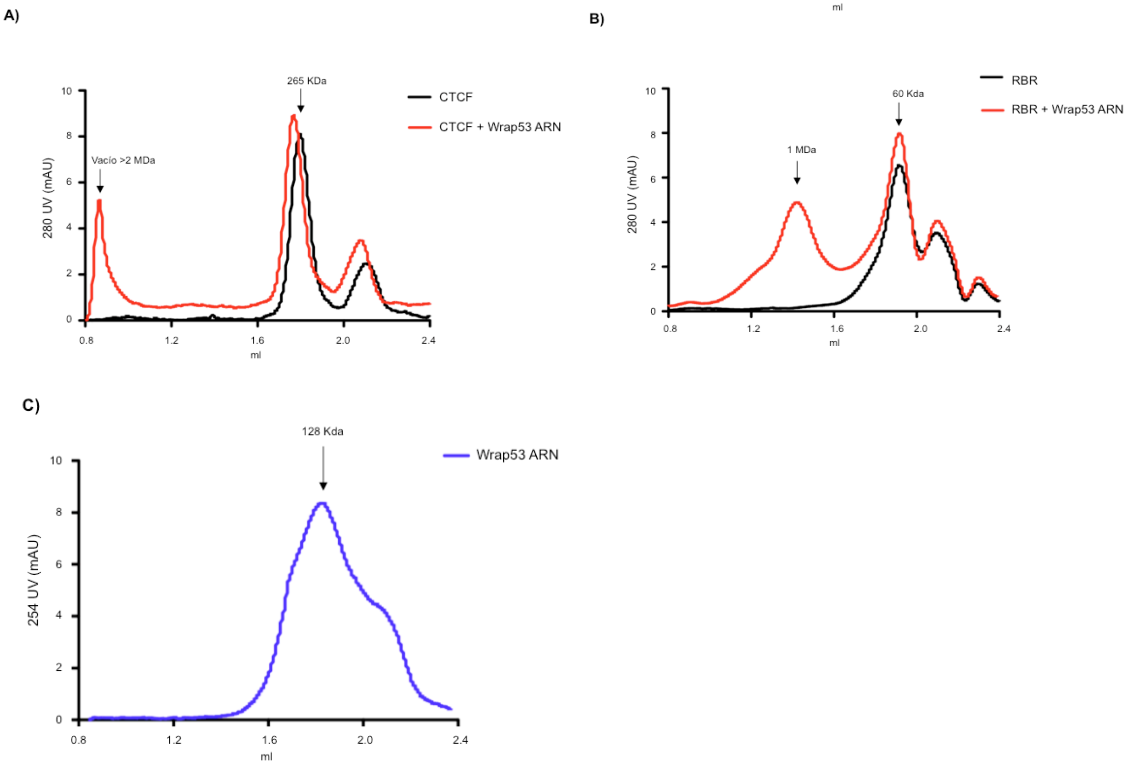


Figura 14. Los perfiles de elución de CTCFwt y RUR.
Perfil de elución de (A) CTCF completo, (B) la RUR en presencia (rojo) o ausencia (negro) del ARN de Wrap53. (C) El perfil de elución del ARN de Wrap53 solo.

En este último caso, los perfiles de elución del ARN y la RUR indican fuertemente su co-elución ya que migran a un peso molecular significativamente mayor que el predicho para la RUR (Figura 14B) o el ARN (Figura 14C) por si mismos. Tomados en conjunto, estos resultados sugieren que el ARN ayuda a la formación de multímeros de CTCF a través de su RUR.

DISCUSIÓN Y CONCLUSIONES

El ARN ha atraído mucha atención recientemente como consecuencia de una multitud de funciones inesperadas que se le han atribuido. Sin embargo, el ARN siempre funciona a través de la interacción con proteínas. Por lo cual es de gran importancia estudiar la compleja relación entre ARNs y proteínas. Los resultados obtenidos en este proyecto de tesis doctoral apoyan la conclusión que el ARN tiene un papel importante en las funciones reguladoras de CTCF. Esto claramente se suma a la gran variedad de funciones que ya se reconocen para CTCF. Mostramos que la unión de *Wrap53* a CTCF está relacionada con la respuesta transcripcional del gen *p53* durante condiciones de daño al ADN. También presentamos la primera base de datos de transcritos que interactúan con CTCF. Esto proporciona información sobre la gran variedad de posibles blancos de CTCF así como la base molecular para la regulación a este nivel. Nuestros hallazgos sugieren una función general donde las interacciones CTCF-ARN regulan la dimerización/multimerización de CTCF. Estos resultados tienen implicaciones potencialmente importantes en la formación y regulación de asas de cromatina.

El gen *p53* humano se encuentra mutado en aproximadamente el 50% de los tumores cancerígenos (National Cancer Institute, 2010). Sin embargo, un gran número de tumores mantiene a p53 sin mutaciones, lo que sugiere que su expresión y/o función puede ser inactivada a través de otros mecanismos. Determinamos anteriormente que la pérdida de CTCF en su sitio de unión en el promotor del gen *p53* causa el silenciamiento epigenético de su expresión mediante el enriquecimiento de marcas de cromatina asociadas a represión

(Soto-Reyes y Recillas-Targa 2010). El *knockdown* de *Wrap53* fue reportado por otro grupo como causa de la regulación negativa de los niveles transcripcionales del gen *p53* (Mahmoudi et al. 2009). Estas observaciones iniciales proporcionaron la base experimental de cómo se regula el locus de p53. En este proyecto, expandimos la evidencia molecular mediante la cual CTCF participa en la regulación transcripcional del gen p53 humano. Nuestra observación que el *knockdown* de CTCF causa el decremento de los niveles de ARNm tanto de *p53* como de *Wrap53* apunta a un mecanismo interdependiente que involucra tanto a CTCF como a *Wrap53* (Figura 4). Nuestra hipótesis inicial que CTCF podría tener contactos directos con el ARN de Wrap53 fue apoyada inicialmente por su co-inmunoprecipitación y luego se confirmó a través de ensayos *in vitro*. Logramos demostrar que CTCF tiene la habilidad de unirse al ARN que no había sido reconocida previamente. Esta requiere de sus DZ 10-11 y el carboxilo terminal y puede ser separada de su bien reconocida capacidad de unión al ADN (Figura 7). Nakahashi y colaboradores reportaron recientemente que CTCF se une principalmente a su secuencia de reconocimiento al ADN con sus DZ 4-7 y que sus DZ restantes tienen más bien un papel estabilizador (Figura 2). También describieron que DZ 9-11 se puede asociar en ~15% de sus sitios de unión con una segunda secuencia conservada río arriba del motivo principal (Nakahashi et al. 2013). Este estudio complementa nuestras observaciones, ya que potencialmente el ~85% de los sitios de unión a CTCF podrían estar libres para hacer contactos adicionales con ARN a través de su RUR o con cualquier CTCF libre antes de unirse al ADN.

El dominio de DZ de CTCF está conservado casi por completo entre especies como el ratón, el pollo y el humano (Ohlsson y Renkawitz, 2001). Este comprende diez DZ de tipo C_2H_2 y un onceavo tipo C_2HC. Los resultados presentados aquí indican que los DZ 10-11 se unen específicamente al ARN. Además de CTCF, hay un número de otras proteínas con DZ en diferentes organismos que se han reportado pueden unirse tanto a ADN como a ARN (Brown, 2005). El TFIIIA en mamíferos es un excelente ejemplo para este tipo de

actividad, esta proteína nos puede proporcionar información importante sobre cómo las proteínas con múltiples DZ de tipo C_2H_2 pueden unirse específicamente tanto al ADN como al ARN. De manera similar a nuestros resultados con CTCF, el TFIIIA utiliza diferentes DZ para alcanzar la especificidad entre el ADN y el ARN. Este requiere los DZ 1-3 para unirse al ADN y sus DZ centrales 4-6 para unirse al ARN (Searles et al. 2000). El TFIIIA en complejo con el ARN ribosomal 5S fue la primera estructura DZ-ARN que fue resuelta usando cristalografía, revelando que la manera en que los DZ unen ARN es diferente a su bien estudiada unión al ADN (Lu et al. 2003). Es importante destacar que estos DZ pueden unirse tanto al ADN como al ARN, tanto a sus versiones de doble cadena o cadena sencilla, pero con diferentes afinidades (Brown, 2005). El reconocimiento del ARN de cadena simple por DZ tipo C_2H_2 se lleva a cabo a través de las cadenas laterales con residuos aromáticos que se intercalan entre las bases de di-nucleótidos que se encuentran espaciados adecuadamente, mientras que el reconocimiento del ADN de doble cadena es a través de las hélices α formando puentes de hidrógeno con las bases de la zurco mayor (Brown, 2005). A diferencia del TFIIIA que muestra una alta especificidad para el ARN ribosomal 5S, nuestros resultados muestran que, al menos *in vitro*, CTCF reconoce al ARN con poca especificidad en su secuencia. Nuestra hipótesis apoya la idea que en lugar de dependencia únicamente en la especificidad de secuencia, los contactos de CTCF-ARN surgen de una manera dependiente a su contexto dentro de la célula y quizás favorecen estructuras secundarias o terciarias complejas del ARN que en la actualidad son difíciles de replicar en experimentos *in vitro*.

Los resultados discutidos anteriormente respaldan que CTCF puede hacer contactos directos con el ARN, sin embargo la relevancia biológica general de esta actividad sigue quedando poco clara. En el contexto de la regulación transcripcional del gen *p53*, nuestros hallazgos sacaron a la luz la importancia funcional de la interacción de CTCF con *Wrap53* durante la respuesta al daño del ADN. En células con *knockdown* de CTCF, una mutante en la unión al ARN

$CTCF_{\Delta 576\text{-}614}$ no puede rescatar los niveles de ARNm adecuados de *p53*, *Wrap53*, *p21* y *PUMA*. En contraste con el caso de $CTCF_{wt}$ que si restablece la respuesta normal al daño del ADN (Figura 11). Estos resultados proporcionan evidencia apoyando que CTCF regula la transcripción adecuada de *p53* y *Wrap53* durante la respuesta al daño del ADN de dos maneras diferentes (1) al unirse a la región promotora del gen p53 (Soto Reyes y Recillas-Targa 2010) y (2) al interactuar con el ARN de *Wrap53* (Figura 15).

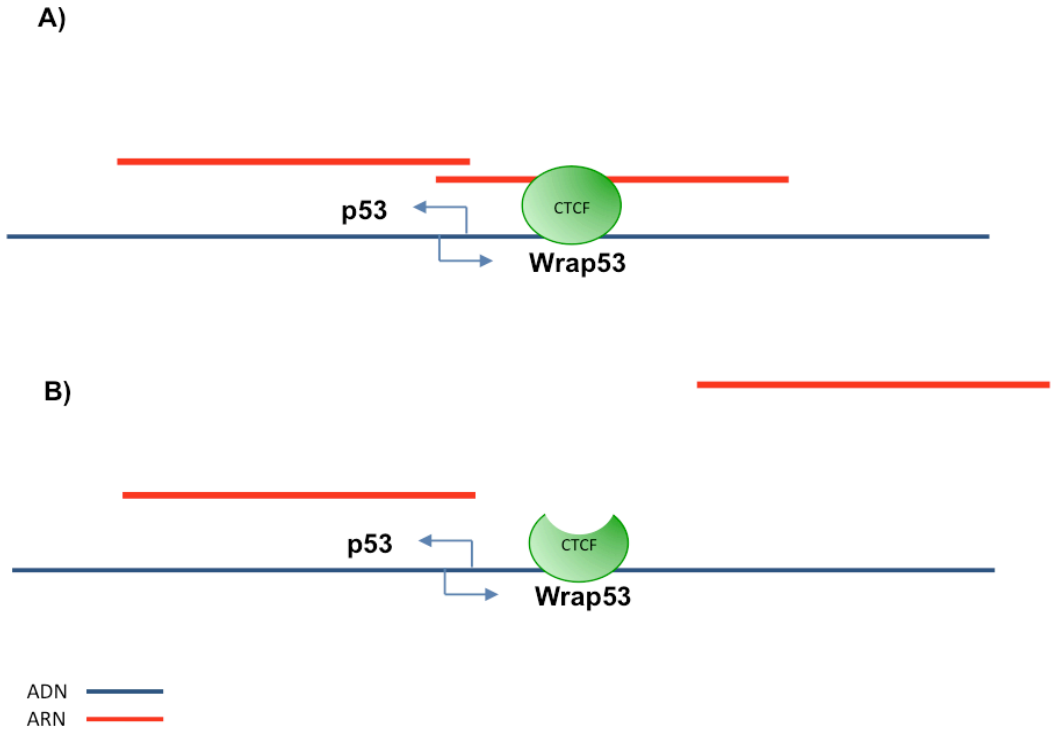


Figura 15. Modelo propuesto para la regulación del gen *p53*.
(A) En condiciones normales, la presencia de CTCF permite que los transcritos de *p53* y *Wrap53* interactúen. Permitiendo que Wrap53 regule positivamente a *p53*. (B) Cuando la asociación de CTCF-Wrap53 se pierde en presencia de $CTCF_{\Delta 574\text{-}614}$ no se puede inducir la expresión de *p53* en respuesta al daño al ADN.

Otra función específica para CTCF se reportó recientemente en el que CTCF se une al ARN no-codificante *JPX* en el contexto de la inactivación del cromosoma X (Sun et al. 2013). Este grupo propuso que CTCF podría ser desalojado de su sitio de unión por la sobre-expresión de *JPX*. Sin embargo, esto no es un mecanismo general de CTCF ya que esta restringido a este sitio de unión específico en un alelo del locus de Xist, siendo irrelevante para los demás sitios de unión a CTCF que se evaluaron. Hasta la fecha, sólo contamos con evidencia

experimental que apoya funciones en loci específicos para interacciones ARN-CTCF. Por lo tanto, la importancia biológica general para estas interacciones sigue siendo difícil de dilucidar.

CTCF fue descrito por primera vez como un factor de transcripción (Filippova et al. 1996) y más tarde un número de estudios la describieron como un *insulator* (Barkess y West 2012). Sin embargo, la literatura reciente sugiere que su función como *insulator* puede ocurrir sólo en algunos casos, probablemente como consecuencia de su función principal como organizador de la cromatina (Handoko et al 2011; Sanyal et al 2012; Phillips-Cremins et al 2013). CTCF está implicada en una gran variedad de funciones en la regulación de genes pero quizás su papel más importante es la formación de asas de cromatina (Phillips y Corces 2009). En este sentido, nuestros datos apoyan la hipótesis donde el ARN nuclear facilita la dimerización/multimerización de CTCF y por lo tanto la formación de asas de cromatina (Figura 16).

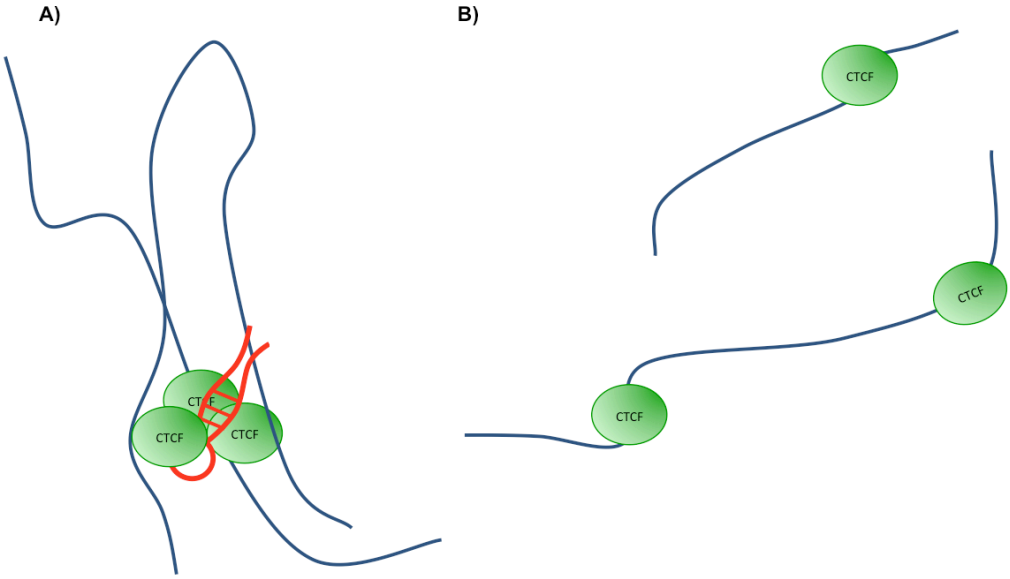


Figura 16. Modelo propuesto para la dimerización/multimerización de CTCF.
(A) En algunas partes del genoma, la presencia de asas de cromatina dependientes de CTCF se forman y/o mantienen con la ayuda de moléculas de ARN. (B) En ausencia de ARN estas asociaciones podrían perderse.

Basamos esta hipótesis en la evidencia que acumulamos donde observamos que el ARN nuclear era esencial para las interacciones CTCF-CTCF mediante pull-downs (Figura 13) y porque CTCF-Wrap53 forma complejos mucho mas grandes de lo esperado cuando utilizamos cromatografía de exclusión por tamaño (Figura 14). La dependencia de ARN en la multimerización de proteínas también se ha observado anteriormente en otros casos. Por ejemplo, la citidina-desaminasa humana APOBEC3G puede oligomerizar sólo en la presencia de ARN (Huthoff et al. 2009). De manera similar, la proteína p53 interactúa con p68/p72 y el complejo Drosha de una manera dependiente al ARN (Suzuki et al. 2009). Además, la interacción física de CTCF con la helicasa de ARN p68 (DDX5) también ha demostrado ser dependiente de un ARN no-codificante asociado, SRA (Yao et al. 2010). Interesantemente, los autores de ese artículo propusieron que otros ARNs pueden sustituir parcialmente a SRA en este contexto, apoyando nuestra hipótesis donde las interacciones ARN-proteína son dependientes del contexto celular y no tanto de la especificidad de secuencia.

Podemos concluir que una característica general de CTCF es la capacidad de unir una gran variedad de ARNs. Sin embargo, al día de hoy sólo conocemos dos funciones específicas que son funcionalmente diferentes para dos ARNs distintos, *JPX* (Sun et al. 2013) y *Wrap53*. Hasta el momento hemos sugerido una posible formación de dimeros/multímeros CTCF dependientes de ARN *in vitro*. Que si se demuestra *in vivo* tendría implicaciones muy importantes en la organización nuclear mediante la regulación de asas de cromatina. Esta hipótesis abre a su vez la puerta a muchas nuevas preguntas. ¿Algunas asas de cromatina se forman, mantienen y regulan gracias a las interacciones CTCF-ARN? ¿Otras proteínas estructurales son dependientes de ARN? ¿Cuáles son las señales regulatorias para que CTCF pueda unir ARN? ¿La unión al ADN es independiente de la unión al ARN, o una precede a la otra? ¿La metilación del ARN puede ser una manera de regular su interacción con CTCF? ¿Cuál es la base molecular para lograr tanto funciones generales como específicas a un

locus? Todas estas son preguntas interesantes y merecen ser desarrolladas en proyectos de investigación en el futuro.

Finalmente, nuestra caracterización bioquímica de la RUR al ARN dentro de CTCF y la descripción de sus ARNs asociados facilitará nuevos estudios para dilucidar nuevas funciones regulatorias. La multimerización de CTCF dependiente de ARN y su potencial en la regulación de asas de cromatina probablemente tendrá implicaciones importantes en investigaciones futuras en estudios sobre organización nuclear y regulación epigenética.

PERSPECTIVAS

Encontrar e interpretar los determinantes genéticos y no genéticos de enfermedades humanas representa una de las principales empresas en la investigación biomédica. Se han realizado avances substanciales utilizando estudios de asociación de todo el genoma hacia la identificación de variaciones genéticas causales en las enfermedades humanas. La función celular en una población sana no sólo depende de la estabilidad genómica, sino también en el mantenimiento de la homeostasis epigenómica. Las numerosas asociaciones de funciones epigenéticas en las enfermedades humanas, sobre todo en el cáncer, apoyan aún más esta noción. Sin embargo, hay un retraso significativo en los estudios epigenómicos ya que la asociación de causa y efecto puede ser muy complicado. Recientemente se han empezado a tomar en cuenta regiones no codificantes del genoma para buscar variaciones que puedan ser causantes de patologías. Estoy interesado en investigar el papel de las interacciones RNA-proteína en el contexto de la organización nuclear y cómo éstas influyen en la etiología de la enfermedad compleja.

Puedo concluir que el hallazgo más importante durante mi investigación doctoral es la relación que observamos entre la organización nuclear y el ARN. Nuestras observaciones sugieren que el ARN sirve para formar/estabilizar la interacción entre proteínas CTCF, favoreciendo su dimerización/multimerización y que esto es independiente de la unión de CTCF al ADN. El hecho que CTCF forme estos complejos y dependa de RNA indica que su formación depende de transcripción. Esto podría estar conectando directamente la actividad transcripcional con el re-modelamiento de asa cromatínicas, a cierto grado, dependientes de CTCF. Esto también se ha sugerido para las fabricas de transcripción en el establecimiento de programas de expresión tejido-específico (Eskiw et al. 2010). Estas interacciones pueden tener importantes implicaciones en la organización nuclear y la regulación de genes. Además, mis observaciones apoyan dos de los mecanismos de acción propuestos para lincARNs, aunque no necesariamente excluye que otros ARNs tanto codificantes y no-codificantes puedan estar

involucrados. 1) Los lincARNs podrían funcionar como andamios que sirven para que dos o más proteínas formen complejos y 2) los lincARNs ayudan a formar asas de cromatina para acercar secuencias regulatorias, como enhancers y promotores, que de otra manera se encuentran a grandes distancias unas de otras.

Como perspectivas directas a esta tesis tengo la intención, como parte de mi investigación posdoctoral, de identificar jugadores adicionales en el proceso por el cual las interacciones ARN-proteína modulan la regulación de genes y la organización nuclear. Existen una variedad de enfoques que puedo abordar como por ejemplo:

- Caracterizar otros RNAs asociados con CTCF e identificar asi otros loci donde pueda tener una función reguladora directa.
- Evaluar la relevancia de las interacciones proteína-ARN, mediante purificación de complejos de proteínas asociadas a CTCF con/sin un tratamiento con RNasas.
- Evaluar la relevancia del ARN para la interacción de cohesinas y Mediador en el contexto de la formación de asas de cromatina.
- Realizar análisis de expresión de lincARNs en diferentes contextos celulares para comparar las condiciones control frente a *knockdown* de CTCF y/o sus co-factores.
- Utilizar Hi-C y ChIA-PET para determinar las interacciones entre asas de cromatina que son dependientes de ARN y determinar su importancia en un contexto biológico relevante.

Finalmente, mi objetivo a largo plazo consiste en explotar los datos acumulados durante el desarrollo de esta línea de investigación para el diseño y elaboración de herramientas de diagnóstico o tratamiento para el cáncer y otras enfermedades complejas.

REFERENCIAS

- Barkess G, West AG. 2012. Chromatin insulator elements: establishing barriers to set heterochromatin boundaries. *Epigenomics* **4**: 67–80.
- Bell AC, West AG, Felsenfeld G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell.* **98(3)**: 387–396.
- Bhinge, AA, Kim J, Euskirchen GM, Snyder M & Iyer VR. 2007. Mapping the chromosomal targets of STAT1 by Sequence Tag Analysis of Genomic Enrichment (STAGE). *Genome Res.* **17**: 910–916
- Brown RS. 2005. Zinc finger proteins: getting a grip on RNA. *Curr Opin Struct Biol* **15**: 94–98.
- Chen H, Tian Y, Shu W, Bo X, Wang S. 2012. Comprehensive Identification and Annotation of Cell Type-Specific and Ubiquitous CTCF-Binding Sites in the Human Genome. *PLoS ONE* **7**: e41374.
- Dignam JD, Lebovitz RM, Roeder RG. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* **11**: 1475–1489.
- ENCODE Project Consortium, 2011. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS biology. **9(4)**: p.e1001046.
- Eskiw CH, Cope NF, Clay I, Schoenfelder S, Nagano T, Fraser P. 2010. Transcription factories and nuclear organization of the genome. *Cold Spring Harb Symp Quant Biol.* **75:**501-6.
- Filippova G, Fagerlie S, Klenova E. 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc

fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Molecular and Cellular Biology* **16(6)**: 2802-13.

- Gross DS, Garrard WT. 1988. Nuclease hypersensitive sites in chromatin. *Annu Rev Biochem* **57**: 159–197.
- Hafner M, Landgraf P, Ludwig J, Rice A, Ojo T, Lin C, Holoch D, Lim C, Tuschl T. 2008. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods* **44**: 3–12.
- Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CWH, Ye C, Ping JLH, Mulawadi F, et al. 2011. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet* **43**: 630–638.
- Holwerda SJB, de Laat W. 2013. CTCF: the protein, the binding partners, the binding sites and their chromatin loops. *Philosophical Transactions of the Royal Society B: Biological Sciences* **368**: 20120369–20120369.
- Huthoff H, Autore F, Gallois-Montbrun S, Fraternali F, Malim MH. 2009. RNA-dependent oligomerization of APOBEC3G is required for restriction of HIV-1. *PLoS Pathog* **5**: e1000330.
- Kaneko S, Li G, Son J, Xu CF, Margueron R, Neubert TA, and Reinberg D. 2010. Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. *Genes Dev.* **24**: 2615–2620.
- Kaneko, S., Son, J., Shen, S.S., Reinberg, D., and Bonasio, R. (2013). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**: 1258–1264.
- Kaneko S, Bonasio R, Saldaña-Meyer R, Yoshida T, Son J, Nishino K, Umezawa A, Reinberg D. 2014. Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to Chromatin. *Mol Cell*.

53(2): 290-300.

- Kim TG, Kraus JC, Chen J, and Lee Y. 2003. JUMONJI, a critical factor for cardiac development, functions as a transcriptional repressor. J. Biol. Chem. **278**: 42247–42255.
- La Rosa-Velázquez De IA, Rincón-Arano H, Benítez-Bribiesca L, Recillas-Targa F. 2007. Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Res* **67**: 2577–2585.
- Li G, Margueron R, Ku M, Chambon P, Bernstein BE, and Reinberg D. 2010. Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev.* **24**: 368–380.
- Lobanenkov VV, Nicolas RH, Adler VV, Paterson H, Klenova EM, Polotskaja AV, Goodwin GH. 1990. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene* **5**: 1743–1753.
- Lobanenkov VV, Nicolas RH, Plumb MA, Wright CA, Goodwin GH. 1986. Sequence-specific DNA-binding proteins which interact with (G + C)-rich sequences flanking the chicken c-myc gene. *Eur J Biochem* **159**: 181–188.
- Lu D, Searles MA, Klug A. 2003. Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature* **426**: 96–100.
- Mahmoudi S, Henriksson S, Corcoran M, Méndez-Vidal C, Wiman KG, Farnebo M. 2009. Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. *Mol Cell* **33**: 462–471.
- Mahmoudi S, Henriksson S, Weibrecht I, Smith S, Söderberg O, Strömblad S, Wiman KG, Farnebo M. 2010. WRAP53 is essential for Cajal body

formation and for targeting the survival of motor neuron complex to Cajal bodies. *PLoS Biol* **8**: e1000521.

- Martin D, Pantoja C, Miñán AF, Valdes-Quezada C, Moltó E, Matesanz F, Bogdanović O, La Calle-Mustienes De E, Domínguez O, Taher L, et al. 2011. Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. *Nat Struct Mol Biol* **18**: 1084–714.
- Mili S, Steitz JA. 2004. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10**: 1692–1694.
- Nakahashi H, Kwon K-RK, Resch W, Vian L, Dose M, Stavreva D, Hakim O, Pruett N, Nelson S, Yamane A, et al. 2013. A Genome-wide Map of CTCF Multivalency Redefines the CTCF Code. *Cell Rep* **3(5)**: 1678-89.
- National Cancer Institute. 2010. Surveillance Epidemiology and End Results [online], http://seer.cancer.gov/
- Ohlsson R, Renkawitz R. 2001. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *TRENDS in Genetics*. **17(9)**: 520-527.
- Pasini D, Cloos PA, Walfridsson J, Olsson L, Bukowski JP, Johansen JV, Bak M, Tommerup N, Rappsilber J, Helin K. 2010. JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**: 306–310.
- Phillips JE, Corces VG. 2009. CTCF: master weaver of the genome. *Cell* **137**: 1194–1211.
- Phillips-Cremins JE, Sauria MEG, Sanyal A, Gerasimova TI, Lajoie BR, Bell

JSK, Ong C-T, Hookway TA, Guo C, Sun Y, et al. 2013. Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell* **153**: 1281–1295.

- Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, Bell AC, Litt, MD, West AG, Gaszner M, Felsenfeld G. 2002. Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proceedings of the National Academy of Sciences of the United States of America.* **99(10)**: 6883–6888.
- Riley KJ, Yario TA, Steitz JA. 2012. Association of Argonaute proteins and microRNAs can occur after cell lysis. *RNA* **18**: 1581–1585.
- Saldaña-Meyer R, Recillas-Targa F. 2011. Transcriptional and epigenetic regulation of the p53 tumor suppressor gene. *Epigenetics* **6**: 1068-1077.
- Sanyal A, Lajoie BR, Jain G, Dekker J. 2012. The long-range interaction landscape of gene promoters. *Nature* **489**: 109–113.
- Searles MA, 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.
- Son J, Shen SS, Margueron R, Reinberg D. 2013. Nucleosome binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes Dev.* **27**: 2663-2677.
- Soto-Reyes E, Recillas-Targa F. 2010. Epigenetic regulation of the human p53 gene promoter by the CTCF transcription factor in transformed cell lines. *Oncogene* **29**: 2217–2227.
- Stadtfeld M, Apostolou E, Akutsu H, Fukuda A, Follett P, Natesan S, Kono T, Shioda T, Hochedlinger K. 2010. Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature*

465: 175–181.

- Sun S, del Rosario BC, Szanto A, Ogawa Y, Jeon Y, Lee JT. 2013. Jpx RNA Activates Xist by Evicting CTCF. *Cell* **153**: 1537–1551.
- Suzuki HI, Yamagata K, Sugimoto K, Iwamoto T, Kato S, Miyazono K. 2009. Modulation of microRNA processing by p53. *Nature* **460**: 529–533.
- Terribilini M, Sander JD, Lee J-H, Zaback P, Jernigan RL, Honavar V, Dobbs D. 2007. RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. *Nucleic Acids Res* **35**: 578–84.
- Valadez-Graham V, Razin SV, Recillas-Targa F. 2004. CTCF-dependent enhancer blockers at the upstream region of the chicken alpha-globin gene domain. *Nucleic Acids Res* **32**: 1354–1362.
- Vousden KH, Prives C. 2009. Blinded by the Light: The Growing Complexity of p53. *Cell* **137**: 413–431.
- Ulitsky I, Shkumatava A, Jan CH, Sive H, Bartel DP. 2011. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell **147**: 1537–1550. Wang, K.C.,
- Xiao T, Wallace J, Felsenfeld G. 2011. Specific Sites in the C Terminus of CTCF Interact with the SA2 Subunit of the Cohesin Complex and Are Required for Cohesin-Dependent Insulation Activity. *Molecular and Cellular Biology* **31**: 2174–2183.
- Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, Lajoie BR, Protacio A, Flynn RA, Gupta RA, et al. 2011. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**: 120–124.

• Yao H, Brick K, Evrard Y, Xiao T, Camerini-Otero RD, 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.

• Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G. 2004. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell* **13**: 291–298.

• Zlatanova J, Caiafa P. 2009. CTCF and its protein partners: divide and rule? *J Cell Sci* **122**: 1275–1284.

ANEXOS

ESTANCIA DE INVESTIGACIÓN

Durante el doctorado tuve la oportunidad de realizar una estancia de investigación en el laboratorio del Dr. Danny Reinberg en la Escuela de Medicina de la Universidad de Nueva York (NYU). Durante esta estancia, desarrollé parte de mi proyecto principal de tesis, simultáneamente también participe en el desarrollo de otro proyecto de investigación sobre interacciones ARN-Proteína que acaba de ser publicado y que resumo a continuación.

Interacciones entre JARID2 y ARNs no-codificantes regulan el reclutamiento del complejo PRC2 a la cromatina

JARID2 es un componente accesorio del complejo represivo Polycomb 2 (PRC2) que es necesario para la diferenciación de las células madre embrionarias (CME). Se ha propuesto un papel para JARID2 en el reclutamiento de PRC2 durante el proceso de diferenciación, pero los detalles moleculares de este mecanismo todavía no se conocen. La evidencia experimental que obtuvimos es la siguiente:

1) Demostramos que JARID2 une ARN *in vitro* e *in vivo*.

2) Identificamos que JARID2 une ARN con una región de 30 aminoácidos localizada en su amino terminal. Esta región no contiene características o dominios anotados, a pesar del hecho de que además de la unión al ARN, también es responsable de las interacciones con SUZ12, EZH2, nucleosomas y de la estimulación de PRC2 (Kim et al. 2003; Li et al. 2010; Pasini et al. 2010; Son et al. 2013).

3) También observamos que la unión al ARN estimula la asociación de JARID2 con EZH2 *in vitro* y también el reclutamiento de EZH2 a algunos sitios de JARID2 *in vivo* (Figura 17A).

4) Comparamos y obtuvimos bases de datos obtenidas por PAR-CLIP-seq de dos proteínas de PRC2, JARID2 y EZH2, que también tiene la habilidad de unir ARN (Kaneko et al. 2010). Esto nos permitió determinar que JARID2 se une a una variedad de lincARNs diferentes a los que se unen a EZH2 lo que demuestra especificidad de unión *in vivo*.

5) De los transcritos que interaccionan con JARID2, centramos nuestro análisis en el transcrito MEG3 por ser un importante regulador del desarrollo (Stadtfeld et al. 2010). Usamos diferentes líneas de células humanas pluripotenciales inducidas (iPSCs) que mostraban diferencias drásticas de MEG3, ya sea altas o bajas, pero no en la expresión de otros genes. Con este sistema demostramos que la falta de MEG3 puede actuar en *trans* y que este transcrito es necesario para reclutar a JARID2 a algunos sitios blanco en el genoma (Figura 17B).

Aunque los resultados presentados anteriormente sugieren un mecanismo molecular por el cual MEG3 ayuda a JARID2 y a la función de PRC2, se necesita más evidencia para abordar algunas preguntas importantes en estudios futuros: (1) cómo este mecanismo se extiende a los demás lincARNs identificados por PAR-CLIP, y (2) cómo los lincARNs regulan el reclutamiento *de novo* de estas proteínas a loci distantes. Nuestros análisis de unión *in vitro* de MEG3 sugieren que JARID2 no se une al ARN de una manera específica de secuencia, consistente con el hecho de que la secuencia primaria de los lincARNs no está bien conservada (Ulitsky et al. 2011).

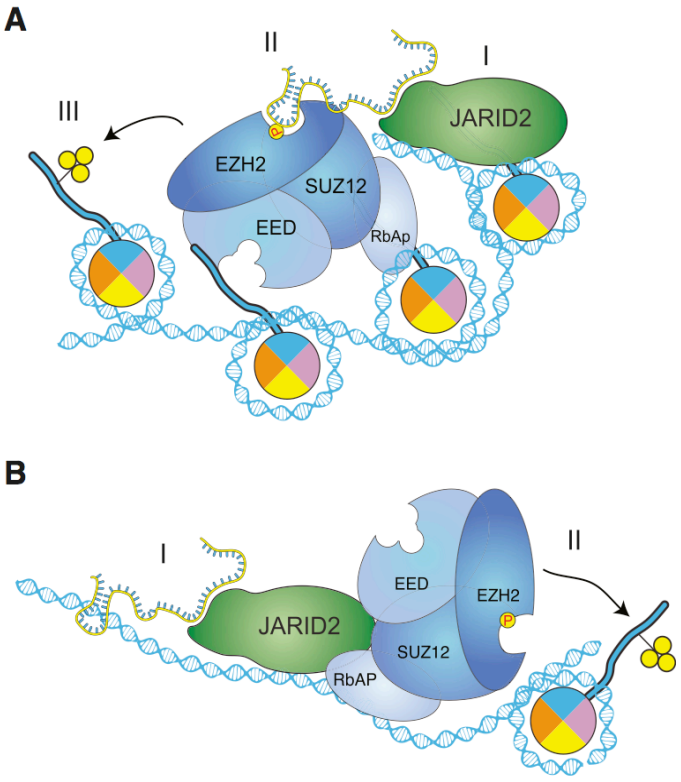


Figura 17. Modelo propuesto para la Interacción de lncARNs, JARID2 y PRC2

(A) En algunos genes blanco, la presencia de JARID2 por sí mismo (I) no es suficiente para el reclutamiento PRC2, que requiere que los lincARNs sirvan como plataforma (II). La presencia de tanto JARID2 y los lincARNs estimulan aún más el reclutamiento y unión de PRC2 en la cromatina, lo que resulta en un aumento de la marca H3K27me3 (III).

(B) En algunos casos, los lincARNs podrían contribuir al reclutamiento inicial de JARID2 a la cromatina (I). Debido a que JARID2 también se une a PRC2 a través de interacciones proteína-proteína, esto se traduce en un mayor reclutamiento de PRC2 y deposición de la marca H3K27me3 (II). (Imagen tomada de Kaneko et al. 2014).

En conclusión, demostramos que JARID2, un componente regulador esencial de PRC2 en CME, contiene una región de unión al ARN que interactúa con el lincARN MEG3. Este lincARN, y posiblemente otros, contribuyen al reclutamiento adecuado de PRC2 en genes blanco y probablemente juega un papel importante en la regulación epigenética de la expresión génica que acompaña la transición de células madre hacia diferentes programas de diferenciación.

ARTÍCULOS PUBLICADOS

CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53

Ricardo Saldaña-Meyer, Edgar González-Buendía, Georgina Guerrero, et al.

Genes Dev. 2014 28: 723-734
Access the most recent version at doi:10.1101/gad.236869.113

CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53

Ricardo Saldaña-Meyer,[1,2] Edgar González-Buendía,[1] Georgina Guerrero,[1] Varun Narendra,[2] Roberto Bonasio,[2,3] Félix Recillas-Targa,[1,4] and Danny Reinberg[2,4]

[1]Instituto de Fisiología Celular, Departamento de Genética Molecular, Universidad Nacional Autónoma de México, México City 04510, México; [2]Howard Hughes Medical Institute, Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, New York 10016, USA

The multifunctional CCCTC-binding factor (CTCF) protein exhibits a broad range of functions, including that of insulator and higher-order chromatin organizer. We found that CTCF comprises a previously unrecognized region that is necessary and sufficient to bind RNA (RNA-binding region [RBR]) and is distinct from its DNA-binding domain. Depletion of cellular CTCF led to a decrease in not only levels of p53 mRNA, as expected, but also those of Wrap53 RNA, an antisense transcript originated from the *p53* locus. PAR-CLIP-seq (photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation [PAR-CLIP] combined with deep sequencing) analyses indicate that CTCF binds a multitude of transcripts genome-wide as well as to Wrap53 RNA. Apart from its established role at the *p53* promoter, CTCF regulates *p53* expression through its physical interaction with Wrap53 RNA. Cells harboring a CTCF mutant in its RBR exhibit a defective *p53* response to DNA damage. Moreover, the RBR facilitates CTCF multimerization in an RNA-dependent manner, which may bear directly on its role in establishing higher-order chromatin structures in vivo.

[*Keywords*: CTCF; RNA binding; Wrap53; p53; multimerization]

Supplemental material is available for this article.

Received December 20, 2013; revised version accepted February 26, 2014.

The CCCTC-binding factor (CTCF) is a remarkably versatile, ubiquitous, and highly conserved zinc finger (ZF) protein (Phillips and Corces 2009). It was originally identified after thorough analyses of factors binding to the chicken c-myc gene promoter (Lobanenkov et al. 1986), later purified (Lobanenkov et al. 1990), and described initially as a transcription factor (Filippova et al. 1996). Since then, CTCF has been implicated in diverse cellular processes, including transcriptional regulation, alternative splicing, insulation, imprinting, X-chromosome inactivation, and higher-order chromatin organization (Phillips and Corces 2009; Holwerda and de Laat 2013). Genome-wide association studies have identified a large number of CTCF-binding sites scattered across the genome, and these sites map to promoters, enhancers, and intergenic regions as well as within gene bodies (Martin et al. 2011; Chen et al. 2012). The associations of CTCF with its DNA target sites as well as with its interacting proteins are thought to contribute to CTCF multifunctionality (Ohlsson and Renkawitz 2001).

Among the various genomic CTCF target sites is the gene encoding the tumor suppressor protein p53. p53 is a sequence-specific transcription factor essential in the cellular response to DNA damage and other types of cellular stress (Vousden and Prives 2009). Contingent on the level of DNA damage, p53 can initiate signaling pathways toward cell cycle arrest, senescence, or apoptosis to avoid oncogenic transformation (Vousden and Prives 2009). Expression of the *p53* gene is regulated by the interplay of a number of transcription factors, microRNAs (miRNAs), the ZF protein CTCF, and its natural antisense transcript, Wrap53 (Mahmoudi et al. 2009; Saldaña-Meyer and Recillas-Targa 2011). *p53* transcription levels are maintained through CTCF binding to the *p53* gene promoter region, allowing for an open chroma-

[3]Present address: Department of Cell and Developmental Biology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA.
[4]Corresponding authors
E-mail danny.reinberg@nyumc.org
E-mail frecilla@correo.ifc.unam.mx
Article is online at http://www.genesdev.org/cgi/doi/10.1101/gad.236869.113.

tin conformation (Soto-Reyes and Recillas-Targa 2010), and upon DNA damage, Wrap53 RNA is essential for induction of *p53* gene expression (Mahmoudi et al. 2009). Wrap53 RNA is an antisense transcript originated from the *p53* locus that positively regulates p53 mRNA levels and, upon DNA damage, is essential for induction of *p53* gene expression (Mahmoudi et al. 2010). The *Wrap53* gene lies upstream of the *p53* gene on the opposite strand and comprises three different transcriptional start sites (TSSs), termed α, β, and γ (see Fig. 1A). Only transcripts originated from exon 1α that directly overlap the first exon of *p53* were found to positively regulate *p53* (Mahmoudi et al. 2009). *Wrap53* can also be translated into protein that was found to be an essential component for Cajal body maintenance (Mahmoudi et al. 2010). Of note, as opposed to Wrap53 RNA, its protein has not been implicated in *p53* regulation.

Here, we report that apart from its established DNA-binding activity, CTCF also has an intrinsic ability to bind RNA. We mapped a domain within CTCF that is necessary and sufficient to bind RNA and is distinct from that required for DNA binding. Depletion of cellular CTCF led to a decrease in not only levels of p53 mRNA, as expected, but also those of Wrap53. With this basis, we addressed the relationship between CTCF and Wrap53 in regulating human *p53* gene expression as well as the role of RNA in CTCF multimerization.

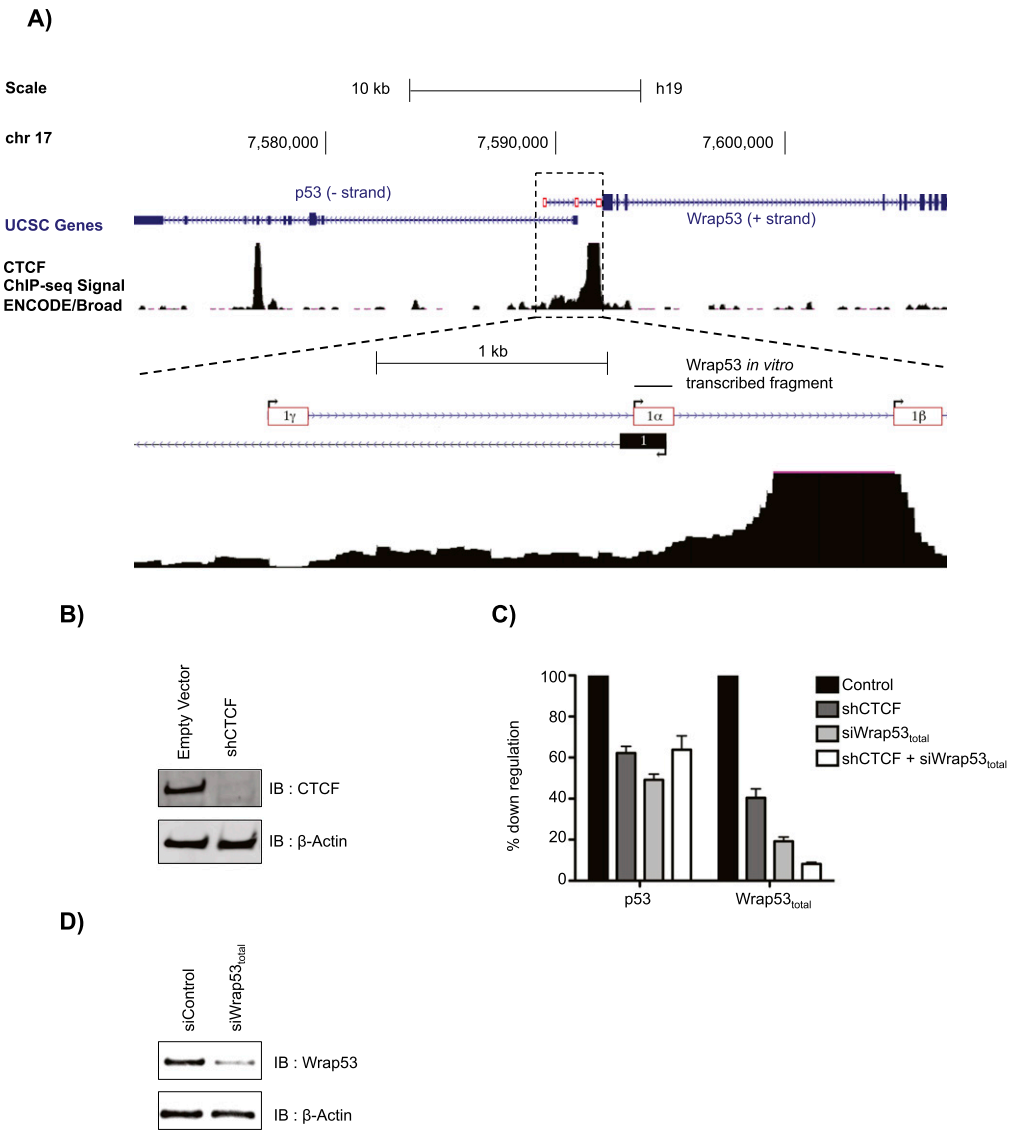


Figure 1. CTCF is required for regulation of the *p53/Wrap53* locus. (*A*) Genomic organization of the *p53/Wrap53* locus. (*B*) Immunoblot after shRNA-mediated knockdown of CTCF or with empty vector (control) in U2OS cells. β-Actin served as loading control. (*C*) RNA was quantified by RT-qPCR after knockdown, as indicated, and normalized to GAPDH levels and is shown as percentage down-regulation. Bars indicate the mean of three biological replicates + SEM. (*D*) As in *B*, but with siRNA-mediated knockdown of Wrap53 or siControl.

Results

CTCF is required for regulation of the p53/Wrap53 locus

The human *p53* gene is mutated in ~50% of tumors (National Cancer Institute Surveillance Epidemiology and End Results Program, http://seer.cancer.gov). However, a large number of tumors carry wild-type *p53*, suggesting that its expression can be disrupted by other mechanisms. Since the CTCF-binding site that lies in the *p53* gene promoter corresponds to the first intron of *Wrap53* on the opposite strand (Fig. 1A), we hypothesized that CTCF regulates the transcription of both *p53* and *Wrap53* genes through binding to its regulatory element and, possibly, to the antisense transcript Wrap53. Upon stable expression of a shRNA against CTCF, resulting in reduced CTCF levels (Fig. 1B), transcripts for p53 and all Wrap53 isoforms ($Wrap53_{total}$) were decreased, based on normalization with the constitutively expressed GAPDH gene (Fig. 1C). Similar decreases were observed with an inducible shRNA targeting the 3′ untranslated region (UTR) of CTCF after 72 h of induction (data not shown). Interestingly, *p53* expression was also down-regulated when the antisense transcript $Wrap53_{total}$ was depleted using siRNAs (Fig. 1C,D; Mahmoudi et al. 2009), and depletion of both CTCF and $Wrap53_{total}$ had similar effects (Fig. 1C), suggesting that transcriptional regulation of sense and antisense RNAs at this locus is tightly coregulated by CTCF.

CTCF binds a variety of RNAs in vivo, including Wrap53

As depletion of either CTCF or $Wrap53_{total}$ gave rise to similar effects, we hypothesized that these two species may be not only functionally but also physically related. To determine whether Wrap53 RNA is associated with CTCF in vivo, we performed RNA immunoprecipitation (RIP) followed by quantitative PCR (qPCR). Consistent with our hypothesis, both p53 and the $Wrap53_{total}$ transcripts were enriched in CTCF-specific RIPs compared with IgG (Supplemental Fig. S1A,B). Moreover, $Wrap53_{total}$ RNA bound specifically to CTCF, as compared with the case of SCML2 (Supplemental Fig. S1A,B), an unrelated protein that also binds RNA in the nucleus (R Bonasio, E Lecona, and D Reinberg, unpub.). Although we observed some enrichment for the $Wrap53_{total}$ RNA using SCML2 as a control, native RIP can yield false positives due to RNA–protein reassociations after cell lysis (Mili and Steitz 2004; Riley et al. 2012). Given these limitations, we reassessed specific and direct binding of Wrap53 to CTCF using an unbiased genome-wide approach involving the PAR-CLIP (photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation) assay, which makes use of a photoactivatable nucleoside analog, 4-thiouridine (4-SU), to selectively and irreversibly cross-link protein to RNA in living cells (Hafner et al. 2008). To avoid possible contamination from other RNA-binding proteins with molecular weights similar to that of CTCF, we performed immunoprecipitations with a buffer containing a zwitterionic detergent, 2% lauryldimethylbetaine, that preserves antibody reactivity while significantly decreasing coprecipitation of CTCF protein partners, such as the DEAD-box RNA helicase p68 (DDX5) and the SA1 subunit of the cohesin complex (Fig. 2A; Supplemental Fig. S2C; Yao et al. 2010; Xiao et al. 2011). We concluded that the radioactive signal obtained was due to RNA specifically cross-linked to CTCF, since it was dependent on the presence of 4-SU in the culture medium (Fig. 2A), was decreased upon CTCF knockdown (Fig. 2B), and was sensitive to RNase treatment (Supplemental Fig. S1C,D). Of note, the preparation of cells for PAR-CLIP seems to result in the partial degradation of full-length (FL) CTCF given the presence of lower-molecular-weight species that are detected with an antibody against CTCF and are diminished upon CTCF knockdown (Supplemental Fig. S1D and Fig. 2B, respectively).

To identify the RNAs bound to CTCF in vivo, the ^{32}P-labeled species corresponding to FL CTCF were excised from the PAR-CLIP membrane, with the cross-linked RNA being eluted and prepared to construct libraries for deep sequencing. We obtained ~1.2×10^6 unique PAR-CLIP tags in each biological replicate. Although *p53* is expressed at low levels and Wrap53 RNA is much less abundant than p53 mRNA, we obtained PAR-CLIP tags from both *Wrap53* α and β TSSs, confirming our initial observations of CTCF interaction with Wrap53 RNA (Fig. 2C). When the average relative enrichment of p53 and $Wrap53_{total}$ transcripts, normalized to those of GAPDH obtained by qPCR, was compared with the PAR-CLIP tags obtained from the two biological replicates, we observed a clear enrichment for $Wrap53_{total}$ over p53 and GAPDH (Fig. 2D). We then inspected the genome-wide distribution of the PAR-CLIP tags and found that ~34% were located at repetitive regions, 30% were located at exons, 28% were located at introns, and 8% were located at intergenic regions (Fig. 2E). There was no significant enrichment for either exonic or intronic reads, indicating that CTCF binds to both nascent and mature transcripts. Using an arbitrary cutoff of five unique PAR-CLIP tags, the data set comprised 17,201 genes, of which 71% are protein-coding, 12% are pseudogenes, 8.3% are antisense, 5% are lincRNAs, and 3.7% accounted for other types (Supplemental Table S1). Thus, $Wrap53_{total}$ RNA is recovered in both native RIP and PAR-CLIP using an antibody against CTCF. The latter analysis provides the first evidence of CTCF binding to transcripts genome-wide.

Identification and mapping of the RNA-binding region (RBR) of CTCF

Given the data collected on RNA binding to CTCF in vivo, we next sought to identify the domain of CTCF required for such interaction. Candidates of CTCF, either FL or deletion mutants, were fused to the human glutathione S-transferase domain (GST), purified from *Escherichia coli*, and then incubated with RNA transcribed in vitro from a 5′-terminal *Wrap53* template spanning nucleotides 1–167 of exon 1α (Fig. 1). RNA-binding activity was evident

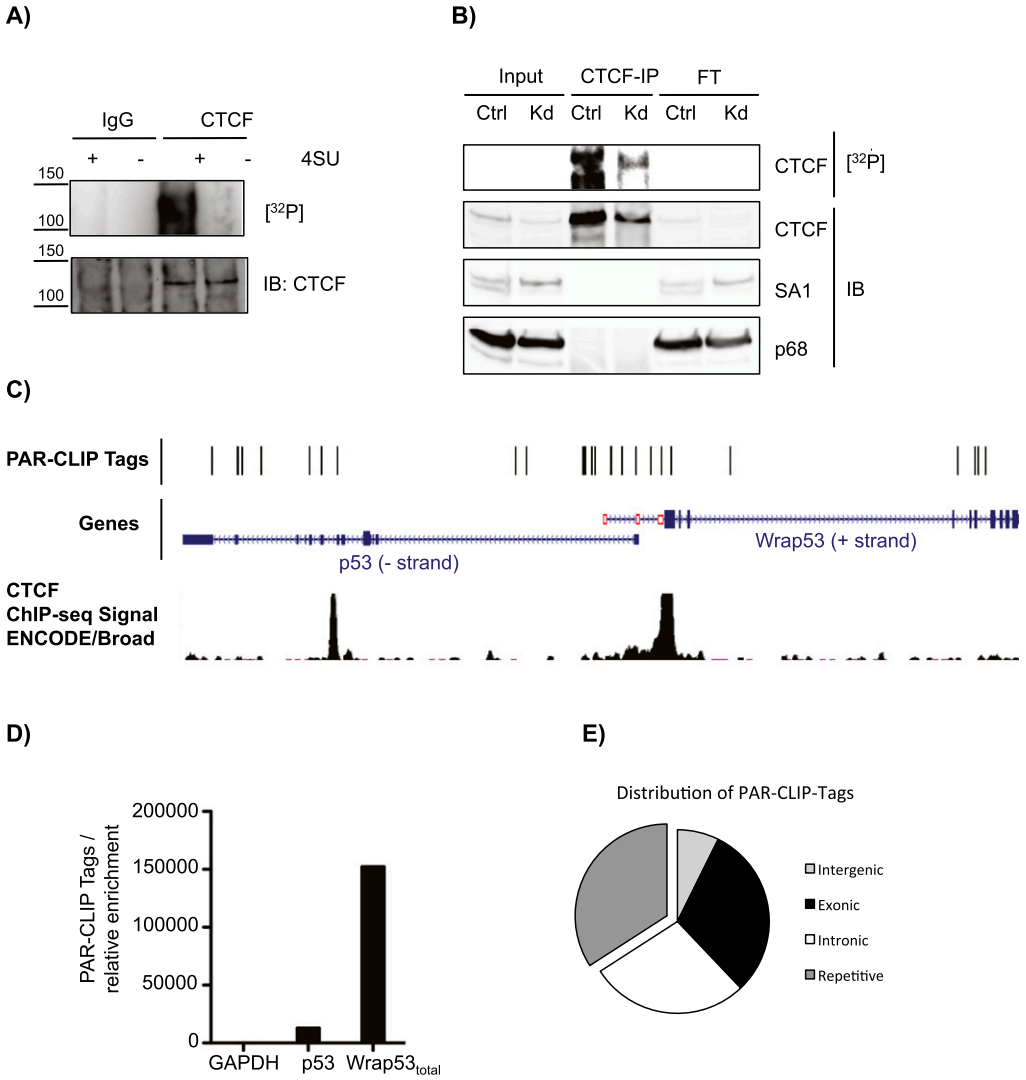


Figure 2. CTCF binds a variety of RNAs in vivo, including Wrap53. (*A*) PAR-CLIP (*top*) and immunoblot for CTCF (*bottom*) in U2OS cells with (+) or without (−) 4-SU incorporation. Whole nuclear lysates were immunoprecipitated with a CTCF antibody or IgG. (*B*) PAR-CLIP (*top*) and immunoblots (*bottom*) for CTCF, DDX5, or SA1 using the same membrane. Immunoprecipitation was performed using an antibody against CTCF in whole nuclear extracts from cells stably transfected with a control shRNA (Ctrl) or shRNA against CTCF (Kd). (*C*) Genome browser view showing PAR-CLIP tags recovered after deep sequencing over the *p53/Wrap53* locus. The CTCF ChIP-seq (chromatin immunoprecipitation [ChIP] combined with deep sequencing) signal from ENCODE is also shown. (*D*) Graphical representation of normalization of PAR-CLIP tags versus the average relative enrichment obtained by qPCR. (*E*) Pie chart showing the genome-wide distribution of PAR-CLIP tags.

from the ZF domain but not from the terminal regions alone (Fig. 3A,B). More detailed mapping demonstrated that ZFs 9–11 exhibit a higher affinity for the RNA than the remaining ZFs (Fig. 3C,D). Finally, systematic deletion of ZFs 9–11 revealed that the optimal region for RNA binding spans amino acids 520–727 and includes ZFs 10–11 and the C-terminal region, henceforth termed the RBR (Fig. 3C,D). While the C terminus alone (576–727 amino acids) was ineffectual in binding, the deletion mutant (520–615 amino acids) lacking most of the C terminus but containing ZFs 10–11 was also ineffectual (Fig. 3C,D), suggesting that, although not sufficient, the C terminus is required for RNA binding.

We next tested the RBR for DNA versus RNA-binding preferences using electrophoretic mobility shift assays (EMSA). Binding of GST-RBR to a probe containing the first *Wrap53* exon was compared in the presence of increasing amounts of RNA or DNA (Fig. 3E). Each of the competing DNAs comprised a distinct CTCF-binding site, as demonstrated previously in EMSA: the CTCF-binding site located in the *p53* gene promoter (Soto-Reyes and Recillas-Targa 2010), the CTCF-binding site in the human *Rb* gene promoter (De La Rosa-Velázquez et al. 2007), and the chicken cHS4 β-globin insulator FII (Valadez-Graham et al. 2004). The RNA competitor was derived by in vitro transcription from the *p53* promoter region described

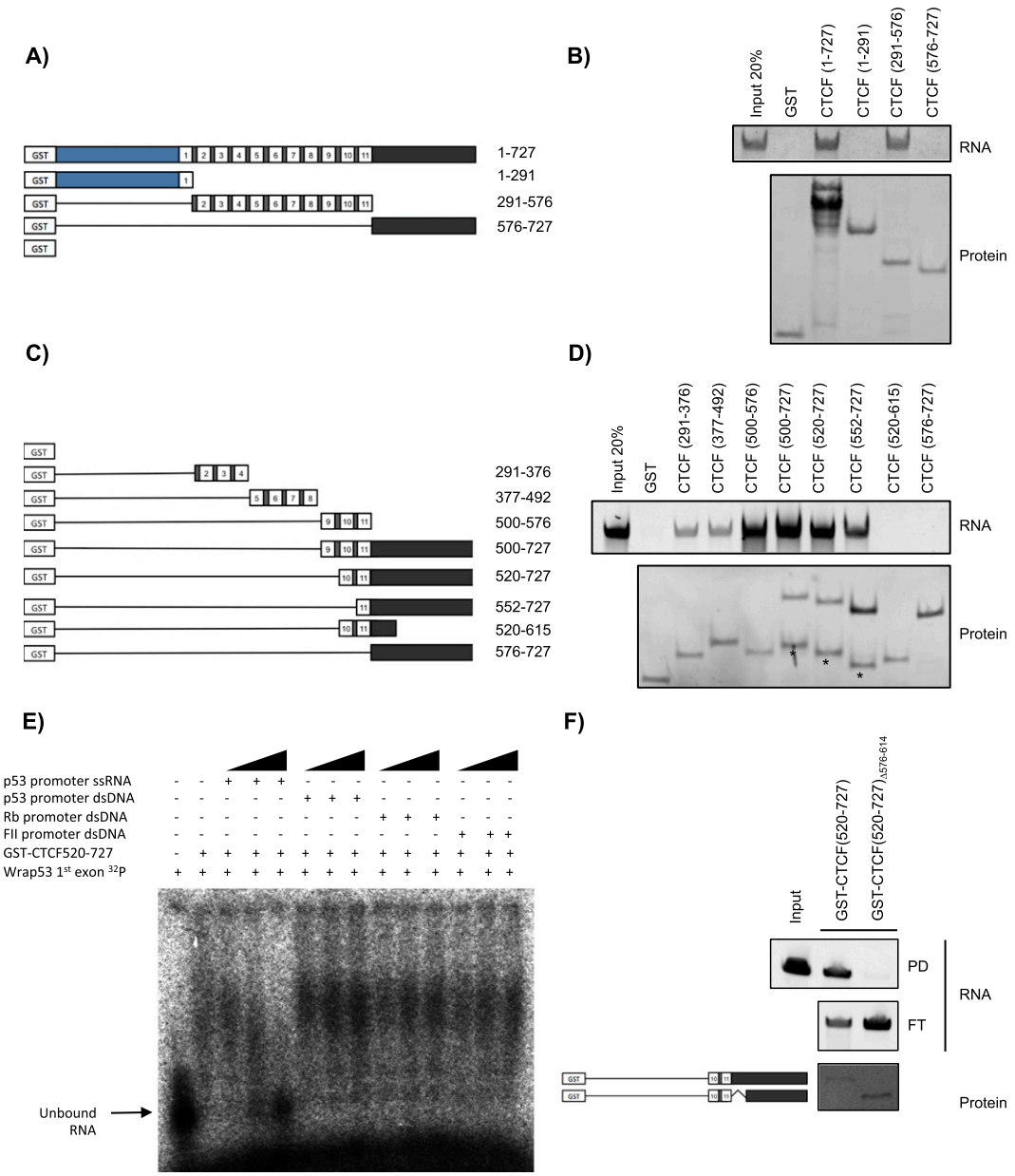


Figure 3. Identification and characterization of the CTCF RBR. (*A*,*C*) Schematic representations of GST-tagged CTCF, either FL or deletion mutant versions, analyzed as shown in *B* and *D*, respectively. (*B*,*D*) RNA-binding assay showing recovered RNA stained with SYBR Gold (*top* panel) and recovered protein stained with SYPRO Red (*bottom* panel). In *D*, the asterisk demarcates degraded protein. (*E*) EMSA using GST-$CTCF_{520–727}$ and a radioactive 5′ probe spanning nucleotides 1–167 of Wrap53 mRNA incubated with increasing amounts of unlabeled DNA or RNA, as indicated. (*F*) As in *B* and *D* but with GST-$CTCF_{520–727}$ or GST-$CTCF(520–727)_{\Delta 576–614}$. (PD) Pull-down; (FT) flow through.

above. While addition of this p53-derived RNA successfully competed GST-RBR binding to the probe containing the *Wrap53* exon, all of the DNA candidates containing a CTCF-binding site were ineffectual (Fig. 3E). To help pinpoint a specific region of CTCF that is necessary for RNA binding, we took advantage of the RNABindR software that is designed to predict putative RBRs (Terribilini et al. 2007). The software predicted with high sensitivity and specificity that the region of the C terminus downstream from the last ZF could bind RNA. Indeed, deletion of residues 576–614 within the predicted region drastically decreased binding to RNA, as demonstrated by RNA-binding assays (Fig. 3F). These data suggest that CTCF contains a previously unrecognized RBR that displays strong binding preferences for p53 and Wrap53 RNAs in vitro.

An internal deletion within the RBR disturbs the DNA damage response

To address the functional relevance of Wrap53 RNA–CTCF interactions in vivo, we generated an inducible cell system in which endogenous CTCF is depleted by shRNA

knockdown and the CTCF-deficient cells are rescued with HA-tagged versions of CTCF, either wild type (HA-$CTCF_{wt}$) or mutant in RNA binding (HA-$CTCF_{\Delta 576–614}$) (Supplemental Fig. S2). Of note, a putative nuclear localization signal is found within 576–614 amino acids of CTCF, specifically in amino acids 590–603 (Klenova et al. 2001). Nonetheless, the Δ576–614 deletion did not affect the ability of CTCF to localize to the nucleus in vivo or its general association to chromatin or to the known interacting proteins DDX5 and the SA1 subunit of the cohesin complex (Supplemental Fig. S2A–C; Yao et al. 2010; Xiao et al. 2011). Furthermore, recombinant forms of both HA-$CTCF_{wt}$ and HA-$CTCF_{\Delta 576–614}$ were able to evict the residual endogenous CTCF from the chromatin fraction after inducing knockdown for 72 h with doxycycline treatment (Supplemental Fig. S2D).

Mahmoudi et al. (2009) reported that both Wrap53 and the p53 transcripts were induced upon DNA damage. They also suggested that Wrap53 not only maintains p53 mRNA levels but also plays a role in stabilizing p53 mRNA in response to DNA damage (Mahmoudi et al. 2009). To ascertain whether a complex between CTCF and Wrap53 RNA regulates *p53* expression, cell lines harboring HA-$CTCF_{wt}$ or HA-$CTCF_{\Delta 576–614}$ were induced with doxycycline for 56 h and then treated for an additional 16 h with doxycycline along with the DNA-damaging drug methyl methanesulfonate (MMS). Consistent with the previous study (Mahmoudi et al. 2009), cells rescued with HA-$CTCF_{wt}$ exhibited up-regulation of both p53 and α and β isoforms of Wrap53 at the mRNA level as well at the protein level after DNA damage treatment (Fig. 4A,B, respectively). Furthermore, genes involved in cell cycle arrest and apoptosis via the p53 pathway (p21 and PUMA, respectively) were also up-regulated (Fig. 4A). In contrast, cells induced for expression of the mutant, RNA-binding-defective HA-$CTCF_{\Delta 576–614}$ failed to rescue these mRNA and protein levels upon DNA damage (Fig. 4). Of note, knockdown of $Wrap53_{total}$ did not affect the recruitment of endogenous CTCF to the *p53* promoter (Supplemental Fig. S3A). Furthermore, when used in the rescue experiment, HA-$CTCF_{\Delta 576–614}$ did not exhibit a defect in binding to the *p53* promoter (Supplemental Fig. S3B) or in the endogenous mRNA levels of p53 or Wrap53 (Supplemental Fig. S3C) but did show a decreased coprecipitation with $Wrap53_{total}$ (Supplemental Fig. S3D,E), highlighting the relevancy of its defect during conditions of DNA damage. Thus, a CTCF mutant defective in RNA binding cannot restore features of the DNA damage response in CTCF-depleted cells, likely due to its failure to bind Wrap53 RNA, thereby preventing transcriptional induction of the *p53* locus.

RNA aids in the formation of CTCF multimers

We speculated that RNA–CTCF interactions might participate in the formation of higher-order multimers of CTCF by bridging the interaction between monomers. CTCF has many binding partners (Zlatanova and Caiafa 2009), including other CTCF molecules (Valadez-Graham et al. 2004; Yusufzai et al. 2004), although the molecular

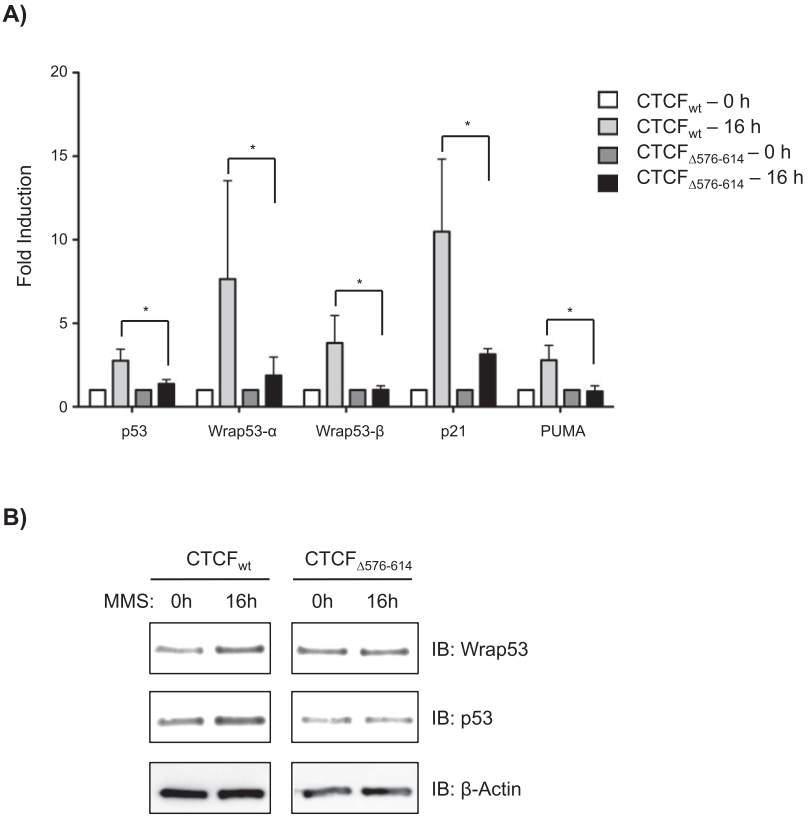


Figure 4. Deletion of the RBR within CTCF disturbs the DNA damage response. Cell lines containing HA-$CTCF_{wt}$ or HA-$CTCF_{\Delta 576–614}$ were induced with doxycycline for 56 h and then treated for an additional 16 h with doxycycline with or without the DNA-damaging drug MMS, as indicated. (*A*) RT-qPCR for the transcripts indicated, shown as fold induction and normalized to GAPDH levels. Bars indicate the mean of three biological replicates + SEM. (*) $P < 0.05$ by Mann-Whitney *U*-test. (*B*) Immunoblot of p53 and Wrap53, with β-actin as loading control.

basis underlying such dimerization/multimerization has not been reported. Because modulation of CTCF multimerization through RNA might have major implications in chromatin looping and interaction with diverse cofactors, we tested the ability of CTCF to dimerize or multimerize in the presence of different nucleases.

To this end, we took advantage of distinctly tagged versions of CTCF. We first tested the ability of FL proteins to interact with each other. Recombinant GST-CTCF$_{wt}$ purified from *E. coli* was added to nuclear extracts derived from the expressing HA-CTCF$_{wt}$ cell line, as a function of DNase or RNase treatment, and then analyzed in pull-down assays. Indeed, GST-CTCF$_{wt}$ could coprecipitate HA-CTCF$_{wt}$ (Fig. 5A). However, the addition of RNase during the incubation period led to an ~50% decrease in the amount of HA-CTCF$_{wt}$ coprecipitated, compared with the untreated case; this decrease was not observed with DNase addition (Fig. 5A). The RBR is sufficient for this dimerization, since incubation with GST-RBR also gave rise to efficient coprecipitation of HA-CTCF$_{wt}$, while the addition of RNase also thwarted HA-CTCF$_{wt}$ recovery in this case, and DNase treatment was again ineffectual (Fig. 5B). Importantly, the ability to interact with RNA was required for the observed CTCF dimerization/multimerization, as evidenced by the absence of such interaction in the case of the mutant, RNA-binding-defective GST-RBR$_{\Delta 576–614}$ incubated with either HA-CTCF$_{wt}$ or HA-CTCF$_{\Delta 576–614}$, (Fig. 5C,D, respectively).

To corroborate this observation, we removed the GST tag from the CTCF$_{wt}$ and performed size exclusion chromatography in the presence or absence of 167 base pairs (bp) of Wrap53 RNA. The elution profile of the CTCF sample alone peaked at an apparent mass of 265 kDa, which would point to the formation of a trimer (Fig. 6A). Interestingly, the elution profile in the presence of Wrap53 RNA suggested the formation of a complex >2 MDa (Fig. 6A). Lending further support to the conclusion that RNA helps CTCF multimerization via its RBR, the elution profile of the RBR fragment (~22 kDa) alone peaked at ~60 kDa but was shifted to 1 MDa in the presence of Wrap53 RNA (Fig. 6B). In the latter case, the RBR and RNA profiles strongly indicated their coelution, with all of the RNA being found in complex with RBR, migrating at a significantly higher molecular weight than that predicted for RBR (Fig. 6B) or RNA alone (data not shown). Taken together, these results suggest that RNA aids in the formation of multimers of CTCF through its RBR.

Discussion

Our findings support the role of RNA binding as an important, additional regulatory function among the spec-

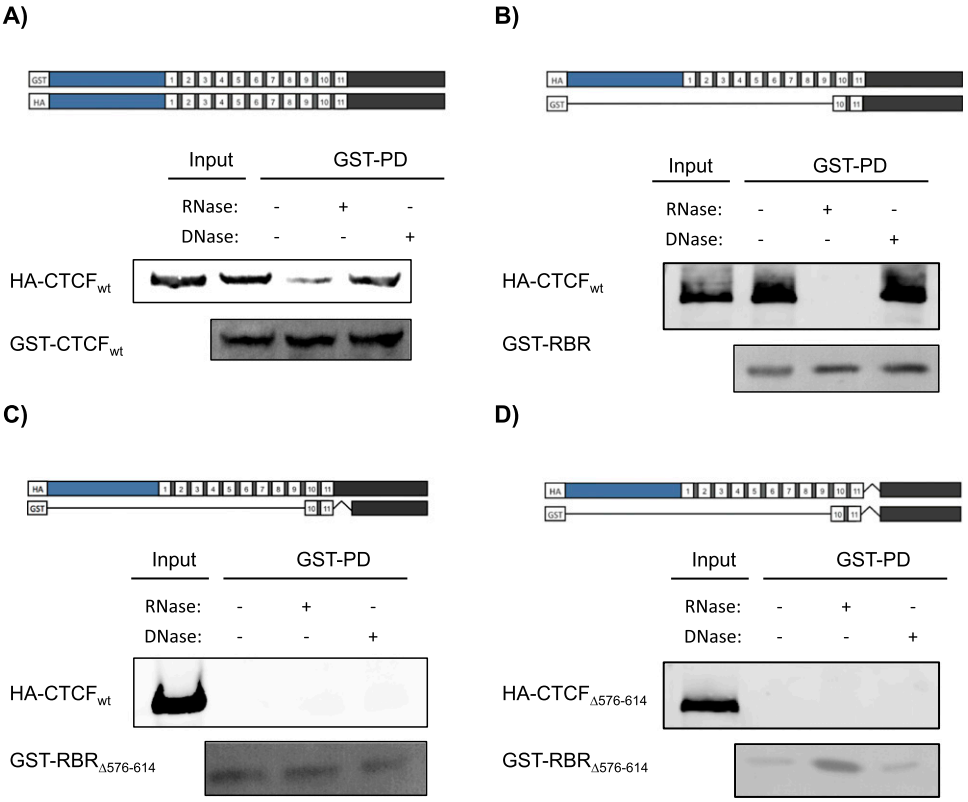


Figure 5. RNA facilitates CTCF dimerization/multimerization. (*A–D*) Shown at the *top* of each panel are schematic representations of CTCF, either wild type (WT) or mutant (CTCF$_{\Delta 576–614}$), or its RBR, either wild type or mutant (RBR$_{\Delta 576–614}$), tagged with HA or GST, as indicated. *Below* are the respective results of GST pull-down experiments performed using nuclear extracts of U2OS cells expressing either HA-CTCF$_{wt}$ or HA-CTCF$_{\Delta 576–614}$, in the presence or absence of nucleases, as indicated. GST-CTCF proteins were stained with Ponceau Red, and HA-CTCF proteins were detected with an antibody against HA.

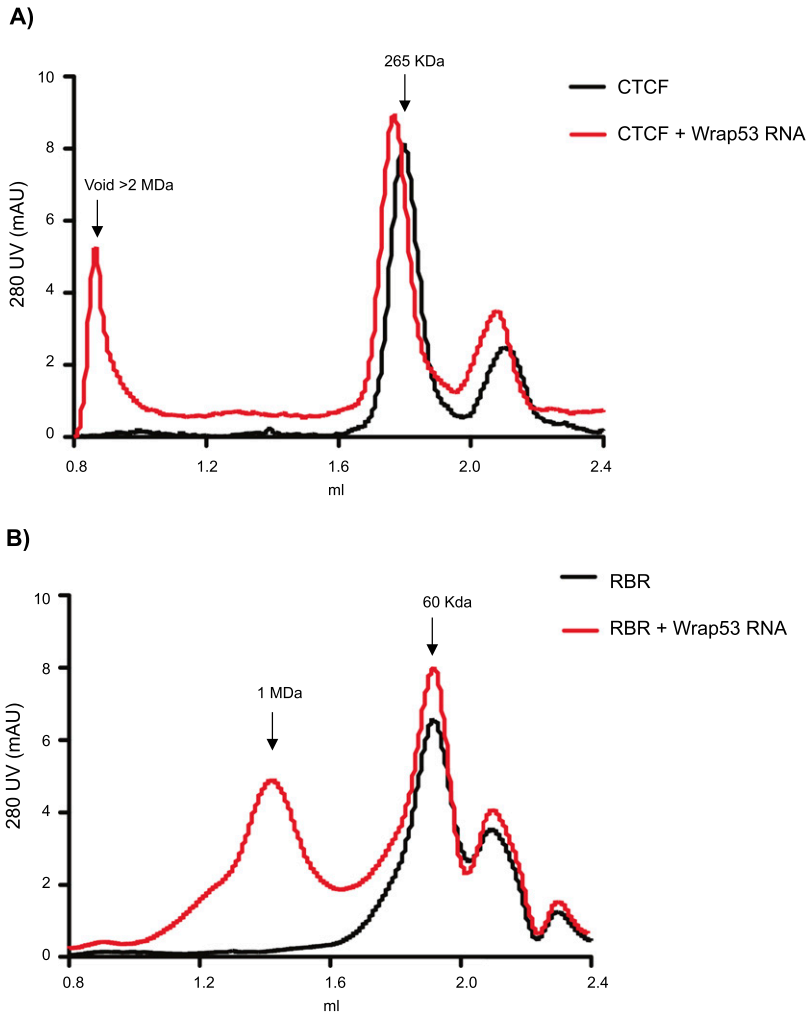


Figure 6. CTCF–Wrap53 RNA forms large-molecular-weight complexes. Purified CTCF, either FL (*A*) or the RBR (*B*) domain, was fractionated in a Superose 6 sizing column in the presence or absence of Wrap53 RNA.

trum of those already recognized for CTCF. As shown here, CTCF binding to Wrap53 is integrally related to the appropriate *p53* transcriptional response to DNA damage. We also presented the first genome-wide array of CTCF-interacting transcripts that may provide insights into the scope of CTCF targets as well as the mechanistic basis for CTCF action. Our findings suggest a general function for such CTCF–RNA interaction in mediating CTCF multimerization that may bear directly on the role of CTCF in chromatin looping.

We previously determined that loss of CTCF binding to target DNA sequences promotes epigenetic silencing of *p53* gene expression through enrichment of repressive chromatin marks in its regulatory region (Soto-Reyes and Recillas-Targa 2010). Knockdown of Wrap53 was reported by another group as leading to down-regulation of *p53* (Mahmoudi et al. 2009). These initial observations provided insights into the regulation of the *p53* locus. Here, we provide an additional means by which CTCF participates in the transcriptional regulation of the human *p53* gene. Our finding that CTCF knockdown down-regulates both p53 and $Wrap53_{total}$ mRNA levels points to an interdependent mechanism involving CTCF and Wrap53 RNA. Our hypothesis that CTCF could make direct contacts with Wrap53 RNA was supported initially by their coimmunoprecipitation and then confirmed through in vitro characterizations. Indeed, as shown here, CTCF comprises a previously unrecognized RNA-binding activity that encompasses its ZFs 10–11 and C-terminal segment independent of its well-established DNA-binding activity (Fig. 3). Nakahashi et al. (2013) reported recently that CTCF binds primarily to its DNA sequence motif with ZFs 4–7 and that its remaining ZFs have a stabilizing role. They also described that ZFs 9–11 can associate with a second conserved, upstream DNA motif at ~15% of its sites (Nakahashi et al. 2013). This study complements our observations, as ~85% of the CTCF proteins bound to DNA would potentially be free to make additional contacts with RNA through their RBRs as well as any unbound CTCF.

The CTCF ZF domain is almost completely conserved among mice, chickens, and humans (Ohlsson and Renkawitz 2001) and comprises 10 C_2H_2-type ZFs and an 11th C_2HC-type ZF. Results presented here show that ZFs 10–11 specifically bind RNA. Besides CTCF, there are a number of other ZF proteins from different organisms that

bind both DNA and RNA (Brown 2005). TFIIIA in mammals is a prime example, providing insight into how proteins with multiple C_2H_2-type ZFs can specifically bind both DNA and RNA. Similar to our findings with CTCF, TFIIIA uses different ZFs to achieve specificity between DNA and RNA, requiring ZFs 1–3 to bind DNA and its central ZFs 4–6 to bind RNA (Searles et al. 2000). TFIIIA in complex with 5S RNA was also the first ZF–RNA structure resolved, revealing that ZF binding to RNA is different from its well-studied binding to DNA (Lu et al. 2003). It is important to emphasize that these ZFs can bind both DNA and RNA and both single-stranded and double-stranded versions but with distinctive affinities (Brown 2005). Recognition of ssRNA by C_2H_2 ZFs is made through aromatic side chains intercalating between appropriately spaced dinucleotide bases, whereas recognition of dsDNA is through the α helix forming hydrogen bonds with the major groove bases (Brown 2005). Unlike TFIIIA, which shows high specificity for 5S RNA, our results show that, at least in vitro, CTCF recognizes RNA with little sequence specificity. Instead of dependence only on sequence specificity, CTCF–RNA contacts may arise in a context-dependent manner, perhaps favoring intricate secondary or tertiary structures that are currently not feasible to reproduce in vitro.

The results discussed above argue for CTCF making direct contacts with RNA, but the general biological relevance of this RNA-binding activity remains unclear. In the context of *p53* transcriptional regulation, our evidence supports CTCF-mediated regulation of *p53* and *Wrap53* transcription during the DNA damage response through two different levels: (1) binding to the *p53* promoter region (Soto-Reyes and Recillas-Targa 2010) and (2) interacting with Wrap53 RNA. Another specific function for CTCF was reported recently in which CTCF binds to the noncoding RNA Jpx in the context of X-chromosome inactivation (Sun et al. 2013). This group proposed that CTCF could be evicted from its binding site by Jpx overexpression. However, this interplay was not a general feature of CTCF, being restricted to an allele-specific site in the Xist locus and irrelevant to the other CTCF-binding sites tested. To date, only locus-specific functions for RNA–CTCF interactions have been supported such that the general biological significance for these interactions remains elusive.

The first role of CTCF described was that of a transcription factor (Filippova et al. 1996), and, later, a number of studies described its function as an insulator (Barkess and West 2012). Nonetheless, accumulating evidence suggests that its insulation function may occur in only a few cases, probably as a consequence of its primary role as a chromatin organizer (Handoko et al. 2011; Sanyal et al. 2012; Phillips-Cremins et al. 2013). CTCF is implicated in a myriad of functions in gene regulation, but perhaps its most crucial role is in the formation of chromatin loops (Phillips and Corces 2009). In this regard, our data support the hypothesis that RNA facilitates CTCF multimerization, as evidenced by the requirement of RNA for CTCF–CTCF interactions by pull-downs (Fig. 5) and the formation of large CTCF–Wrap53 RNA complexes using size exclusion chromatography (Fig. 6). This RNA dependence for protein multimerization has been observed in other cases as well. For example, the human cytidine deaminase APOBEC3G (A3G) was shown to oligomerize only in the presence of RNA (Huthoff et al. 2009), and, similarly, the p53 protein can interact with p68/p72 and the Drosha complex in an RNA-dependent manner (Suzuki et al. 2009). Moreover, the physical interaction of CTCF with the DEAD-box RNA-binding protein p68 (DDX5) was also shown to be dependent on its associated noncoding RNA, SRA (Yao et al. 2010). Interestingly, the investigators proposed that other RNAs might partially substitute for SRA in this context, hinting at context-dependent interactions as opposed to sequence specificity. Nonetheless, our results show that the FL CTCF treated with RNase decreased the recovery of additional CTCF molecules by only ~50% (Fig. 5A). This observation could be due to incomplete degradation of the RNA or contributions to CTCF oligomerization by CTCF-interacting proteins.

We can conclude that the ubiquitous architectural protein CTCF binds a large array of different RNA species genome-wide. This seems to constitute a general feature for the "master weaver" of the genome. However, currently, we have insight into only two functionally different and locus-specific roles for CTCF in its interaction with Jpx (Sun et al. 2013) or Wrap53 RNA (this study). We provided some insight into the possible formation of RNA-dependent CTCF multimers that would influence nuclear organization by regulating chromatin loops, and this in turn begets new sets of questions. Are some chromatin loops formed, maintained, and regulated by CTCF–RNA interactions? Do other architectural proteins interact via RNA? What are the regulatory cues for CTCF in RNA binding? Does DNA binding precede, being required for RNA binding, or vice versa? Is RNA methylation a means of regulating its interaction with CTCF? What is the molecular basis to achieve both locus-specific and general functions? All of these are interesting questions and warrant further investigation. Finally, the biochemical characterization of the RBR within CTCF and the genome-wide description of its associated RNAs will facilitate screens for other relevant locus-specific and cell type-specific regulatory functions. The RNA-dependent multimerization of CTCF and its potential in mediating chromatin looping will likely have major implications in future research in nuclear organization and epigenetic regulation.

Materials and methods

Antibodies and oligonucleotide primers

Detailed information for antibodies and oligonucleotides used in this study is reported in Supplemental Tables S3 and S4, respectively.

Protein constructs

DNA encoding FL CTCF or $CTCF_{\Delta 576–614}$ was cloned into the pINTA-N3 system (Kaneko et al. 2014), and expression was activated in a doxycycline-dependent manner by the rtTA trans-

activator contained in the pTRIPZ vector expressing a doxycycline-dependent shRNA against the 3′ UTR of CTCF (Thermo Scientific V3THS_409881). To produce GST-tagged truncated CTCF proteins in *E. coli*, a series of human CTCF cDNAs were PCR-amplified and cloned into pGEX6-P1.

Native RIP

Nuclear extracts were obtained using an established protocol (Dignam et al. 1983) with minor modifications to minimize RNase activity. Briefly, cells were washed with PBS and then with buffer A (10 mM Tris at pH 7.9 at 4°C, 1.5 mM $MgCl_2$, 10 mM KCl, protease inhibitors, phosphatase inhibitors) and lysed in buffer A with 0.2% IGEPAL CA-630 for 5 min on ice. Nuclei were isolated by centrifugation at 2500*g* for 5 min and lysed in buffer C (20 mM Tris at pH 7.9 at 4°C, 25% glycerol, 400 mM NaCl, 1.5 mM $MgCl_2$, 10 mM EDTA, 0.4 U/μL murine RNase inhibitor, protease inhibitors, phosphatase inhibitors) for 30 min at 4°C. Lysates were cleared at 20,000*g* for 30 min. For immunoprecipitation, lysates were diluted to 1 mg/mL in RIP buffer (20 mM Tris at pH 7.9 at 4°C, 200 mM KCl, 0.05% IGEPAL CA-630, 10 mM EDTA), cleared by centrifugation at 20,000*g* for 10 min, and incubated with a predetermined, depleting amount of antibody for 3 h at 4°C. Immunocomplexes were recovered as follows. Five microliters of protein G-coupled Dynabeads (Invitrogen) was added per microgram of antibody used and incubated for 1 h at 4°C. Beads were washed in RIP-W buffer (20 mM Tris at pH 7.9 at 4°C, 200 mM KCl, 0.05% IGEPAL CA-630, 1 mM $MgCl_2$) twice and incubated with 2 U of TURBO DNase (Ambion) in 20 μL of RIP-W buffer for 10 min at room temperature to eliminate potential bridging effects of protein–DNA and RNA–DNA interactions. After two additional washes in RIP-W buffer, RNA was eluted and purified with TRIzol (Invitrogen), and the residual DNA was removed with an additional TURBO DNase treatment.

PAR-CLIP and PAR-CLIP-seq (PAR-CLIP combined with deep sequencing)

PAR-CLIP was performed as described with some modifications (Kaneko et al. 2014). Briefly, U2OS cells were grown under standard conditions and pulsed with 400 μM 4-SU (Sigma) for 16–24 h. After washing the plates with PBS, cells were cross-linked with 400 mJ/cm^2 UVA (365 nm) using a Stratalinker UV cross-linker (Stratagene). Whole nuclear lysates (WNLs) were obtained by fractionating cytoplasm and nuclei by a standard method (Dignam et al. 1983), and nuclei were then incubated for 10 min at 37°C in an appropriate volume of CLIP buffer (20 mM HEPES at pH 7.4, 5 mM EDTA, 150 mM NaCl, 2% lauryldimethylbetaine) supplemented with protease inhibitors, 20 U/mL Turbo DNase (Life technologies), and 200 U/mL murine RNase inhibitor (New England Biolabs). After clearing the lysate by centrifugation, immunoprecipitations were carried out using 200 μg of WNLs, appropriate antibody, and protein G-coupled Dynabeads (Life Technologies) in the same CLIP buffer overnight at 4°C, after which, when required, the extracts were treated with various concentrations of RNase A+T1 cocktail (Ambion) for 5 min at 37°C. Contaminating DNA was removed by treating the beads with Turbo DNase (2 U in 20 μL). Cross-linked RNA was labeled by successive incubation with 5 U of Antarctic phosphatase (New England Biolabs) and 5 U of T4 PNK (New England Biolabs) in the presence of 10 μCi [γ-^{32}P] ATP (PerkinElmer). Labeled material was resolved on 8% Bis-Tris gels, transferred to nitrocellulose membranes, and visualized by autoradiography.

For PAR-CLIP-seq experiments, 1 mg of WNLs was employed. 3′-blocked DNA adapter (100 pmol/μL) was ligated to the RNA after dephosphorylation and before 5′-^{32}P end-labeling by incubation of the beads with T4 RNA ligase 1 (New England Biolabs) for 1 h at 25°C. Labeled material was resolved on 8% Bis-Tris gels, transferred to nitrocellulose membranes, and visualized by autoradiography. The band of interest was excised, and RNA was eluted from the membrane by treatment with 4 mg/mL proteinase K (Roche) for 30 min at 37°C and then with proteinase K in the presence of 3.5 M urea for 30 min at 55°C. After phenol/chloroform extraction, custom-designed 5′ RNA adapters were ligated, the products were size-selected on polyacrylamide gels, and libraries were amplified and sequenced on an Illumina HiSeq 2000 sequencing system.

For the initial PAR-CLIP-seq mapping, we clipped adapter sequences from PAR-CLIP reads and kept those >17 nt. The resulting reads were collapsed to remove duplicate sequences with the FASTX toolkit and then mapped with Bowtie –v2 –m40 –best –strata to the hg19 assembly. Two separate replicates were pooled and then processed with an arbitrary cutoff of five unique reads. The repetitive elements listed by RepeatMasker (as downloaded from the University of California at Santa Cruz Genome Browser Web site on November 1, 2013) were discarded, leaving 17,201 genes. The list of these genes can be found in Supplemental Table S1, and additional sequencing information can be found in Supplemental Table S2. Annotation-based analysis was performed using the ENSEMBL database.

Sequencing data

All sequencing data have been deposited to the National Center for Biotechnology Information Gene Expression Omnibus as SuperSeries GSE53554

In vitro binding assays

RNA probes were synthesized from template DNA. Binding assays were carried out in the presence of GST-tagged versions of FL CTCF or its deletion mutants (100 pmol) and Wrap53 cRNA in 100 μL of binding buffer (50 mM Tris-HCl at pH 7.9, 100 mM KCl, 0.1% Nonidet P-40, 1.5 mM $MgSO_4$). A master mix was prepared to ensure equal loading of RNA for all samples, and the reaction mixture was then incubated for 10 min at 4°C. GST beads were then added to the reaction mixture and incubated for 30 min at 4°C. CTCF–RNA complexes were pulled down using GST beads, washed twice with binding buffer, and divided in half for protein and RNA analysis. Protein was separated by SDS-PAGE and detected by SYPRO Red staining. RNA was purified using TRIzol (Invitrogen), separated by urea 8M-PAGE, and detected by SYBR Gold staining.

Dimer assays

A master mix was prepared with GST-tagged proteins (100 pmol) and 300 μg of NE in 300 μL of binding buffer (50 mM Tris-HCl at pH 7.9, 100 mM KCl, 0.1% Nonidet P-40, 1.5 mM $MgSO_4$) and then divided into three aliquots: no treatment, RNase A treatment (1 μg; Ambion), and Turbo DNase treatment (5 U). The reaction mixture was then incubated for 1 h at 4°C. GST beads were then added to the reaction mixture and incubated for 30 min at 4°C. CTCF complexes were pulled down using GST beads, washed twice with binding buffer, separated by SDS-PAGE, transferred to nitrocellulose membranes, stained with Ponceau Red, and then blotted using antibody against HA.

Cell culture

Inducible U2OS cells expressing human CTCF and $CTCF_{\Delta 576–614}$ were generated by transfecting pTRIPZ-shCTCF and the relevant pINTA-N3 constructs (described above) using Lipofectamine 2000 (Invitrogen) followed by selection with 50 μg/mL Zeocin (Invitrogen). Cells were treated with doxycycline for 72 h at a concentration calibrated to allow reduction of the endogenous protein levels and give rise to levels of CTCF and $CTCF_{\Delta 576–614}$ similar to that of untreated cells.

Gel filtration chromatography

Recombinant CTCF or RBRs were loaded onto a Superose 6 PC 3.2/30 (GE Life Sciences) 2.4-mL sizing column with or without Wrap53 RNA and fractionated.

EMSA

Recombinant protein was incubated at room temperature with increasing concentrations of competitors prior to adding the RNA probe. RNA probe was then added and incubated for 30 min at room temperature in 15 μL of binding buffer containing 50 mM Tris-HCl (pH 7.5) at 4°C, 150 mM NaCl, 5 mM $MgCl_2$, 0.1 mM ZnSO4, 1 mM DTT, 10% glycerol, 0.1% Tween-20, 8 U of RNase inhibitor (New England Biolabs), and 1 μg of yeast transfer RNA (tRNA). Samples were resolved at 4°C by 0.5× TGE and 4% PAGE and visualized by autoradiography.

Acknowledgments

We are grateful to Dr. L. Vales for critical reading of the manuscript. We thank Dr. Pedro P. Rocha, Ramya Raviram, and Dr. Jane Skok for discussion and advice. We also thank the Genome Technology Center at New York University for help with sequencing. This work was supported by grants from the National Institute of Health (NIH) (GM-64844 and R37-37120) and the Howard Hughes Medical Institute to D.R., and the Dirección General de Asuntos del Personal Académico-Universidad Nacional Autónoma de México (IN209403 and IN203811) and Consejo Nacional de Ciencia y Tecnología, México (CONACyT; 42653-Q and 128464) to F.R.-T. R.S.-M. was supported by a PhD and an international research stay fellowship from CONACyT (213029). E.G.-B was supported by a PhD fellowship from CONACyT and Dirección General de Estudios de Posgrado-Universidad Nacional Autónoma de México (DGEP; 207989). R.B. was supported by a Helen Hay Whitney Foundation post-doctoral fellowship and the Helen L. and Martin S. Kimmel Center for Stem Cell Biology post-doctoral fellow award. V.N. is supported by an NIH training grant from the Graduate Program in Cellular and Molecular Biology (T32 GM007238). Additional support was provided by the PhD Graduate Program "Doctorado en Ciencias Bioquímicas" to the Instituto de Fisiología Celular and the Universidad Nacional Autónoma de México.

References

Barkess G, West AG. 2012. Chromatin insulator elements: establishing barriers to set heterochromatin boundaries. *Epigenomics* **4:** 67–80.

Brown RS. 2005. Zinc finger proteins: getting a grip on RNA. *Curr Opin Struct Biol* **15:** 94–98.

Chen H, Tian Y, Shu W, Bo X, Wang S. 2012. Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PLoS ONE* **7:** e41374.

De La Rosa-Velázquez De IA, Rincón-Arano H, Benítez-Bribiesca L, Recillas-Targa F. 2007. Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Res* **67:** 2577–2585.

Dignam JD, Lebovitz RM, Roeder RG. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* **11:** 1475–1489.

Filippova G, Fagerlie S, Klenova E. 1996. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol Cell Biol* **16:** 2802–2813.

Hafner M, Landgraf P, Ludwig J, Rice A, Ojo T, Lin C, Holoch D, Lim C, Tuschl T. 2008. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods* **44:** 3–12.

Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CWH, Ye C, Ping JLH, Mulawadi F, et al. 2011. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet* **43:** 630–638.

Holwerda SJB, 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.

Huthoff H, Autore F, Gallois-Montbrun S, Fraternali F, Malim MH. 2009. RNA-dependent oligomerization of APOBEC3G is required for restriction of HIV-1. *PLoS Pathog* **5:** e1000330.

Kaneko S, Bonasio R, Saldaña-Meyer R, Yoshida T, Son J, Nishino K, Umezawa A, Reinberg D. 2014. Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. *Mol Cell* **53:** 290–300.

Klenova EM, Chernukhin IV, El-Kady A, Lee RE, Pugacheva EM, Loukinov DI, Goodwin GH, Delgado D, Filippova GN, Leon J, et al. 2001. Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. *Mol Cell Biol* **21:** 2221–2234.

Lobanenkov VV, Nicolas RH, Plumb MA, Wright CA, Goodwin GH. 1986. Sequence-specific DNA-binding proteins which interact with (G + C)-rich sequences flanking the chicken c-myc gene. *Eur J Biochem* **159:** 181–188.

Lobanenkov VV, Nicolas RH, Adler VV, Paterson H, Klenova EM, Polotskaja AV, Goodwin GH. 1990. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5′-flanking sequence of the chicken c-myc gene. *Oncogene* **5:** 1743–1753.

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

Mahmoudi S, Henriksson S, Corcoran M, Méndez-Vidal C, Wiman KG, Farnebo M. 2009. Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. *Mol Cell* **33:** 462–471.

Mahmoudi S, Henriksson S, Weibrecht I, Smith S, Söderberg O, Strömblad S, Wiman KG, Farnebo M. 2010. WRAP53 is essential for Cajal body formation and for targeting the survival of motor neuron complex to Cajal bodies. *PLoS Biol* **8:** e1000521.

Martin D, Pantoja C, Miñán AF, Valdes-Quezada C, Moltó E, Matesanz F, Bogdanović O, de la Calle-Mustienes E, Domínguez O, Taher L, et al. 2011. Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. *Nat Struct Mol Biol* **18:** 708–714.

Mili S, Steitz JA. 2004. Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses. *RNA* **10:** 1692–1694.

Nakahashi H, Kwon K-RK, Resch W, Vian L, Dose M, Stavreva D, Hakim O, Pruett N, Nelson S, Yamane A, et al. 2013. A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep* **3:** 1678–1689.

Ohlsson R, Renkawitz R. 2001. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet* **17:** 520–527.

Phillips JE, Corces VG. 2009. CTCF: master weaver of the genome. *Cell* **137:** 1194–1211.

Phillips-Cremins JE, Sauria MEG, Sanyal A, Gerasimova TI, Lajoie BR, Bell JSK, Ong C-T, Hookway TA, Guo C, Sun Y, et al. 2013. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153:** 1281–1295.

Riley KJ, Yario TA, Steitz JA. 2012. Association of Argonaute proteins and microRNAs can occur after cell lysis. *RNA* **18:** 1581–1585.

Saldaña-Meyer R, Recillas-Targa F. 2011. Transcriptional and epigenetic regulation of the p53 tumor suppressor gene. *Epigenetics* **6:** 1068–1077.

Sanyal A, Lajoie BR, Jain G, Dekker J. 2012. The long-range interaction landscape of gene promoters. *Nature* **489:** 109–113.

Searles MA, 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.

Soto-Reyes E, Recillas-Targa F. 2010. Epigenetic regulation of the human p53 gene promoter by the CTCF transcription factor in transformed cell lines. *Oncogene* **29:** 2217–2227.

Sun S, del Rosario BC, Szanto A, Ogawa Y, Jeon Y, Lee JT. 2013. Jpx RNA activates Xist by Evicting CTCF. *Cell* **153:** 1537–1551.

Suzuki HI, Yamagata K, Sugimoto K, Iwamoto T, Kato S, Miyazono K. 2009. Modulation of microRNA processing by p53. *Nature* **460:** 529–533.

Terribilini M, Sander JD, Lee J-H, Zaback P, Jernigan RL, Honavar V, Dobbs D. 2007. RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. *Nucleic Acids Res* **35:** 578–584.

Valadez-Graham V, Razin SV, Recillas-Targa F. 2004. CTCF-dependent enhancer blockers at the upstream region of the chicken α-globin gene domain. *Nucleic Acids Res* **32:** 1354–1362.

Vousden KH, Prives C. 2009. Blinded by the light: the growing complexity of p53. *Cell* **137:** 413–431.

Xiao T, Wallace J, Felsenfeld G. 2011. Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol Cell Biol* **31:** 2174–2183.

Yao H, Brick K, Evrard Y, Xiao T, Camerini-Otero RD, 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.

Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G. 2004. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell* **13:** 291–298.

Zlatanova J, Caiafa P. 2009. CTCF and its protein partners: divide and rule? *J Cell Sci* **122:** 1275–1284.

Epigenetics 6:9, 1068-1077; September 2011; © 2011 Landes Bioscience

Transcriptional and epigenetic regulation of the p53 tumor suppressor gene

Ricardo Saldaña-Meyer and Félix Recillas-Targa*

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

Key words: p53, DNA methylation, chromatin, histone modification, CTCF, Wrap53, epigenetics, cancer

The p53 tumor suppressor is one of the most studied molecules in cancer research. Despite the fact that there is a detailed understanding involving multiple aspects of the *p53*-associated biology, many aspects of its transcriptional regulation are still not well clarified. Limited information is available on how the *p53* gene is transcriptional and epigenetically regulated. The *p53* gene expression is tightly controlled through a variety of transcription factors, miRNAs, its anti-sense RNA Wrap53, the insulator protein CTCF and very likely by other genetic and epigenetic mechanisms. It's the intent of this article to review in depth important aspects concerning the transcriptional regulation of the *p53* gene and perhaps serve as a stepping-stone to begin a conceptual change on how future p53 research can be approached.

Overview

Over 30 years have past since the discovery of p53 and the relevance of the so-called "guardian of the genome" has grown exponentially. This can be appreciated from the plethora of research articles providing new insights and reviews written trying to integrate the complexity of *p53*. We believe one of the latest reviews on *p53*, written by Karen H. Vousden and Carol Prives titled "Blinded by the Light," describes perfectly how the current situation can overwhelm anyone dedicated to *p53* research.[1] Despite the fact that there is a detailed understanding about multiple aspects of the *p53*-associated biology, many areas still remain only partially explored.

It's remarkable how a key mechanism like the transcriptional regulation of the human *p53* gene, has been so grossly overlooked. With this in mind, it's our intent to review in depth all aspects concerning the transcriptional regulation of the *p53* gene and integrate current research. We'll discuss novel mechanisms that in our opinion should be the main focus for future research pertaining to the guarding of the genome.

*Correspondence to: Félix Recillas-Targa; Email: frecilla@ifc.unam.mx
Submitted: 05/26/11; Accepted: 06/23/11
DOI: 10.4161/epi.6.9.16683

Transcriptional Regulation

Transcription factors targeting the p53 gene promoter regulation. Most of what we know about the normal *p53* gene promoter since its characterization by Tuck and Crawford on 1989 was produced in the late 1990s and focused in the mouse and human *p53* gene promoters.[2] Early studies used reporter gene assays and in vitro gel-shift and more recently chromatin immunoprecipitation (ChIP) to validate that *p53's* gene expression is regulated by direct binding of several types of transcription factors (see below). To date, the mouse and human promoters have at least nine evolutionary conserved transcription factor binding motifs. **Figure 1** illustrates both the human and mouse promoters with the distribution of the transcription factor binding motifs characterized to be involved in *p53* gene expression regulation.

Early reports showed that Myc/Max heterodimers and USF could trans-activate both the murine and the human *p53* gene promoter.[3-6] Additionally, AP-1 and NFκB were also found to trans-activate the human *p53* gene promoter.[5] Furlong et al. (1996) observed that YY1 and NF1 activate the human *p53* gene promoter while competing for the same binding site.[6] The binding motifs for Pax in the human and murine promoters were found to inhibit the *p53* gene promoter activity when PAX5, PAX2 or PAX8 were overexpressed.[8] These were the first transcription factors described to negatively regulate *p53's* gene expression. Of note, the independent mutation of the AP-1, the NFκB or the Pax binding motif abolished completely the promoter's activity suggesting that these nuclear factors are essential for the *p53* gene promoter basal activity. Another conserved transcription factor, HOXA5, was found that when overexpressed in breast cancer cell lines promotes apoptotic cell death by upregulating *p53* expression. The binding of HOXA5 to the ATTA containing sequence in the murine promoter was confirmed in vitro by gel-shift.[9] It is worth mentioning, that the murine and human *p53* genes have two interferon-stimulated response elements (ISRE) within their promoter or their first-intron regions, respectively.[10] The ISREs are activated by the IFN-activated transcription factor ISGF3, a heterotrimeric complex formed by Stat1, Stat2 and IRF-9. Although, only the human *p53* gene has an ISRE in its promoter, the mouse sites are found inside the first intron region. Nevertheless, both in human and mouse IFNα/β caused transcriptional activation of the *p53* gene.[10] IFNα/β, are known to be essential cytokines for antiviral immunity response and are in

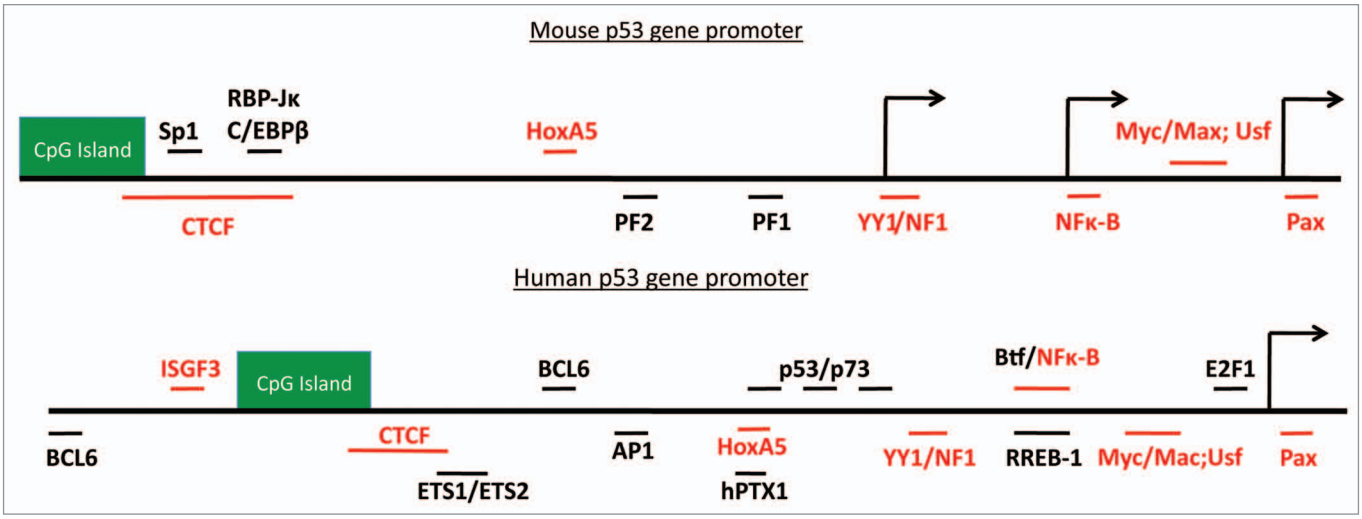


Figure 1. Comparison between the human and the mouse *p53* gene promoters. The binding motifs for transcription factors are shown. Evolutionary conserved motifs are highlighted in red and species-specific motifs are in black. CpG islands and the transcriptional start sites (TSS),[2,85] are represented as green boxes and arrows, respectively (the image is not to scale; for exact coordinates for the binding motifs see **Table 1**).

fact used for the treatment of some cancers, although the molecular basis for the treatment is still unclear.[11]

A recent discovery showed that both the human and murine promoters have a CTCF binding site downstream of a CpG island that has been experimentally demonstrated to bind to DNA in vivo and influence *p53*'s transcriptional regulation.[12,13] Our group and others have proposed that CTCF's role is apparently not as a classic transcription factor but instead as a barrier against transcriptional silencing. Because of its relevance, the role of CTCF in *p53*'s transcriptional regulation will be discussed further below.

Additionally, the human promoter has ten other specific transcription factors. Phan et al. (2004) provided evidence that the human proto-oncogene BCL6 has a repressive effect over *p53* gene transcription, by binding two specific DNA motifs within its promoter. BCL6 is abundant in germinal-center B cells where there is no *p53* gene expression.[14] Of note, BCL6 seems to be normally repressed by CTCF binding however during lymphomagenesis BCL6 expression is positively regulated by DNA methylation preventing CTCF-mediated silencing creating a potentially negative regulatory feedback.[15] In contrast, Wang et al. (2006) provided evidence of a positive feedback loop where the p73 or p53 protein directly binds to the *p53* gene promoter and trans-activate its expression both in physiological conditions and in response to cellular stress. These brings to light the importance of p53's autoregulation and that when disrupted it can lead to defects in cell cycle regulation and suppression of p53-mediated apoptosis.[16]

Moreover, the Pituitary homeobox 1 (hPitx1) was found to bind two DNA consensus motifs in the *p53* gene promoter and to activate its transcription.[17] hPitx1 is primarily expressed in the anterior pituitary and plays an important role as a tumor suppressor. Furthermore, it suppresses RAS activity and tumorigenicity, is crucial in the development of the vertebrate hindlimb and Pitx1-deficient mice present severe craniofacial abnormalities, pituitary defects and neonatal lethality.[18-21] Another transcription factor, Protein kinase C δ (PKCδ), was found to interact with the death-promoting transcription factor Btf and co-occupy the human *p53* gene core promoter element (CPE) causing its transactivation resulting in p53-dependent apoptosis in response to DNA damage.[22] Nevertheless, the direct binding to DNA was only determined for Btf and not for PKCδ.[22] PKCδ, a ubiquitously expressed member of the novel subfamily of PKC isoforms, is a common mediator of apoptosis in response to different stimuli.

In addition, the human promoter has two ETS-binding sites (EBS) in the form of a palindrome separated by 4 nucleotides (**Fig. 1**). ETS1 and ETS2 were found to activate transcription of a reporter gene by 5- to 10-fold in COS cells. In NIH3T3 cells a correlation was found between high amounts of *p53* and *ETS2* transcripts, suggesting that ETS2 upregulates *p53* expression in vivo.[23] ETS1 also binds cooperatively to the palindromic sequence on the human *p53* gene promoter avoiding the need of a binding partner.[24] Although ETS1 and ETS2 are highly related, their regulatory function differs significantly.[25] ETS1 has a restricted expression and it has been implicated in development of vascular systems, whereas ETS2 is ubiquitously expressed during murine development as well as in the adult.[26,27] Recently, the human *p53* CPE was shown to be a target for the Ras-responsive-element binding protein-1 (RREB-1). It was found to bind and trans-activate p53 to induce apoptosis in response to DNA damage. Its silencing also significantly reduces *p53* expression and notably suppresses the expression of the p53 target genes.[28] RREB-1 is expressed ubiquitously but its exact function remains to be elucidated. It was originally described to trans-activate the human calcitonin gene promoter-reporter construct during Ras- or Raf-induced differentiation in a medullary thyroid cancer cell line.[29] The most recent transcription factor found to directly bind the human *p53* gene promoter is E2F1. The authors suspected a direct role for E2F1 in the induction of mutant p53 and using

Table 1. Summary of binding motifs in the human and mouse *p53* gene promoters

Transcription factor	Sequence	Position	Reference
Human			
p53 and p73:	**1** TTA CTT GCC C **2** TGT CAT GGC G **3** TGT CCA GCT T	-118 to -108 -103 to -93 -91 to -81	Wang, et al, 2006[16]
hPitx1:	**1** GGA TTA **2** GGG ATT G	-121 to -115 -54 to -47	Liu, et al. 2007[17]
YY1/NF1:	GTC ATG GCG A	-102 to -92	Furlong, et al. 1996[7]
Myc/Max;USF:	CAT GTG	-33 to -27	Roy, et al. 1994[4]; Reisman, et al. 1993[3]
BCL6:	**1** CTC CTA GAA **2** TTC CTA AAA	-1903 to -1894 -379 to -370	Phan, et al. 2004[14]
HOXA5	GGA TTA CT	-121 to -113	Raman, et al. 2000[9]
NFκB and Btf (CPE):	AGG GGT TGA TGG GAT TGG GGT TTT CCC CT	-64 to -35	Krisch, et al. 1999[5]; Liu, et al. 2007[22]
EBS (ETS1, ETS2):	CGG AAA GCC TTC CT	-388 to -374	Venanzoni, et al. 1996[23]
ISGF3(ISRE1):	TGG AGA GAG AAA CTG	-1384 to -1369	Takaoka, et al. 2003[10]
AP1:	TGA CTC T	-173 to -166	Krisch, et al. 1999[5]
PAX:	GTG ACA CG	+95 to +103	Stuart, et al. 1995[8]
E2F1:	ACT GGC GC	-20 to -13	Bug, et al. 2011[31]
CpG Island:	CGG GAG GAG AGG CGA ACA GCG GAC GCC A…	-1028 to -683	UCSC
CTCF:	CGC TTC TAT CTT GGC GAG AAG CGC CTA C…	-781 to -381	UCSC; Barski, et al. 2007[72]
Mouse			
C/EBPβ and RBP-Jκ:	TTG GGA AA	-613 to -605	Boggs, et al. 2007[39]; Boggs, et al. 2008[41]
YY1:	TTG CCC TTA CTT GTT ATG GCG ACT ATC CAG	-12 to +18	Furlong, et al. 1996[7]
NF1:	GTT ATG GCG ACT ATC CAG CTT TGT G	0 to +25	Ginsberg, et al. 1990[37]
Pax:	GTC ACG C	+195 to 201	Stuart, et al. 1995[8]
PF1:	CTG ACT CT	-64 to -56	Ginsberg, et al. 1990[37]
PF2:	CCA CTC TTT ATA CTT GAC ACA	-195 to -174	Hale, et al. 1995[38]
HOXA5:	AAA ACT TAA	-210 to -201	Raman, et al. 2000[9]
NFκB/ETF:	TGG GAT TGG GAC TTT CCC CT	+47 to +67	Wu, et al. 1994[86]
Myc/Max; Usf:	CAC GTG	+69 to +75	Reisman, et al. 1993[3]; Ronen, et al. 1991[6]
CpG Island:	CGG AGC TAA GAG TCG CTC CTC CGA CGT C…	-1280 to -950	UCSC
Sp1:	CCC CGC C	-685 to -678	Putative Binding Site[3]
CTCF:	GGC TCT CCG CAT CCT CCT CCG ATT CCG A…	-999 to -370	Su, et al. 2009[12]

the Transfac algorithm they were able to predict two potential evolutionary conserved E2F1 binding sites BS-I and BS-II, within the *p53* gene promoter.[30] Only one site, BS-I, could bind E2F1 and was found to be necessary, but not sufficient, for efficient mutant p53 mRNA induction in response to chemotherapeutics.[31] Taking into account the novel interaction for E2F1 with the p53 promoter its important to point out that there are numerous interactions between these two proteins that are well established and have been reviewed in depth elsewhere in reference 32. For example, upon cellular stress E2F1 and p53 are activated and stabilized to induce apoptosis through common regulators like the CDKN2A locus and the p300 and PCAF acetylases.[33,34] Also, increased nuclear p53 has been observed when E2F1 directly binds to p53 phosphorylated at Ser315 improving p53's DNA binding and transactivation of its target genes.[35] In addition, there is widespread crosstalk of the Rb-E2f and Mdm2-p53 pathways through indirect interactions to control cell cycle arrest and senescence.[36]

The murine *p53* gen promoter has five other specific transcription factors regulating its expression. A putative binding site for Sp1 is found in the murine *p53* gene promoter; although no reports are available that experimentally demonstrate its binding.[3] Moreover, a conserved element matching 7 out of 8 bases from the AP1-binding site consensus sequence is found in the promoter. This AP1-like site was named PF1 and it appears to be a novel transcription factor since in vitro gel-shift assays AP1 couldn't compete efficiently for the PF1 binding site.[37] Another novel transcription factor, called PF2, was found to bind the murine *p53* gene promoter. The deletion of the PF2 site was observed to entirely inactivate the promoter activity, suggesting a critical role in basal transcription regulation.[38] Even though, both PF1 and PF2 were characterized on the early 1990's, to date, there are no additional reports about their role in the transcriptional regulation of the murine *p53* gene expression or any others genes. Boggs et al.[39] found that C/EBPβ-2 efficiently binds to the

murine *p53* gene promoter.[39] C/EBPβ-2 is a transcription factor critical for the normal growth and differentiation of various cell types.[40] It induces its expression in response to mitogen stimulation and as cells enter S phase. The same regulatory site also binds RBP-Jκ but instead it has a repressor effect over *p53* gene expression.[41] A more recent study, showed that C/EBPβ-2 and RBP-Jκ bind differentially in vivo. C/EBPβ-2 enhances p53 transcription during the transition from growth arrest to the entry into S-phase while RBP-JK represses p53 transcription during the same transition. These results suggest that both factors may work together or in coordination to regulate p53's activity during the cell cycle.[42]

As we can see, the *p53* gene promoter is highly regulated by the direct binding of a number of transcription factors involved in tumor suppression, oncogenic transformation, development, differentiation and apoptosis. It comes as no surprise that the *p53* gene promoter has such a variety of factors involved in essential cellular mechanisms controlling its expression. Nevertheless, p53's transcriptional regulation via a variety of transcription factors is merely one aspect of its regulation. Currently, very little is known about other genetic and epigenetic mechanisms of regulation for the *p53* gene. We believe that regulation through transcription factors need to be addressed in conjunction with epigenetic mechanisms and regulatory RNAs.

Epigenetic silencing via DNA methylation of the p53 gene promoter. Some in vitro studies have correlated DNA hypermethylation in the *p53* gene promoter region with a decrease in transcription.[43,44] Schroeder et al. provided the first experimental evidence for *DNA* methylation dependent silencing of the *p53* gene promoter using a reporter gene construct.[43] By means of in vitro DNA methylation, they were able to obtain a 90% reduction of gene expression. On a similar approach, on the rat *p53* gene promoter, single-site DNA methylation could reduce gene expression by 85%.[44] It's noteworthy that at that time no reports were published on a naturally occurring DNA hypermethylation event on the *p53* gene promoter and even to date there are only a few examples in human cancers that will be discussed in this section.

Some studies have documented in patient samples from primary hepatocellular carcinoma a correlation between aberrant *p53* gene promoter DNA hypermethylation with low levels of mRNA production.[45] Similar results were also found in 32% of acute lymphoblastic leukemia (ALL) patients and 40% of chronic lymphocytic leukemia (CLL) patients.[46] It's interesting to point out that *p53* mutations in ALL are very rare, only reported in 2–3% of patients. Suggesting that, at least in ALL, DNA hypermethylation is a silencing mechanism for p53 inactivation more frequent in the wild-type *p53* gene. In contrast, in a case study on sporadic adrenocortical cancers (ACCs), no DNA methylation was found in the *p53* gene promoter in ten patient samples with ACC and, as expected, in five normal adrenal samples examined.[47]

The aberrant regulation of *p53* gene promoter by DNA methylation remains controversial due to several inconsistencies in the literature and an apparent lack of direct methylation over the *p53* core promoter (see below). A meta-analysis shows that *p53* inactivation caused by mutation is infrequent (about 20% of the cases) in human breast cancer, this favors the prediction that *p53* silencing in this cancer type is controlled by epigenetic mechanisms.[48] A case study in breast cancer samples analyzing DNA methylation and mutation status showed that the *p53* gene promoter was exclusively methylated in 19% of the cases with no mutations detected.[49] Interestingly, there was no extensive DNA hypermethylation but instead methylation sites were restricted to 4 out of the 16 potential CpG dinucleotides around the core promoter. Additionally, a case study evaluated the DNA methylation status of BRCA1, BRCA2 and *p53* tumor suppressor gene promoters and focused in four groups: (1) a group of Jewish BRCA1/BRCA2 mutation carriers with or (2) without breast cancer (3) a group with sporadic breast cancer and (4) a group of healthy individuals. They found no DNA methylation in the BRCA2 and *p53* promoters in any case. In contrast, the BRCA1 promoter presented DNA methylation in only a small percentage (5%) of individuals in each group.[50] It's worth mentioning that the *BRCA1* gene promoter can be DNA hypermethylated in breast cancer and that CTCF protects the promoter against aberrant methylation.[51]

In a study for low-grade gliomas, high percentages of *p53* gene promoter DNA methylation were found in 60% of low-grade astrocytomas (29/48), in 61% of oligoastrocytomas (11/18), and 74% of oligodendrogliomas (31/42).[52] These results are in contrast with previous reports showing low proportion of *p53* DNA methylation in 8% of astrocytomas (2/24), in 2% of glioblastomas (1/43),[53] and 0% of oligodendroglial tumors (0/41).[54] It's relevant to the discussion that the primers used for methylation-specific PCR were the same in all three studies, but no systematic distribution of CpG dinucleotide methylation was determined. In contrast, in a case-control study of two *p53* single nucleotide polymorphisms (SNPs), with increased risk of tumor development in extra-axial brain tumors, the *p53* gene promoter was hypermethylated in 37.5% of meningiomas, 30% of schwannomas and in 52.6% of metastases.[55] We must take into account that most of the DNA methylation reports in cancer patients didn't carried out a sodium bisulfite conversion and sequencing. Thus, making the degree of promoter methylation impossible to determine and a detailed chromatin status of the *p53* gene promoter was lacking in every case.

In the study by Amatya et al. (2005) DNA methylation was also found in the *p53* gene promoter in three out of six malignant glioma cell lines (U87MG, LNT-229, T98G were methylated and U138MG, LN-18, LN-308 were unmethylated).[52] The DNA methylation status described for T98G transformed cell line is in controversy with recent results by our research group in which we found no DNA methylation of the *p53* gene promoter or in surrounding sequences in a range of 3 kb.[13] Intriguingly, incubation with the DNA methylation inhibitor (5-aza-2'-deoxicytidine) reactivated *p53* expression in the glioma T98G cell line in both studies.[13,52] The reactivation results in glioma cells by our research group bring to light the potential relevance of the genomic organization of the *p53* gene. Considering an structural role for CTCF, a possible explanation for our results is that aberrant DNA methylation can be incorporated "outside" of the promoter region, upstream and downstream relative to the

CTCF binding site, and that such distal genomic regions can contribute negatively to the *p53* gene expression in glioma cell lines. It's probable that an indirect effect of DNA methylation could be affecting the expression of p53 regulators or genomic loci directly or indirectly linked to *p53* gene regulation. This is an attractive possibility, which merits to be explored that can even include microRNAs, antisense RNAs, ncRNAs or Polycomb group proteins.[56-62]

In the basis of all these evidence, we conclude that in different types of tumors DNA methylation of the promoter region is not a widespread mechanism for transcriptional silencing of the *p53* gene, suggesting that other mechanisms may be involved. Furthermore, the effect of DNA methylation of genomic regions different from promoters on the transcriptional regulation of genes is not well explored. This might be of great significance since the methylation-sensitive factor CTCF is associated with the formation of intra- and interchromosomal contacts throughout the genome.[63]

miRNAs directly regulating p53. Another novel aspect involving p53's transcriptional regulation is its direct regulation via miRNAs. Le et al. (2009) used an in silico approach to search for potential miRNA-binding sites in the 3'-UTR region of the p53 mRNA. They found that miRNA-125b represses p53 post-trancriptionally in a manner dependent on its binding site in *p53* 3'-UTR.[58] This was the first report of a miRNA directly regulating *p53*. Shortly after, another research group published a similar interaction for the isoform miRNA-125a.[59] These two miRNAs gained much interest since a role in stem cell and differentiation was found. Interestingly, the decreased expression of miRNA-125a and 125b in embryonic stem cells correlated with a higher expression of *p53*. Conversely during differentiation the p53 protein and RNA levels are decreased in agreement with the increased expression of miRNA-125a and miRNA-125b.[60] These observations suggest that the regulation via miRNAs of p53 might be critical for stem cell ability to differentiate and for its self-renewal capacity.

Furthermore through another in silico screen, miRNA-504 was identified to posses a recognition site in the 3'-UTR of p53. This miRNA was found to directly interfere with p53 functions in processes such as apoptosis and cell cycle arrest in human cell lines and also to promote tumorigenicity.[61] The regulation of miR-504 and the biological significance of the regulation of p53 by miR-504 are still unclear. Tian et al. used a luciferase reporter screen to determine if p53 could be targeted by miRNAs.[62] From a pool of 107 miRNAs tested, only miR-1285 targeted the p53 3' UTR. It is noteworthy that in this assay miRNA-125b had no effect over p53 in disagreement with the results from Le et al.[58] To date, very little is known about miR-1285 and simply some expression profiles show that it is expressed in SHS5Y neuroblastoma cells and HepG2 liver cancer cells.[64]

Using a miRNA library with a dual-reporter screening assay and a computational analysis miRNA-25 and miRNA-30d were found to negatively regulate p53 by targeting the 3'-UTR of the human *p53* gene.[65] miRNA-25 and miRNA-30d were sufficient to reduce cellular apoptosis, perturb cell cycle arrest and senescence by downregulating the p53 protein and its target genes p21, Bax, Puma and Gadd45a. The screening assay used by this group didn't score miRNA-125b and miRNA-504 as possible candidates, even though these miRNAs are two known p53 regulators.[58,61] In light of this result, the authors concluded that the assay they used to find miRNAs targeting specific genes is not exhaustive and can have tissue-specific restrictions.[65]

It is our opinion, that the negative transcriptional regulation through miRNAs is an essential mechanism for the cell to control *p53*'s expression both in normal and aberrant conditions. With the exception of BCL6 and Pax, all known transcription factors that bind the human *p53* gene promoter positively regulate its expression. This fact renders miRNAs as one of the most likely processes that can negatively regulate *p53* gene expression and if deregulated can promote malignant transformation.[66] As described here, only six miRNAs are known so far to directly target the human *p53* gene transcript but very likely more will be identified in the near future. Specially taking into account that different approaches are used to find miRNAs targets and that all of these can offer different results.

p53 transcriptional regulation through the anti-sense RNA Wrap53. Recently, an unexpected mechanism for the transcriptional regulation of *p53* has been documented on the basis of the characterization of an anti-sense transcript named Wrap53.[57] After a series of thorough experiments Mahmoudi et al. concluded that the *p53* and *Wrap53*α transcripts interact in a head to head fashion in vitro, and that this RNA duplex interaction is necessary for the proper transcription of *p53*.[57] When knocking down the Wrap53α transcript or blocking the potential Wrap53/p53 RNA hybrids using 2-O-oligonucleotides, the levels of p53 mRNA were drastically reduced. On the other hand, overexpression of the overlapping exon 1α sequence upregulated p53 mRNA levels. Interestingly, no effect over the expression of Wrap53 was observed when either overexpressing or knocking down *p53*.

At the time of the original report, no function for the Wrap53 protein was determined. Although, its role in p53's post-transcriptional regulation was excluded in the report since the overexpression of the protein had no effect on p53's transcription or protein levels.[57] Not long after, the Wrap53 protein was characterized as a Cajal body protein that binds and directs small Cajal body-specific RNAs (scaRNAs), including telomerase RNA, to Cajal bodies.[67,68] These studies ascertain Wrap53 as an essential protein for the localization and processing of nuclear ribonucleoproteins. Furthermore, no apparent relationship between the p53 and Wrap53 proteins was reported. This presents additional evidence to the ever-growing notion that RNAs are essential regulators at many levels. It should be clear that Wrap53α is a coding RNA that was determined to have a positive effect over p53 transcription. Nevertheless, we cannot exclude the possibility that it may have additional roles yet to be discovered that might be similar from those of some ncRNAs. For example, long ncRNAs are known to be a requirement for the activation of gene expression of some critical regulators of development and differentiation.[55] On the contrary, another set of recently identified ncRNAs, termed long intervening noncoding RNA or lincRNAs, can have a silencing effect over transcription by interacting with PcG proteins from the PRC2 complex.[69] LincRNA-p21 is another example of a

different mechanism for repression by lincRNAs. Although, its exact mechanism remains unresolved, its clear that lincRNA-p21 after p53 induction interacts with the heterogeneous ribonucleoprotein K (hnRNP-K) to downregulate a large set of p53 target genes that block programmed cell death specifically but not cell cycle arrest.[70]

In a more clinical approach Bug and Dobbelstein (2011) showed that the use of different chemotherapeutics could have diverging effects on Wrap53 and mutant p53 transcriptional regulation.[31] In short, their results show that three topoisomerase II inhibitors, doxorubicin, idarubicin and etoposide, could exclusively increase Wrap53α mRNA levels.[31] Out of these drugs, doxorubicin was able to accumulate mutant p53 mRNA levels, nevertheless, idarubicin and etoposide failed to do so. It's important to remember that a positive correlation between p53 and Wrap53α mRNA levels was found in wild-type p53 cells, in contrast with the results in mutant p53 cells treated with some chemotherapeutics.[31,57] In light of this, the role of topoisomerase II may be crucial in regulating the chromatin structure for the *Wrap53* and *p53* genes.[71] These gives way to an interesting discussion about the relevance of the higher order organization of the *Wrap53* and *p53* genes that in our opinion could be essential for the proper transcription and response to cellular stress. This issue will be addressed in the analysis of the chromatin organization in the section below.

Comparative Analysis of the Chromatin Organization on the Mouse and Human *p53* Gene Promoter Regions

Chromatin organization on the mouse p53 gene promoter. As mentioned before a large amount of data has been generated on the p53 protein but it's until recently that its transcriptional regulation, more particularly, at the epigenetic level has been addressed. Su et al. (2009) were the first research group to approach p53's transcription with an epigenetic perspective.[12] Their findings were a direct consequence of a carefully and systematical experimental design. Dividing the murine *p53* gene in 25 regions for ChIP analysis, they observed that the mouse *p53* gene was organized in two very distinct epigenetic sub-domains. The first was completely unmethylated and presented epigenetic modifications characteristic of active genes, such as histone acetylation and H3K4me3. In addition, the second sub-domain was entirely DNA methylated and enriched with deacetylated histones (**Fig. 2A**).[12] Furthermore, they found seven regulatory factors bound to the gene body associated to chromatin remodeling and maintenance that are particularly enriched in the transition sites between each sub-domain. These factors are: CTCF, PARP1, topoisomerase II, HDAC1 and Dnmt1 that were found in both sites, while topoisomerase I and Brg1 were associated only with the second site. It is noteworthy that two binding sites for CTCF were found located coincident with the boundaries of each sub-domain (**Fig. 2A**). CTCF is the only insulator protein reported in mammals and it has been implicated in gene regulation in many levels working as a transcription factor, having insulator activity, involved in genomic imprinting and long-range chromatin interactions.[63] In the *p53* mouse gene CTCF was shown to possess insulator activity, apparently serving as a barrier between two chromatin sub-domains. Interestingly, CTCF was dissociated from the second sub-domain when cells were synchronized in G_2/M or treated with a PARP1 inhibitor, however it remained bound to the first sub-domain. The loss of CTCF in the second sub-domain correlated with low *p53* expression and the formation of a higher-order structure and the incorporation of histone H1.[12] These results suggest that CTCF's binding to the first sub-domain isn't dependent on poly(ADP-ribosyl)ation but more importantly is essential for transcription and the maintenance of the epigenetic sub-domains of the *p53* gene.

This study provides support for the idea that formation of distinct and regulated epigenetic sub-domains, determined by CTCF binding, within one gene is possible.[12] Such organization adds a novel functional level for CTCF and its partners in gene regulation at the chromatin domain level. These raise several new questions, for example, how common and relevant is the distribution of CTCF inside the gen body? Nonetheless, at the same time this evidence provides a glimpse for a much more complete understanding of the transcriptional regulation for the mouse *p53* gen.

Chromatin organization on the human p53 gene promoter. To the best of our knowledge no studies have focused in the chromatin organization of the human *p53* gene. The lack of a detailed study of the region made us analyze the information available on the UCSC database deposited from a chromatin immunoprecipitation genome wide sequencing (ChIP-seq) data.[72] The study by Barski et al. (2007) focused mainly on histone methylation but also included CTCF in the chromatin immunoprecipitation.[72] After analyzing the information from the human databases we encountered some similarities with the mouse *p53* gene organization. As expected, both promoters were enriched with common markers for active promoters like H3K4me3 and RNA Pol-II and also, downstream of its associated CpG island, the insulator protein CTCF (**Fig. 2B**). Additionally, the human promoter was enriched with the histone variant H2A.Z on the promoter region found both upstream and downstream of transcriptional start sites.[72] Especially, one similarity between the two promoters caught our eye. Two CTCF binding sites were also found in the human *p53* gene, one on the promoter region and another well inside the gene body on intron 9. Interestingly, the enrichment of CTCF on the internal binding site correlated with a local enrichment of the repressive chromatin mark H3K9me3. This suggests that CTCF might be allowing a chromatin configuration essential for active transcription. It's important to explore if the case for human might have similar regulatory consequences as in the mouse, since a comparable organization of CTCF bindings sites is present at the human *p53* gene.

One example of an internal CTCF site controlling gene expression has recently been described for the pro-apoptotic gene PUMA.[73] In this case CTCF represses basal PUMA expression by recruiting a Cohesin complex needed to define an intragenic chromatin boundary, more precisely at intron 4. Both the *PUMA* and *p53* genes have internal CTCF sites that appear to have opposite structural and regulatory functions. Interestingly,

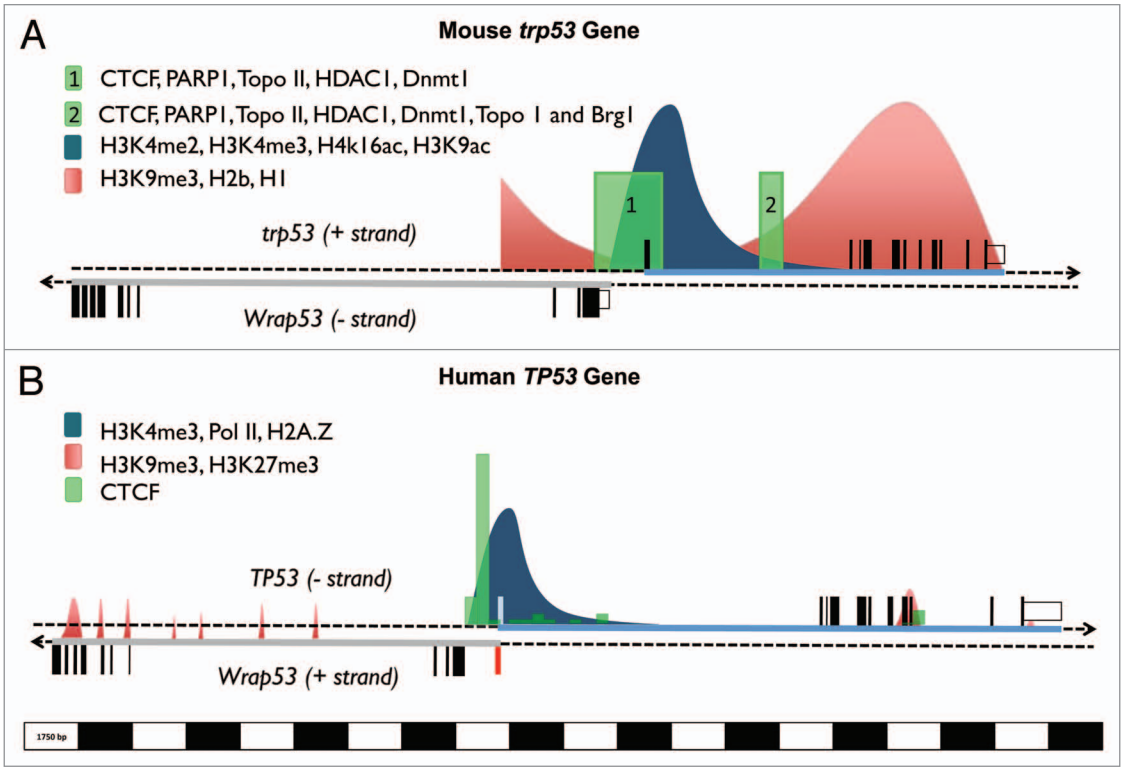


Figure 2. The genomic organization of the human and mouse *p53* gene. (A) Both the *Wrap53* and *trp53* murine genes are illustrated at scale. The exons are shown as vertical black lines and the UTR's as empty boxes. The enrichment of active histone marks are shown in dark green, the repressive histone mark H3K9me3, the histone variant H2b and the heterochromatin histone H1 are shown in red, the binding of chromatin modification and transcription regulatory factors are shown in light green. All enrichments shown were obtained through conventional ChIP experiments with a 700 base resolution.[12] (B) The *Wrap53* and the *p53* human genes are illustrated at scale. The first exon for *Wrap53*α is shown as a red vertical line and the p53 first exon as a blue vertical line; all other exons are shown as vertical black lines and the UTR's as empty boxes. Alternative exons and TSS for Wrap53 and Tp53 are not shown for simplicity. Enrichment for repressive chromatin marks are illustrated in red color and the active histone mark H3K4me3, the histone variant H2A.Z and RNA Pol-II in dark green; the binding of the multifunctional CTCF nuclear factor is shown in light green. All enrichments shown were obtained through ChIP-Seq experiments with a single-nucleosome resolution.[72]

PUMA appears to be in an arrested transcriptional state with one active and one repressed chromatin sub-domain. Meanwhile, during cellular stress conditions CTCF is dissociated and gene expression is activated.[74] This is similar to the murine *p53* gene that is also organized in two clear epigenetic sub-domains, except than in contrast gene expression is repressed upon dissociation of CTCF from the second site.[12] The case of the *p53* gene could be closer related to a recent study where the intragenic binding of CTCF functions as a "late" checkpoint in early stages of elongation at the U2 snRNA and β-actin genes.[73] Intragenic CTCF binding sites seem to be common, 30% of genes, nonetheless the importance and relevance of these sites is still lacking.[75]

Our analysis shows that the chromatin organization of the human and mouse *p53* present some similarities (**Fig. 2**). The CTCF binding sites suggest that this organization could be essential for transcriptional regulation and might be similar in both the mouse and the human *p53* genes. Another significant similarity is that both genes have an associated antisense gene. Wrap53 has been determined to have an essential positive regulatory role on p53 transcription, however the role for its murine counterpart hasn't yet been explored.[57] In contrast, the main difference found in the human *p53* gene organization is that apparently it's not arranged in two distinct epigenetic sub-domains. Further research will be essential to better our understanding of the epigenetic mechanisms involved in the *p53* gene regulation.

Epigenetic regulation of the human p53 gene promoter on transformed cell lines. Our research group reported a novel epigenetic mechanism for *p53* silencing. The transcription factor CTCF in the human *p53* gene promoter was found to contribute to its transcriptional expression and had the ability to maintain the promoter in a local open chromatin configuration.[13] When CTCF was knocked-down on HeLa cells the expression of the *p53* gene and its protein synthesis was decreased significantly. When the chromatin was analyzed an enrichment of histone H3K9me3 and H4K20me3 was found in the distal portion of the promoter (>1 kb from the TSS), with a notable decrease of histone H3 acetylation. In contrast, in the core promoter, no changes in histone H3 acetylation were found but an increase in H3K9me3, H4K20me3 and particularly H3K27me3 repressive histone marks was observed, suggesting that the PcG proteins could be involved in the abnormal epigenetic silencing of the human *p53* gene promoter. In two glioma cell lines, T98G and U87MG, CTCF is not bound to the promoter region in vivo, correlating with the lack of expression for the *p53* gene. Equally,

in this in vivo context the chromatin showed an aberrant gain of repressive histone marks, an effect more clearly observed in the glioma T98G cell line. These results suggested that CTCF protects the *p53* gene promoter against repressive histone marks.[13] Remarkably, no direct correlation was detected between repression and DNA hypermethylation. This is in contrast with what has been described for other tumor suppressor genes such as BRCA1, Rb and p16.[51,76,77] In conclusion, CTCF doesn't seem only to protect genomic areas from DNA methylation, but it also seems capable to protect against the repressive effects of the Polycomb group proteins and their heterochromatic influence.

Conclusions and Perspectives

There is extensive information about the role of the p53 protein however what we know concerning its transcriptional regulation and its chromatin organization, especially in humans, is very scarce. As described above, the *p53* gene expression is tightly regulated through a variety of transcription factors, miRNAs, its anti-sense RNA Wrap53, the insulator protein CTCF and very likely by other genetic and epigenetic mechanisms. The *p53* gene is mutated and expressed at significantly elevated levels in many tumors.[78] Although, the lack of expression of the wild-type *p53* gene in tumor samples or transformed cell lines is not well understood. Only few examples correlated the lack of expression of *p53* with the overexpression of a transcription factor or miRNAs.[14,62]

We believe that the role of the transcription factors involved in *p53*'s transcriptional regulation should be addressed taking into consideration its chromatin context and the epigenetic mechanisms needed to activate or silence gene expression. We are convinced that particular attention should be paid to well-known transcription factors that bind directly to the *p53* gene promoter like Myc, YY1 and p53 that have been implicated in the regulation of the chromatin structure. For example, Myc was found to upregulate chromatin acetyltransferases, regulate a widespread euchromatin program and also that its binding is dependent on the chromatin context of its target promoter region.[79-81] Furthermore, YY1 in association with other proteins is participating in histone modification and chromatin remodeling.[82,83] The p53 protein has been implicated in a number of epigenetic processes. It's essential for the recruitment of histone methyltransferases, histone acetyltransferases, histone desacetylases and also of the histone variant H2A.Z to p53 target genes.[84] Intriguingly, even though a positive feedback loop for *p53* gene expression was established several years ago none of the studies concerning the transcriptional regulation of *p53* has focused on the many factors where p53 is involved in histone modifications and its effects on gene expression.[16] In light of this, the regulation of its own transcription suggests that p53 could be working not only as a classical transcription factor but also by recruiting histone-modifying enzymes to allow a permissive chromatin configuration.

Furthermore, a novel aspect of *p53* transcriptional and epigenetic regulation emerges from its relationship with the multifunctional CTCF nuclear factor. When knocking-down CTCF the human *p53* gene loses its expression in transformed cell lines supporting its relevant contribution to p53 expression regulation, although the exact mechanism by which this occurs remains unexplored.[13] All of these observations rise more than a few questions. Is CTCF the critical cause for the *p53* gene silencing? Is this silence occurring exclusively in abnormal conditions? Is DNA methylation involved directly or indirectly? How are histone-modifying enzymes recruited to the promoter? Is p53 essential for the chromatin configuration of its own promoter? Could ncRNAs be involved in the epigenetic silencing of the *p53* gene? Potentially and at many levels, these p53 regulators could represent effective therapeutic targets. Nonetheless, many of these questions still need to be addressed and answered to better understand the regulation of the "guardian of the genome" before significant clinical advances can be achieved.

Acknowledgments

This work was supported by the Dirección General de Asuntos del Personal Académico-Universidad Nacional Autónoma de México (IN203811) and Consejo Nacional de Ciencia y Tecnología (CONACyT: 128464). R.S.M. is supported by a Ph.D. fellowship from CONACyT (213029). Additional support was provided by the Ph.D., Graduate Program of Doctorado en Ciencias Bioquímicas, UNAM.

References

1. Vousden KH, Prives C. Blinded by the light: the growing complexity of p53. Cell 2009; 137:413-31.
2. Tuck SP, Crawford L. Characterization of the human p53 gene promoter. Mol Cell Biol 1989; 9:2163-72.
3. Reisman D, Elkind NB, Roy B, Beamon J, Rotter V. c-Myc trans-activates the p53 promoter through a required downstream CACGTG motif. Cell Growth Differ 1993; 4:57-65.
4. Roy B, Beamon J, Balint E, Reisman D. Transactivation of the human p53 tumor suppressor gene by c-Myc/Max contributes to elevated mutant p53 expression in some tumors. Mol Cell Biol 1994; 14:7805-15.
5. Kirch HC, Flaswinkel S, Rumpf H, Brockmann D, Esche H. Expression of human p53 requires synergistic activation of transcription from the p53 promoter by AP-1, NFkappaB and Myc/Max. Oncogene 1999; 18:2728-38.
6. Ronen D, Rotter V, Reisman D. Expression from the murine p53 promoter is mediated by factor binding to a downstream helix-loop-helix recognition motif. Proc Natl Acad Sci USA 1991; 88:4128-32.
7. Furlong EE, Rein T, Martin F. YY1 and NF1 both activate the human p53 promoter by alternatively binding to a composite element, and YY1 and E1A cooperate to amplify p53 promoter activity. Mol Cell Biol 1996; 16:5933-45.
8. Stuart ET, Haffner R, Oren M, Gruss P. Loss of p53 function through PAX-mediated transcriptional repression. EMBO J 1995; 14:5638-45.
9. Raman V, Martensen SA, Reisman D, Evron E, Odenwald WF, Jaffe E, et al. Compromised HOXA5 function can limit p53 expression in human breast tumours. Nature 2000; 405:974-8.
10. Takaoka A, Hayakawa S, Yanai H, Stoiber D, Negishi H, Kikuchi H, et al. Integration of interferon-alpha/beta signalling to p53 responses in tumour suppression and antiviral defence. Nature 2003; 424:516-23.
11. Pfeffer LM, Dinarello CA, Herberman RB, Williams BR, Borden EC, Bordens R, et al. Biological properties of recombinant alpha-interferons: 40th anniversary of the discovery of interferons. Cancer Res 1998; 58:2489-99.
12. Su CH, Shann YJ, Hsu MT. p53 chromatin epigenetic domain organization and p53 transcription. Mol Cell Biol 2009; 29:93-103.
13. Soto-Reyes E, Recillas-Targa F. Epigenetic regulation of the human p53 gene promoter by the CTCF transcription factor in transformed cell lines. Oncogene 2010; 29:2217-27.
14. Phan RT, Dalla-Favera R. The BCL6 proto-oncogene suppresses p53 expression in germinal-centre B cells. Nature 2004; 432:635-9.
15. Lai AY, Fatemi M, Dhasarathy A, Malone C, Sobol SE, Geigerman C, et al. DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas. J Exp Med 2010; 207:1939-50.
16. Wang S, El-Deiry WS. P73 or p53 directly regulates human p53 transcription to maintain cell cycle checkpoints. Cancer Res 2006; 66:6982-9.

17. Liu DX, Lobie PE. Transcriptional activation of p53 by Pitx1. Cell Death Differ 2007; 14:893-907.
18. Szeto DP, Rodriguez-Esteban C, Ryan AK, O'Connell SM, Liu F, Kioussi C, et al. Role of the Bicoid-related homeodomain factor Pitx1 in specifying hindlimb morphogenesis and pituitary development. Genes Dev 1999; 13:484-94.
19. Kolfschoten IG, van Leeuwen B, Berns K, Mullenders J, Beijersbergen RL, Bernards R, et al. A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity. Cell 2005; 121:849-58.
20. DeLaurier A, Schweitzer R, Logan M. Pitx1 determines the morphology of muscle, tendon, and bones of the hindlimb. Dev Biol 2006; 299:22-34.
21. Quentien MH, Barlier A, Franc JL, Pellegrini I, Brue T, Enjalbert A. Pituitary transcription factors: from congenital deficiencies to gene therapy. J Neuroendocrinol 2006; 18:633-42.
22. Liu H, Lu ZG, Miki Y, Yoshida K. Protein kinase C delta induces transcription of the Tp53 tumor suppressor gene by controlling death-promoting factor Btf in the apoptotic response to DNA damage. Mol Cell Biol 2007; 27:8480-91.
23. Venanzoni MC, Robinson LR, Hodge DR, Kola I, Seth A. ETS1 and ETS2 in p53 regulation: spatial separation of ETC binding sites (EBS) modulate protein:DNA interaction. Oncogene 1996; 12:1199-204.
24. Baillat D, Laiten C, Leprivier G, Margerin C, Aumercier M. Ets-1 binds cooperatively to the plindromic Ets-binding sites in the p53 promoter. Biochem Biophys Res Commun 2009; 378:213-7.
25. Seth A, Robinson L, Panayiotakis A, Thompson DM, Hodge DR, Zhang XK, et al. The EndoA enhancer contains multiple ETS binding site repeats and is regulated by ETS proteins. Oncogene 1994; 9:469-77.
26. Paradanaud L, Dieterlen-Lievre F. Expresión of C-ETS1 in early chick embryo mesoder: relationship to the hemaglioblastic lineage. Cell Adhs Commun 1993; 1:151-60.
27. Kola I, Brookes S, Green AR, Graber R, Tymms M, Papas TS, Seth A. The Ets1 transcription factor is widely expressed during murine embryo development and is associated with mesodermal cells in morphogenetic processes such as organ formation. Proc Natl Acad Sci USA 1993; 90:7588-92.
28. Liu H, Hew HC, Lu ZG, Yamaguchi T, Miki Y, Yoshida K. DNA damage signalling recruits RREB-1 to the p53 tumour suppressor promoter. Biochem J 2009; 422:543-51.
29. Thiagalingam A, De Bustros A, Borges M, Jasti R, Compton D, Diamond L, et al. RREB-1, a novel zinc finger protein, is involved in the differentiation response to Ras in human medullary thyroid carcinomas. Mol Cell Biol 1996; 16:5335-45.
30. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, et al. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res 2006; 34:108-10.
31. Bug M, Dobbelstein M. Anthracyclines induces the accumulation of mutant p53 through E3F1-dependent and -independent mechanisms. Oncogene 2011; [ahead pub].
32. Polager S, Ginsberg D. p53 and E2f: partners in life and death. Nat Rev Cancer 2009; 9:738-48.
33. Lowe SW, Sherr CJ. Tumor suppression by Ink4a-Arf: progress and puzzles. Curr Opin Genet Dev 2003; 13:77-83.
34. Pediconi N, Ianari A, Costanzo A, Belloni L, Gallo R, Cimino L, et al. Differential regulation of E2F1 apoptotic target genes in response to DNA damage. Nat Cell Biol 2003; 5:552-8.
35. Murray-Zmijewski F, Slee EA, Lu X. A complex barcode underlies the heterogeneous response of p53 to stress. Nat Rev Mol Cell Biol 2008; 9:702-12.
36. Polager S, Ginsberg D. E2F—at the crossroads of life and death. Trends Cell Biol 2008; 18:528-35.
37. Ginsberg D, Oren M, Yaniv M, Piette J. Protein-binding elements in the promoter region of the mouse p53 gene. Oncogene 1990; 5:1285-90.
38. Hale TK, Braithwaite AW. Identification of an upstream region of the mouse p53 promoter critical for transcription expression. Nucleic Acids Res 1995; 23:663-9.
39. Boggs K, Reisman D. C/EBPbeta participates in regulating transcription of the p53 gene in response to mitogen stimulation. J Biol Chem 2007; 282:7982-90.
40. Johnson PF. Molecular stop signs: regulation of cell cycle arrest by C/EBP transcription factors. J Cell Sci 2005; 118:2545-55.
41. Boggs K, Henderson B, Reisman D. RBP-Jkappa binds to and represses transcription of the p53 tumor suppressor gene. Cell Biol Int 2009; 33:318-24.
42. Polson A, Takahashi P, Reisman D. ChIP (Chromatin Immunoprecipitation) analysis demonstrates coordinated binding of two transcription factors to the promoter of the p53 tumor suppressor gene. Cell Biol Int 2001; 34:883-91.
43. Schroeder M, Mass MJ. CpG methylation inactivates the transcriptional activity of the promoter of the human p53 tumor suppressor gene. Biochem Biophys Res Commun 1997; 235:403-6.
44. Pogribny IP, Pogribna M, Christman JK, James SJ. Single-site methylation within the p53 promoter region reduces gene expression in a reporter gene construct: possible in vivo relevance during tumorigenesis. Cancer Res 2000; 60:588-94.
45. Pogribny IP, James SJ. Reduction of p53 gene expression in human primary hepatocellular carcinoma is associated with promoter region methylation without coding region mutation. Cancer Lett 2002; 176:169-74.
46. Agirre X, Novo FJ, Calasanz MJ, Larráyoz MJ, Lahortiga I, Valgañón M, et al. Tp53 is frequently altered by methylation, mutation and/or deletion in acute lymphoblastic leukaemia. Mol Carcinog 2003; 38:201-8.
47. Sidhu S, Martin E, Gicquel C, Melki J, Clark SJ, Campbell P, et al. Mutation and methylation analysis of Tp53 in adrenal carcinogenesis. Eur J Surg Oncol 2005; 31:549-54.
48. Pharoah PD, Doy NE, Caldas C. Somatic mutations in the p53 gene and prognosis in breast cancer: a meta-analysis. Br J Cancer 1999; 80:1968-73.
49. Kang JH, Kim SJ, Noh DY, Park IA, Choe KJ, Yoo OJ, Kang HS. Methylation in the p53 promoter is a supplementary route to breast carcinogenesis: correlation between CpG methylation in the p53 promoter and the mutation of the p53 gene in the progression from ductal carcinoma in situ to invasive ductal carcinoma. Lab Invest 2001; 81:573-9.
50. Kontorovich T, Cohen Y, Nir U, Friedman E. Promoter methylation patterns of ATM ATR, BRCA1, BRCA2 and p53 as putative cancer risk modifiers in Jewish BRCA1/BRCA2 mutation carriers. Breast Cancer Res Treat 2009; 116:195-200.
51. Butcher D, Rodenhiser D. Epigenetic inactivation of BRCA1 is associated with aberrant expression of CTCF and DNA methyltransferase (DNMT3B) in some sporadic breast tumours. Eur J Cancer 2007; 43:210-9.
52. Amatya VJ, Naumann U, Weller M, Ohgaki H. Tp53 promoter methylation in human gliomas. Acta Neuropathol 2005; 110:178-84.
53. Gonzalez-Gomez P, Bello MJ, Arjona D, Lomas J, Alonso ME, De Campos JM, et al. Promoter hypermethylation of multiple genes in astrocytic gliomas. Int J Oncol 2003; 22:601-8.
54. Alonso ME, Bello MJ, Gonzalez-Gomez P, Arjona D, Lomas J, de Campos JM, et al. Aberrant promoter methylation of multiple genes in oligodendrogliomas and ependymomas. Cancer Genet Cytogenet 2003; 144:134-42.
55. Almeida LO, Custódio AC, Pinto GR, Santos MJ, Almeida JR, Clara CA, et al. Polymorphisms and DNA methylation of gene Tp53 associated with extra-axial brain tumors. Genet Mol Res 2009; 8:8-18.
56. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, et al. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 2007; 129:1311-23.
57. Mahmoudi S, Henriksson S, Corcoran M, Méndez-Vidal C, Wiman KG, Farnebo M. Wrap53, a natural p53 antisense transcript required for p53 induction upon DNA damage. Mol Cell 2009; 33:462-71.
58. Le MT, The C, Shyh-Chang N, Xie H, Zhou B, orzh V, et al. MicroRNA-125b is a novel negative regulator of p53. Genes Dev 2009; 23:862-76.
59. Zhang Y, Gao JS, Tang X, Tucker LD, Quesenberry P, Rigoutsos I, Ramratnam B. MicroRNA 125a and its regulation of the p53 tumor suppressor gene. FEBS Lett 2009; 583:3725-30.
60. Solozobova V, Blattner C. Regulation of p53 in embryonic stem cells. Exp Cell Res 2010; 316:2434-46.
61. Hu W, Chan CS, Wu R, Zhang C, Sun Y, Song JS, et al. Negative regulation of tumor suppressor p53 by microRNA miR-504. Mol Cell 2010; 38:689-99.
62. Tian S, Huang S, Wu S, Guo W, Li J, He Z. MicroRNA-1285 inhibits the expression of p53 by directly targeting its 3' untranslated region. Biochem Biophys Res Commun 2010; 396:435-9.
63. Phillips JE, Corces VG. CTCF: master weaver of the genome. Cell 2009; 137:1194-211.
64. Hausser J, Berninger P, Rodak C, Jantscher Y, Wirth S, Zavolan M. MirZ: an integrated microRNA expression atlas and target prediction resource. Nucleic Acids Res 2009; 37:266-72.
65. Kumar M, Lu Z, Tkwi AA, Chen W, Callander NS, Ramos KS, Young KH, Li Y. Negative regulation of the tumor suppressor p53 gene by microRNAs. Oncogene 2011; 30:843-53.
66. Barbarotto E, Schmittgen TD, Calin GA. MicroRNAs and cancer: profile, profile, profile. Int J Cancer 2008; 122:969-77.
67. Venteicher AS, Abreu EB, Meng Z, McCann KE, Terns RM, Veenstra TD, et al. A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. Science 2009; 323:644-8.
68. Tycowski KT, Shu MD, Kukoyi A, Steitz JA. A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. Mol Cell 2009; 34:47-57.
69. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 2009; 448:223-7.
70. Huarte M, Guttman M, Fledser D, Garber M, Koziol MJ, et al. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell 2010; 142:409-19.
71. Zhou K, Choe KT, Zaidi Z, Wang Q, Mathews MB, Lee CG. RNA helicase A interacts with dsDNA and topoisomerase IIalpha. Nucleic Acids Res 2003; 31:2253-60.
72. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. Cell 2007; 129:823-37.
73. Gomes NP, Espinosa JM. Gene-specific repression of the p53 target gene PUMA via intragenic CTCF-Cohesin binding. Genes Dev 2010; 24:1022-34.
74. Egloff S, Al-Rawaf H, O'Reilly D, Murphy S. Chromatin structure is implicated in "late" elongation checkpoints on the U2 snRNA and beta-actin genes. Mol Cell Biol 2009; 29:4002-13.
75. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 2007; 128:1231-45.
76. De La Rosa-Velázquez I, Rincón-Arano H, Benítez-Brbiesca L, Recillas-Targa F. Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. Cancer Res 2007; 67:2577-85.

77. Witcher M, Emerson B. 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.
78. Oren M, Rotter V. Mutant p53 gain-of-function in cancer. Cold Spring Harb Perspect Biol 2010; 2:1107.
79. Knoepfler PS, Zhang XY, Cheng PF, Gafken PR, McMahon SB, Einsenman RN. Myc influences global chromatin structure. EMBO J 2006; 25:2723-34.
80. Cotterman R, Jin VX, Krig SR, Lemen JM, Wey A, Farnham PJ, Knoepfler PS. N-Myc regulates a widespread euchromatic program in the human genome partially independent of its role as a classical transcription factor. Cancer Res 2008; 68:9654-62.
81. Guccione E, Martinato F, Finocchiaro G, Luzi L, Tizzoni L, Dall' Olio V, et al. Myc-binding-site recognition in the human genome is determined by chromatin context. Nat Cell Biol 2006; 8:764-70.
82. Rezai-Zadeh N, Zhang X, Namour F, Fejer G, Wen YD, Yao YL, et al. Targeted recruitment of a histone H4-specific methyltransferase by the transcription factor YY1. Genes Dev 2003; 17:1019-29.
83. Cai Y, Jin J, Yao T, Gottschalk AJ, Swanson SK, Wu S, et al. YY1 functions with INO80 to activate transcription. Nat Struct Mol Biol 2007; 14:872-4.
84. Beckerman R, Prives C. Transcriptional regulation by p53. Cold Spring Harbor Perspect Biol 2010; 2:935.
85. Bienz-Tadmor B, Zakut-Houri R, Libresco S, Givol D, Oren M. The 5' region of the p53 gene: evolutionary conservation and evidence for a negative regulatory element. EMBO J 1985; 4:3209-13.
86. Wu H, Lozano G. NFkappaB activation of p53. A potential mechanism for suppressing cell growth in response to stress. J Biol Chem 1994; 269:20067-74.

Molecular Cell
Article

Interactions between JARID2 and Noncoding RNAs Regulate PRC2 Recruitment to Chromatin

Syuzo Kaneko,[1,5] Roberto Bonasio,[1,4,5] Ricardo Saldaña-Meyer,[1] Takahaki Yoshida,[2] Jinsook Son,[1] Koichiro Nishino,[3] Akihiro Umezawa,[2] and Danny Reinberg[1,*]
[1]Howard Hughes Medical Institute and NYU School of Medicine, Department of Molecular Pharmacology and Biochemistry, New York, NY 10016, USA
[2]National Research Institute for Child Health and Development, Department of Reproductive Biology, Tokyo 157-8535, Japan
[3]University of Miyazaki, Faculty of Agriculture, Laboratory of Veterinary Biochemistry and Molecular Biology, Miyazaki 889-2192, Japan
[4]Present address: Department of Cell and Developmental Biology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA
[5]These authors contributed equally to this work
*Correspondence: Danny.Reinberg@nyumc.org
http://dx.doi.org/10.1016/j.molcel.2013.11.012

SUMMARY

JARID2 is an accessory component of *Polycomb* repressive complex-2 (PRC2) required for the differentiation of embryonic stem cells (ESCs). A role for JARID2 in the recruitment of PRC2 to target genes silenced during differentiation has been put forward, but the molecular details remain unclear. We identified a 30-amino-acid region of JARID2 that mediates interactions with long noncoding RNAs (lncRNAs) and found that the presence of lncRNAs stimulated JARID2-EZH2 interactions in vitro and JARID2-mediated recruitment of PRC2 to chromatin in vivo. Native and crosslinked RNA immunoprecipitations of JARID2 revealed that Meg3 and other lncRNAs from the imprinted *Dlk1-Dio3* locus, an important regulator of development, interacted with PRC2 via JARID2. Lack of *MEG3* expression in human induced pluripotent cells altered the chromatin distribution of JARID2, PRC2, and H3K27me3. Our findings show that lncRNAs facilitate JARID2-PRC2 interactions on chromatin and suggest a mechanism by which lncRNAs contribute to PRC2 recruitment.

INTRODUCTION

Polycomb group (PcG) genes are key epigenetic regulators in multicellular organisms, as they maintain transcriptional repression of lineage-specific genes throughout development, thus contributing to the stability of cell identity (Schwartz and Pirrotta, 2007). All mammalian PcG protein complexes identified so far perform their epigenetic function by acting on chromatin (Lanzuolo and Orlando, 2012); in particular, the *Polycomb* repressive complex 2 (PRC2) is responsible for di- and trimethylation of lysine 27 in histone H3 (H3K27me2/3) (Margueron and Reinberg, 2011), a hallmark of facultative heterochromatin (Trojer and Reinberg, 2007).

One of the outstanding questions regarding mammalian PRC2 function is that of specificity of action: how are certain genes selected for repression while others are unaffected? How can the same molecular machinery silence different genes in different cell lineages? Because none of the core components of PRC2 (EZH2, EED, SUZ12, RBBP4/7) possess a DNA binding domain (Margueron and Reinberg, 2011), it is believed that chromatin targeting must be specified elsewhere, by interactions with DNA-binding factors (Boulay et al., 2012; Kim et al., 2009), preexisting histone methylation (Margueron et al., 2009), chromatin-associated long noncoding RNAs (lncRNAs) (Rinn et al., 2007; Tsai et al., 2010), or a combination thereof (Margueron and Reinberg, 2011).

One essential factor for proper recruitment of PRC2 during the early phases of embryonic stem cell (ESC) differentiation is the Jumonji family, ARID domain-containing protein JARID2 (Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009), which is often deleted in chronic myeloid malignancies (Puda et al., 2012). In the absence of JARID2, PRC2 is recruited late and incompletely to its target genes and its enzymatic function is diminished (Li et al., 2010; Son et al., 2013), which results in failure to follow the differentiation program. Although JARID2 target sites are enriched for CGG- and GA-containing sequences (Peng et al., 2009), its DNA binding preferences lack the specificity to explain its distribution on chromatin (Li et al., 2010). Therefore, the nature of the recruitment pathway for JARID2 and the mode by which JARID2 regulates downstream steps of PRC2 assembly and function remain unclear.

Noncoding RNAs have been implicated in the regulation of epigenetic pathways, from early work on the lncRNA Xist in X chromosome inactivation (Brockdorff et al., 1992; Brown et al., 1992) and antisense transcripts in imprinted loci (John and Surani, 1996) to the more recent discovery of HOTAIR (Rinn et al., 2007) and its proposed role as a scaffold for chromatin-modifying "supercomplexes" (Tsai et al., 2010). Mammalian genomes contain thousands of lncRNAs (Guttman et al., 2009), most of which remain functionally uncharacterized. Because of their large size, potential for tertiary structure formation, and ability to form sequence-specific interactions with DNA, lncRNAs appear well suited to exchange information between

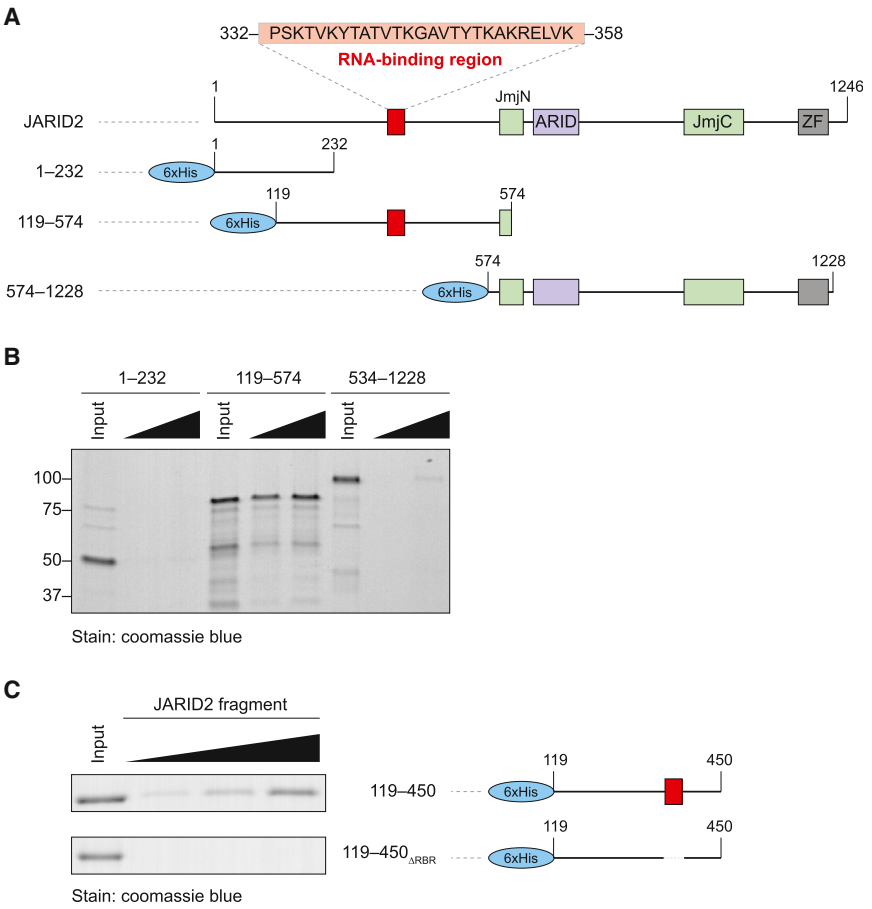


Figure 1. Identification of the RNA-Binding Region of JARID2

(A) Domain organization of human JARID2 and scheme of the 6xHis-fused truncations utilized in the mapping experiments.

(B) In vitro streptavidin pull-down after incubation of increasing concentrations of the indicated JARID2 recombinant fragments with biotinylated $HOTAIR_{1–333}$. Input, 2 μg; titration, 2 and 4 μg.

(C) High-resolution mapping of the residues of JARID2 necessary for RNA binding in vitro. The two indicated fragments (right) were incubated with $HOTAIR_{1–333}$ and assayed as in (B). Input, 2 μg; titration, 1, 2, and 4 μg.

See also Figure S1.

chromatin-modifying complexes and the genomic sequence (Bonasio et al., 2010; Rinn and Chang, 2012). *Polycomb* repressive complex-1 (PRC1), PRC2, and the MLL complex interact with the lncRNAs ANRIL, HOTAIR, and HOTTIP, respectively, and these interactions facilitate their recruitment to chromatin (Rinn et al., 2007; Wang et al., 2011; Yap et al., 2010). However, the molecular details and downstream consequences of these RNA-protein interactions remain poorly understood. For example, the RNA-binding activity of PRC2 has been attributed to both EZH2 (Kaneko et al., 2010; Zhao et al., 2010) and SUZ12 (Kanhere et al., 2010), and, in unbiased analyses, large portions of the transcriptome were reported to bind to PRC2 (Kaneko et al., 2013; Khalil et al., 2009; Zhao et al., 2010), raising the question of how specificity is achieved in vivo.

Here, we show that the PRC2 accessory subunit JARID2 binds to lncRNAs in vivo and in vitro and that its interaction with MEG3, an lncRNA encoded by the imprinted *DLK1–DIO3* locus, is necessary for proper recruitment and assembly of PRC2 at a subset of target genes in pluripotent stem cells.

RESULTS

JARID2 Binds to RNA In Vitro

Despite a requirement for JARID2 during development (Takeuchi et al., 1995) and its key role in PRC2 recruitment and function during differentiation of mouse ESCs (Li et al., 2010; Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009), the mechanisms by which JARID2 is targeted to chromatin and orchestrates PRC2 function remain poorly understood. Given that several PcG and PcG-associated proteins interact with lncRNAs, which in some cases regulate their recruitment to chromatin (Kanhere et al., 2010; Rinn et al., 2007; Yap et al., 2010), and based on our own preliminary observations in vitro (Kaneko et al., 2010), we hypothesized that lncRNAs might also regulate the function of JARID2.

We previously mapped an RNA-binding region (RBR) of EZH2, a core component of PRC2, and found that phosphorylation of a threonine within that region stimulated binding to lncRNAs (Kaneko et al., 2010). We performed similar in vitro RNA-binding assays on JARID2 using a bait spanning nucleotides 1–333 of HOTAIR, a lncRNA that regulates PRC2 function (Tsai et al., 2010), and detected an affinity for RNA within an internal fragment of JARID2, but not in the N-terminal or C-terminal regions (Figures 1A and 1B). Further mapping experiments revealed that the deletion of residues 332–358 resulted in a severe decrease of RNA binding in vitro (Figure 1C; Figures S1A and S1B available online). The sequence spanning these residues is conserved strongly in vertebrates (Figure S1C), but only very weakly with *Drosophila* (Figure S1D), suggesting that JARID2 may have acquired an additional layer of regulation in vertebrates.

JARID2 binds to EZH2, the catalytic component of PRC2 (Margueron and Reinberg, 2011) and stimulates its histone methyltransferase activity (Li et al., 2010; Son et al., 2013). These functions require the same internal fragment that contains the RBR (Figure S1E) but can be uncoupled from the latter, given that deletion of residues 332–358 did not affect the ability of JARID2 to interact with PRC2 in vivo (Figure S1F) or to stimulate its enzymatic activity (Figure S1G), whereas the 349–574 fragment did not bind to RNA (Figure S1A) but retained the ability to interact with nucleosomes (Figure S1E) (Son et al., 2013).

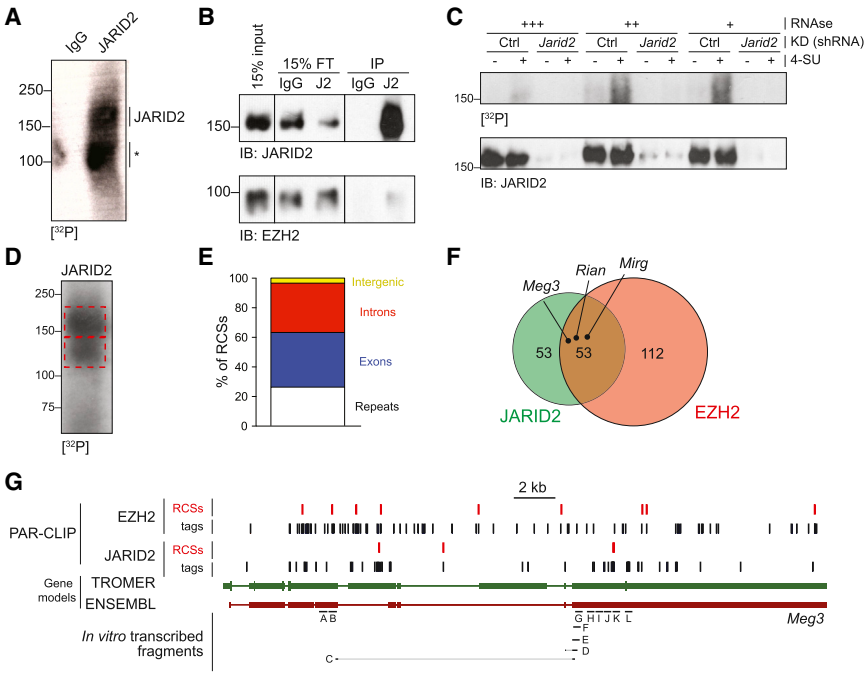


Figure 2. JARID2 and EZH2 Share Interacting lncRNAs In Vivo, Including Meg3

(A) PAR-CLIP with JARID2 antibodies or IgG in E14 ESC cells. The position of full-length JARID2 is indicated. The asterisk marks a presumed degradation product.

(B) Immunoblot on the same material utilized for the autoradiography in (A). J2, anti-JARID2 antibody.

(C) PAR-CLIP (top) and immunoblot for JARID2 (bottom) performed in cells pulsed (+) or not pulsed (−) with 4-SU and stably transfected with an shRNA against *Jarid2* or a control shRNA. Extracts were treated with increasing concentration of a cocktail of DNase-free RNase A and T1.

(D) PAR-CLIP-seq blot for JARID2.

(E) Distribution of JARID2 RCSs identified by PARalyzer in the genome. The stacked columns represent % of total RCSs. "Repeats" include all features listed in the RepMask database.

(F) Venn diagram of lncRNAs containing RCSs for JARID2, EZH2, or both. PAR-CLIP data for EZH2 were taken from GSE49433 (Kaneko et al., 2013).

(G) Genome browser view of JARID2 and EZH2 CLIP tags (black bars) or RCSs identified by PARalyzer (red bars) mapping to the Meg3 lncRNA. Gene models for *Meg3* according to both TROMER and ENSEMBL are shown. Meg3 fragments tested for in vitro binding are indicated at the bottom.

See also Figure S2 and Table S1.

Therefore, although partially overlapping regions of JARID2 are required for these various functions, the only activity that we could uniquely attribute to the 332–358 fragment is that of binding RNA in vitro. Henceforth, we will refer to these residues as the RBR of JARID2 and to the mutant protein lacking these residues as JARID2$_{\Delta RBR}$.

JARID2 and EZH2 Bind to lncRNAs In Vivo, Including Meg3

To determine whether JARID2 makes direct contacts with RNA in vivo, we utilized the photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) technique, which crosslinks RNA that has incorporated 4-thiouridine (4-SU) to proteins in vivo (Hafner et al., 2010). Consistent with our hypothesis, we detected a strong PAR-CLIP signal in immunoprecipitations (IPs) with our JARID2 antibody (Li et al., 2010) in extracts from embryonic day 14 (E14) ESCs (Figure 2A). These IPs were conducted in presence of 2% lauryldimethylbetaine, a zwitterionic detergent that almost completely abolished the PRC2-JARID2 interaction while preserving antibody reactivity (Figure 2B). Furthermore, the radioactive signal we observed must have originated from RNA crosslinked to JARID2, because it was dependent on the incorporation of 4-SU and was erased by treatment with increasing concentrations of RNase and by *Jarid2* knockdown (Figure 2C).

To identify the RNAs bound to JARID2 in vivo, we excised ^{32}P-labeled bands from the PAR-CLIP membranes (Figure 2D), eluted the crosslinked RNA, and sequenced it. In addition to the major, full-length band, we excised a faster migrating band that also reacted with JARID2 antibodies in WT but not in *Jarid2*$^{-/-}$ *cells* (data not shown). We obtained 90,000–200,000 unique CLIP tags in each replicate and analyzed their distribution using the PARalyzer software, which takes advantage of the T→C transitions caused by 4-SU crosslinking to discriminate signal from noise (Corcoran et al., 2011). PARalyzer identified 9,050 putative RNA-protein contact sites (RCSs) for JARID2, of which ~26% overlapped by more than 50% with repeats (Figure 2E) and were discarded. Among the 2,057 lncRNAs annotated in the mouse genome (ENSEMBL release 67), we identified 106 that contained at least one nonrepetitive RCS for JARID2 (Table S1). This bioinformatic pipeline applied to our previously generated EZH2 PAR-CLIP data (Kaneko et al., 2013) revealed that, in the same ESCs, EZH2 interacted with 165 lncRNAs, of which 53 were in common with JARID2 (Figure 2F; Table S1), including Meg3/Gtl2, Rian, and Mirg, three lncRNAs encoded within the imprinted *Dlk1-Dio3* locus.

We focused our attention on the lncRNA Meg3 (also known as Gtl2), because of previous reports linking it to pluripotency (Stadtfeld et al., 2010), imprinting (da Rocha et al., 2008), and PRC2 function (Zhao et al., 2010). Consistent with our PARalyzer analysis, several tags and RCSs from both JARID2 and EZH2 PAR-CLIP data mapped to *Meg3* (Figure 2G). Although some of the JARID2 CLIP tags mapped to a 5′ region annotated as an intron by ENSEMBL (red track), the existence of an exon in this region is supported by the TROMER database (Benson et al., 2004) (green track), and our own RNA-seq (data not shown). Another cluster of JARID2 CLIP tags mapped to the 3′ exon and accumulated in a region where PARalyzer identified an RCS (Figure 2G). Few CLIP tags mapped to highly expressed genes such as *Nanog* (Figure S2A) or *Gapdh* (Figure S2B), suggesting that the presence of *Meg3* tags reflected direct interactions in vivo.

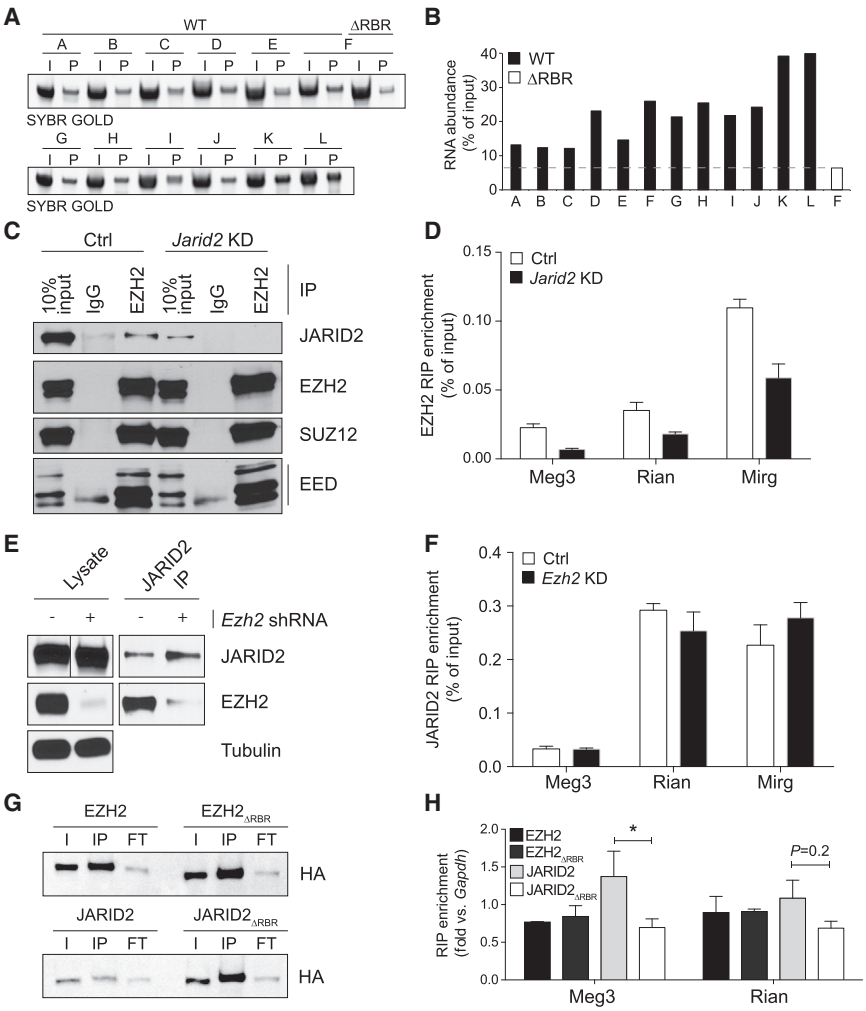


Figure 3. The JARID2 RBR Mediates Interactions of Meg3 with PRC2

(A) In vitro pull-down assay with different fragments of Meg3 using 4 μg of GST-JARID2 119–450 WT or ΔRBR.

(B) Quantification of bands shown in (A).

(C) RIPs for EZH2 were performed in E14 ESCs stably transfected with an shRNA against *Jarid2* (KD) or empty vector (ctrl). Coprecipitated proteins were revealed by western blot. The 10% input and IgG lanes are shown as controls.

(D) qRT-PCR on EZH2 RIPs from control E14 ESCs (white bars) or *Jarid2* knockdown E14 ESCs (black bars). Data are shown as percentage of RIP input. Bars represent the mean of four replicates + SD.

(E and F) As in (C) and (D) but *Ezh2* was knocked down and the RIP were performed with JARID2 antibodies.

(G) Western blots for HA RIPs from nuclear extracts of KH2 transiently transfected with N3-tagged EZH2, $EZH2_{\Delta RBR}$ (top), JARID2, and $JARID2_{\Delta RBR}$ (bottom). I, input; IP, HA immunoprecipitation; FT, flow-through.

(H) qRT-PCR normalized to *Gapdh* levels on HA RIPs described in (E). Bars indicate the mean of three biological replicates + SEM. *p < 0.05 by Mann-Whitney *U* test.

See also Figures S3 and S4.

The RBR of JARID2 Contributes to Meg3 Binding In Vitro and In Vivo

Having identified Meg3 as a JARID2-interacting lncRNA by CLIP, we next tested this interaction in vitro. We performed pull-down assays with in vitro-transcribed fragments of Meg3 (Figure 2G, bottom) obtained from a previously generated clone (Zhao et al., 2010). Although all tested fragments bound to a recombinant JARID2 fragment spanning the RBR (Figure 3A), the interaction was stronger for fragments originating from the 3′ end of the clone (Figure 3B), including one (fragment "K") that spanned the only RCS identified in this region (Figure 2G). We analyzed structural predictions for these fragments but found no obvious similarities among those that bound with higher affinity, except a trend for less stable structures (Figure S3). Importantly, a JARID2 fragment lacking the RBR displayed a pronounced reduction in binding affinity, at least toward the Meg3 fragment tested (fragment "F"; Figures 3A and 3B).

To validate these JARID2-RNA interactions with a technique more quantitative than PAR-CLIP, we resorted to native RNA immunoprecipitations (RIPs) followed by quantitative PCR (qPCR). Consistent with our PAR-CLIP results and previous reports (Zhao et al., 2010), Meg3 was enriched in PRC2 RIPs performed with EZH2 antibodies, and so were two other lncRNAs encoded within the same imprinted locus, Rian and Mirg (Figures 3C and 3D). However, when we depleted JARID2 by small hairpin RNA (shRNA)-mediated knockdown, we observed a considerable decrease in the amounts of these lncRNAs coprecipitating with PRC2 (Figures 3C and 3D), and similar results were obtained with SUZ12 antibodies (Figures S4A and S4B). Importantly, knockdown of *Ezh2* did not affect the ability of JARID2 to bind to Meg3, Rian, or Mirg (Figures 3E and 3F), suggesting that the JARID2-Meg3 interaction makes the largest contribution to the affinity of Meg3 for PRC2.

We then asked whether Meg3 bound to JARID2 via the RBR. To this end, we transiently expressed in mouse ESCs hemagglutinin (HA)-tagged EZH2, EZH2 lacking the previously identified RBR ($EZH2_{\Delta RBR}$) (Kaneko et al., 2010), JARID2, and $JARID2_{\Delta RBR}$ and performed HA RIPs followed by quantitative RT-PCRs (qRT-PCRs). Consistent with the results presented above, the interaction of Meg3 with JARID2 was significantly decreased when its RBR was removed, whereas EZH2 and $EZH2_{\Delta RBR}$ coprecipitated Meg3 with equal efficiencies (Figures 3G and 3H). We detected a similar trend for Rian, although in that case the difference did not reach statistical significance (Figure 3H). Similarly, human JARID2, but not human $JARID2_{\Delta RBR}$, bound to Meg3 in mouse ESCs (Figures S4C and S4D).

These data supported the conclusion that Meg3 interacts with PRC2 mainly through the RBR of JARID2, which led us to speculate that this lncRNA may participate in the function of JARID2 on chromatin.

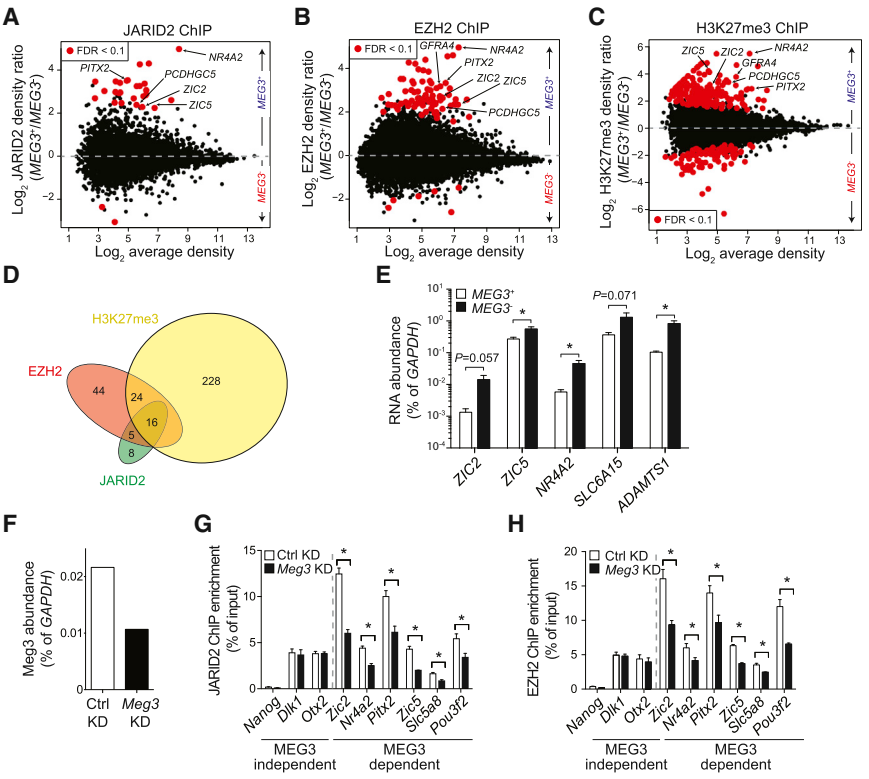


Figure 4. Decreased Occupancy of PRC2 at Some Chromatin Targets in *MEG3*$^{-}$ Cells

(A–C) MA plots for JARID2 (A), EZH2 (B), and H3K27me3 (C) occupancy, as determined by the normalized and input-corrected read densities in *MEG3*$^{+}$ (above dotted line) versus *MEG3*$^{-}$ (below dotted line) hiPSC lines. Each dot represents an ER in common between at least two hiPSC lines. DBRs with an FDR < 0.1 are displayed in red.

(D) Venn diagram for DBRs with FDR < 0.1.

(E) qRT-PCR analysis of PRC2 targets in *MEG3*$^{+}$ and *MEG3*$^{-}$ hiPSCs. Bars represent the mean RNA abundance (as percentage of *GAPDH*) in the five *MEG3*$^{+}$ and three *MEG3*$^{-}$ lines tested. *p < 0.05, as calculated by Mann-Whitney *U* test.

(F) qRT-PCR for *Meg3* 24 hr after transfection of KH2 ESCs with control (white bar) or *Meg3* siRNAs.

(G and H) ChIP-qPCR for JARID2 (G) or EZH2 (H) with primers mapping to PRC2 peaks near the indicated genes in KH2 ESCs treated with control (white bars) or *Meg3* (black bars) siRNAs. Bars represent the mean of three replicates + SEM. *p < 0.05 by Mann-Whitney *U* test.

See also Figures S5–S7 and Tables S2 and S3.

MEG3 Regulates PRC2 Occupancy In *Trans*

In light of the connection between MEG3 and pluripotency (Stadtfeld et al., 2010), we turned our attention to a set of eight human induced pluripotent stem cell (hiPSC) lines that differ greatly in their levels of *MEG3* expression (Nishino et al., 2011), thus offering a natural experimental system in which to study the effects of MEG3 on PRC2 function. We classified these lines into 5 *MEG3*$^{+}$ and 3 *MEG3*$^{-}$ (Figures S5A and S5B) and confirmed that *EZH2*, *SUZ12*, and *JARID2* were expressed at similar levels in *MEG3*$^{+}$ and *MEG3*$^{-}$ lines (Figures S5C–S5E). Importantly, pluripotent markers *OCT3/4* and *NANOG* were also expressed at comparable levels in all lines and at much higher levels than in differentiated cells, such as foreskin fibroblasts (Figures S5F and S5G).

Next, we performed chromatin immunoprecipitation (ChIP) followed by deep sequencing (ChIP-seq) for JARID2, EZH2, and H3K27me3, the product of PRC2 catalysis. We identified enriched regions (ERs) for all three features and compared normalized read densities in *MEG3*$^{+}$ versus *MEG3*$^{-}$ cells. The genome-wide analysis revealed 29, 89, and 268 differentially bound regions (DBRs) with a false discovery rate (FDR) < 0.1 in *MEG3*$^{+}$ versus *MEG3*$^{-}$ cells for JARID2, EZH2, and H3K27me3, respectively (Figures 4A–4C). At most of these *MEG3*-dependent DBRs, PRC2 exhibited stronger binding in the hiPSC lines that expressed *MEG3*, suggesting that the lncRNA had a stimulatory function (Figures 4A–4C; Figures S6A–S6C). Genes near the *MEG3*-dependent DBRs were enriched for functional terms related to the regulation of transcription during embryonic development and differentiation (Table S2), even taking into account the enrichment of developmental and transcription-related terms in the background population comprising all JARID2, EZH2, and H3K27me3 ERs (Table S3).

Of the 29 *MEG3*-dependent DBRs for JARID2, 21 overlapped with EZH2 DBRs (Figure 4D), and of these 16 also overlapped with H3K27me3 DBRs, a fraction much larger than expected by chance alone (p value < 10^{-20}, hypergeometric distribution). Among the regions with lower JARID2 occupancy in the absence of *MEG3* were the loci encoding the transcription factors *ZIC5*, *NR4A2*, *ZIC2*, and *PITX*, as well as the neuronal gene *PCDHGC5* (Figure S6D and data not shown), all of which function in differentiating or differentiated cells and must therefore be silenced in pluripotent stem cells. In all loci tested, the *MEG3*-dependent loss of PRC2 targeting resulted in transcriptional derepression (Figure 4E).

It has been suggested that, in mouse ESCs, Meg3 exerts a *cis*-repressive effect on the adjacent *Dlk1* gene by recruiting PRC2 (Zhao et al., 2010). To determine whether a similar mechanism was conserved in hiPSCs, we analyzed the *DLK1* promoter in *MEG3*$^{+}$ versus *MEG3*$^{-}$ hiPSCs. Unexpectedly, we observed no differences in JARID2 or PRC2 occupancy (data not shown) and no evidence of a reciprocal correlation in the levels of *MEG3* and *DLK1* RNA in these cells (Figure S5H) or with the protein-coding RNA at the other extremity of the imprinted locus, *DIO3* (Figure S5J).

Because the *DLK1-DIO3* locus encodes multiple ncRNAs (*MEG3*, *RIAN*, and *MIRG*), all repressed in *MEG3*$^{-}$ hiPSCs, we examined the effect of altering *MEG3* levels alone on PRC2 localization. We overexpressed *MEG3* (or GFP RNA as a control) in *MEG3*$^{-}$ hiPSCs and analyzed the distribution of JARID2 and EZH2 by ChIP-seq. A caveat of this experiment is that *MEG3* was expressed at levels ~10 times higher than in the average *MEG3*$^{+}$ line (Figure S7A); nonetheless, we observed increased

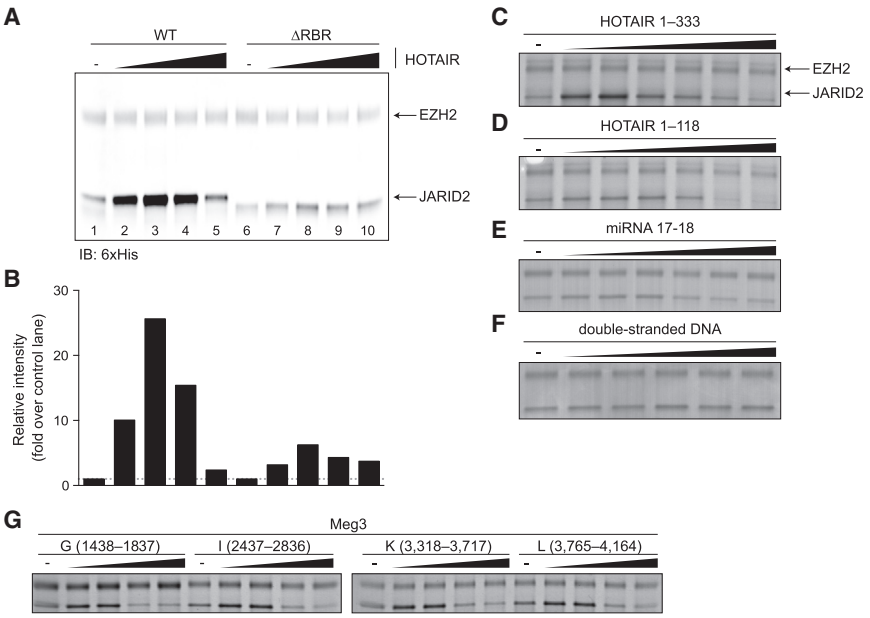


Figure 5. HOTAIR Stimulates EZH2–JARID2 Interactions via the JARID2 RBR

(A) FLAG-6xHis-tagged recombinant EZH2 (20 pmol) was incubated with 6xHis-tagged $JARID2_{119–450}$ WT or ΔRBR (40 pmol) in presence of increasing amounts of $HOTAIR_{1–333}$ (0–12 pmol). Pull-down was performed with anti-FLAG beads and proteins revealed by 6xHis immunoblot.

(B) Densitometric quantification of the signal for JARID2 shown in (A). WT and ΔRBR lanes were normalized each to lane 1 and lane 6 (no HOTAIR control), respectively.

(C–F) In vitro interaction stimulation assays with recombinant 6xHis-tagged $JARID2_{119–574}$ and FLAG-6xHis-tagged full-length EZH2 in presence of increasing concentrations of $HOTAIR_{1–333}$ (C), a smaller 5′ truncation of HOTAIR (D), a microRNA (E), or double-stranded DNA (F). Pull down was performed with anti-FLAG beads and proteins stained with Coomassie blue.

(G) Same as (C)–(F) using Meg3 fragments.

densities of JARID2 at some *MEG3*-dependent DBRs, but not at control targets (Figure S7B). However, we did not detect recovery of JARID2 occupancy at all JARID2 DBRs (Figure S7C), suggesting that the regulation afforded by endogenous MEG3 could not be fully recapitulated by providing large amounts of the lncRNA in *trans*. Interestingly, PRC2 was preferentially depleted from the EZH2 DBRs upon overexpression of *MEG3* (Figure S7D). Although this depletion of EZH2 was unexpected, the fact that the EZH2 DBRs were preferentially depleted by the manipulation of MEG3 levels is consistent with the idea that occupancy at these sites is selectively regulated by this lncRNA.

To further verify that the observed changes in PRC2 occupancy were directly related to the changes in MEG3 levels, we performed transient knockdown of the orthologous lncRNA in mouse ESCs. Despite only partial knockdown efficiency (~50%; Figure 4F), six out of nine DBRs tested lost JARID2 and EZH2 after Meg3 depletion (Figures 4G and 4H), suggesting that the regulation of PRC2 by Meg3 is conserved between human and mouse. PRC2 occupancy at *Plek2*, *Mbnl3*, and *Cdx4* was not affected by *Meg3* knockdown (data not shown), despite the fact that orthologous loci were among the MEG3-dependent DBRs in hiPSCs. However, it is not surprising that subtle differences in gene regulation would exist across this species barrier, especially given that mouse and human stem cells are not equivalent (Ginis et al., 2004).

Together, our ChIP experiments in hiPSCs and mouse ESCs support the conclusion that MEG3 acts in *trans* on PRC2 and JARID2 by facilitating their recruitment to a subset of target genes.

RNA Facilitates JARID2-PRC2 Interactions

Given that loss of MEG3 caused defects in PRC2 recruitment and that several lncRNAs crosslinked to both JARID2 and EZH2 in vivo, we hypothesized that RNA-mediated scaffolding could stabilize JARID2-PRC2 interactions. To test this hypothesis in vitro, we first examined the lncRNA HOTAIR, which functions as a scaffold between the LSD1/CoREST/REST complex and PRC2 (Tsai et al., 2010). Low amounts of HOTAIR stimulated the interaction between recombinant EZH2 and JARID2 fragments in vitro (Figures 5A and 5B, compare lanes 1–3), whereas higher concentrations of the RNA resulted in a return to baseline interaction levels, suggesting the possibility of squelching (Figures 5A and 5B, compare lanes 3–5). This stimulation was considerably reduced for JARID2 fragments that lacked the RBR (Figures 5A and 5B, lanes 6–10) but retained the ability to interact with EZH2 (Figure 5A, compare lanes 1 and 6) or when we replaced the HOTAIR lncRNA fragment with a shorter version, a control pre-microRNA that forms two stem-loops, or double-stranded DNA containing the HOTAIR lncRNA sequence (Figures 5C–5F). Stimulation of binding was also observed when we incubated recombinant EZH2 and JARID2 with those Meg3 fragments that displayed maximum affinity in the in vitro pull-down assay (Figure 5G). Therefore, stimulation of JARID2-EZH2 interactions may be a general mechanism of action for lncRNAs that bind to these proteins.

The JARID2 RBR Stimulates PRC2 Assembly on Chromatin

Having discovered that lncRNAs stimulate JARID2-PRC2 interactions in vivo (Figure 4) and in vitro (Figure 5) and knowing that the RBR was required for the latter (Figures 5A and 5B), we asked whether it was also required for the former. To this end, we overexpressed JARID2 or $JARID2_{\Delta RBR}$ in human foreskin fibroblasts, which express high levels of *HOTAIR* (Rinn et al., 2007), and measured its accumulation at known genomic targets by ChIP-qPCR. Wild-type (WT) and mutant JARID2 were expressed at comparable levels in lentivirally transduced fibroblasts (Figure 6A, top) and did not affect EZH2 levels (Figure 6A, middle). Upon JARID2 overexpression, we observed increased accumulation

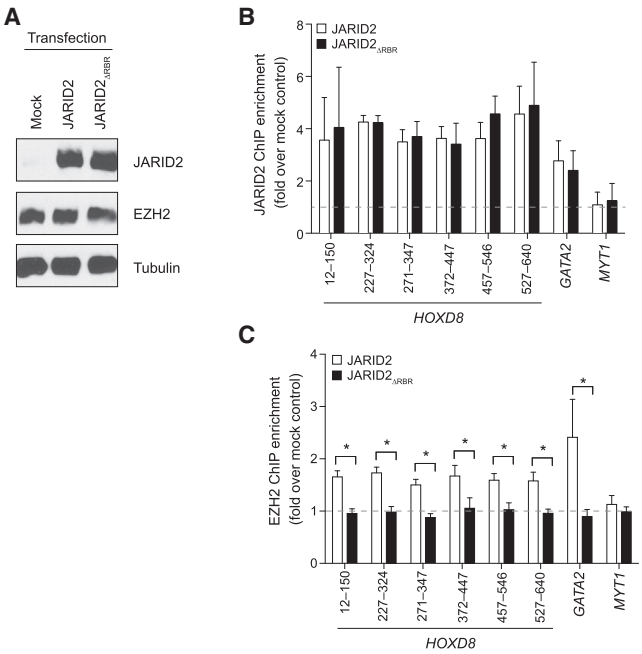


Figure 6. JARID2 Recruits EZH2 to Chromatin in an RBR-Dependent Manner

(A) Western blots for foreskin fibroblasts transduced with lentiviruses expressing JARID2, JARID2$_{\Delta RBR}$, or mock transduced.

(B and C) ChIP-qPCR with antibodies against JARID2 (B) or EZH2 (C) at known PRC2 chromatin targets in foreskin fibroblasts transduced as in (A). Several primer sets are shown for *HOXD8*. The primers for *MYT1* were designed at a distal location, devoid of PRC2, and serve as a negative control. The ChIP enrichment is normalized against that obtained with the same antibodies in mock-transfected control cells (dotted line). Bars represent mean of four technical replicates + SD (B) or three biological replicates + SEM (C). *$p < 0.05$ by Mann-Whitney *U* test.

of the protein at known chromatin targets, such as *HOXD8* and *GATA2*, compared with controls (Figure 6B). JARID2$_{\Delta RBR}$ accumulated at these targets with the same efficiency as the WT protein but, unlike the WT, it was incapable to recruit additional EZH2, which remained at the same levels as observed in the mock-transfected cells (Figure 6C). Given that these targets were already silent in the mock-transfected controls, we did not attempt to detect further repression at the transcriptional level.

These results suggest that RNA-protein interactions via the JARID2 RBR contribute to the recruitment and assembly of PRC2 on chromatin. Interestingly, depletion of HOTAIR in foreskin fibroblasts also impairs PRC2 recruitment at these loci (Rinn et al., 2007; Tsai et al., 2010), suggesting that it may act through interactions with the JARID2 RBR.

The fact that JARID2$_{\Delta RBR}$ is incapable of recruiting EZH2 to target sites is compatible with a model by which lncRNAs modulate JARID2-PRC2 interactions and orchestrate the distribution and activity of PRC2 on chromatin.

DISCUSSION

The results presented above allow us to add JARID2 to the growing list of chromatin-associated proteins that interact with lncRNAs and offer further support to the hypothesis that lncRNAs are a component of the *Polycomb* axis in mammals.

We mapped the JARID2 RBR to an N-terminal fragment that contains no annotated features or domains, despite the fact that, in addition to RNA binding, it is also responsible for interactions with SUZ12, EZH2, nucleosomes, and PRC2 stimulation (Kim et al., 2003; Li et al., 2010; Pasini et al., 2010; Son et al., 2013). Although we assigned these biochemical activities to distinct protein fragments (Figure S1), they do map to adjacent regions, and it is tempting to speculate that this vicinity might reflect a functional crosstalk, by which, for example, RNA binding could stimulate nucleosomal binding and contribute to PRC2 regulation.

Among the lncRNAs identified by PAR-CLIP, we focused our functional analysis on Meg3 because of its connections with ESC pluripotency, imprinting, and PRC2 function (da Rocha et al., 2008; Stadtfeld et al., 2010; Zhao et al., 2010). We demonstrated JARID2-Meg3 interactions using RIP-qPCR, PAR-CLIP, and in vitro pull-down assays. Importantly, we also found RCSs for EZH2 within *Meg3* in our previously published EZH2 PAR-CLIP data set (Kaneko et al., 2013) (Figures 2F and 2G). This is consistent with earlier results obtained by native and UV-cross-linked RIP (Zhao et al., 2010) and supports a model by which Meg3 contacts both JARID2 and EZH2 and stimulates their interaction. Without JARID2 the interaction between Meg3 and PRC2 is much weaker (Figure 3D), suggesting that of the two contact points, the one on JARID2 makes the larger contribution to the affinity for Meg3. We note that by using *Ezh2*$^{-/-}$ cells as a control, Zhao et al. could not have detected this requirement for JARID2, because in absence of EZH2 the IP performed with an anti-EZH2 antibody would not have recovered JARID2.

We envision a model in which some lncRNAs function as scaffold to stimulate assembly of PRC2 at JARID2 target sites (Figure 7A). In addition, the existence of DBRs that lose JARID2 occupancy in absence of MEG3 (Figure 4A) suggests that, at certain sites, lncRNAs might also be required for the initial recruitment of JARID2 (Figure 7B). In both scenarios, the net result of lncRNA action is to increase PRC2 occupancy and H3K27me3 deposition (Figures 7A and 7B). We have demonstrated this model using MEG3 and genomic targets in hiPSCs and mouse ESCs (Figure 4); however, the recovery of other lncRNAs cross-linked to both JARID2 and EZH2 by PAR-CLIP-seq (Figure 2F; Table S1) allows us to speculate that this mode of action might not be limited to MEG3. In fact, the JARID2$_{\Delta RBR}$ mutant does not recruit PRC2 to the *HOX* locus in foreskin fibroblasts, despite the fact that this locus is not MEG3 dependent. We propose that other lncRNAs function as scaffold for JARID2-PRC2 interactions in this setting and we note that foreskin fibroblasts express high levels of *HOTAIR* (Rinn et al., 2007), which is also able to stimulate JARID2-EZH2 interactions in vitro (Figure 5).

Not all lncRNAs bind equally to JARID2 and EZH2. Our reanalysis of EZH2 PAR-CLIP tags (Kaneko et al., 2013) identified a number of lncRNAs not shared with JARID2 (Figure 2F; Table S1) that, likely, regulate PRC2 function in JARID2-independent ways. In addition to lncRNAs, EZH2 binds to a variety of coding transcripts in vivo and in vitro with high affinity but seemingly low specificity (Davidovich et al., 2013; Kaneko et al., 2013). Those interactions appear to constitute a distinct regulatory

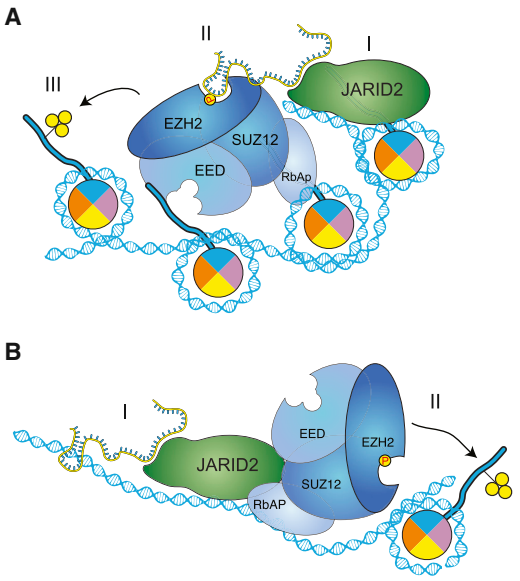


Figure 7. Proposed Model for the Interplay of lncRNAs, JARID2, and PRC2

(A) At some target genes, the presence of JARID2 by itself (I) is not sufficient for maximum PRC2 recruitment, which requires scaffolding by lncRNAs (II). The presence of both JARID2 and lncRNAs stimulates further recruitment and assembly of PRC2 on chromatin, resulting in increased H3K27me3 (III). The structure of lncRNAs bound to JARID2 (and PRC2) remains to be elucidated, and the one shown here is only for the purpose of illustration.

(B) In some cases, lncRNAs might contribute to the initial recruitment of JARID2 to chromatin (I). Because JARID2 also binds PRC2 via protein-protein interactions, this results in increased PRC2 recruitment and H3K27 methylation (II).

mechanism from the one described here, as they occur mostly at promoters of transcribed genes, which have low PRC2 occupancy (Davidovich et al., 2013; Kaneko et al., 2013) and are largely devoid of JARID2 (data not shown). However, we cannot exclude that nascent RNAs could compete with lncRNAs for binding to EZH2, which might help explain their apparent inhibitory function.

Our gain-of-function experiments in hiPSCs confirmed that at least some JARID2 DBRs identified in *MEG3*$^+$ versus *MEG3*$^-$ hiPSCs can be rescued by supplying MEG3 lncRNA in *trans* to otherwise *MEG3*-negative hiPSCs (Figure S7B). The ectopic expression of MEG3 lncRNAs caused EZH2 depletion rather than increased occupancy at the previously identified MEG3-dependent DBRs (Figure S7D), which was in contrast to our expectations. However, considering that lentiviral transduction resulted in 10-fold higher *MEG3* levels compared to endogenous *MEG3* (Figure S7A), we speculate that such an excess of RNA molecules acted in a dominant negative fashion through squelching, as seen in vitro (Figure 5). It is also possible that the random integration of the lentivirally encoded *MEG3* at ectopical sites within the genome may explain its unphysiological behavior in these experiments, given that the genomic location of lncRNA genes appears to play a pivotal role in their function (Ulitsky et al., 2011). Nonetheless, the fact that MEG3-dependent DBRs were selectively affected confirmed that occupancy at these targets is indeed regulated by MEG3, which was further supported by the results of transient *Meg3* knockdown in mouse ESCs (Figures 4G and 4H).

Genomic imprinting of *MEG3* is unstable in human ESCs, and several hiPSC lines, regardless of their parental cell type, maintain repression at this locus even after continuous passaging (Nishino et al., 2011). This is likely mediated by aberrant DNA hypermethylation (Nishino et al., 2011) (data not shown). Therefore, it is possible that current reprogramming protocols fail to set the appropriate epigenetic state at the *DLK1-DIO3* locus, which is an important consideration given that improper regulation of this imprinted regions leads to developmental abnormalities in mice (Stadtfeld et al., 2010; Takahashi et al., 2009) and humans (Kagami et al., 2008). The mechanistic details of the epigenetic alteration at this imprinted locus during reprogramming remains elusive, but our data suggest that MEG3 interactions with PRC2 might play an important role.

Although our findings suggest a molecular mechanism by which MEG3 contributes to JARID2 and PRC2 function, further investigations are required to address (1) whether and how this mechanism extends to the other lncRNAs identified by PAR-CLIP, and (2) whether and how lncRNAs regulate de novo recruitment of their interacting proteins to distant loci. Our in vitro binding analyses of Meg3 lncRNA suggest that JARID2 does not bind to RNA in a sequence-specific manner, consistent with the fact that the primary sequence of lncRNAs is not well conserved (Ulitsky et al., 2011).

In conclusion, we have demonstrated that JARID2, an essential regulatory component of PRC2 in pluripotent stem cells, contains an RNA-binding region that mediates, at least in part, its interaction with the imprinted lncRNA MEG3. This—and possibly other—RNA interaction contributes to proper recruitment and assembly of PRC2 on target genes and likely plays an important role in orchestrating the epigenetic regulation of gene expression that accompanies the transition from stem cell pluripotency to differentiation.

EXPERIMENTAL PROCEDURES

For information about antibodies, oligonucleotides, and plasmids see Tables S4–S6 and the Supplemental Experimental Procedures.

Cells

HiPSCs were cultured as described previously (Nishino et al., 2011). Cells were harvested at passage 34 (UtE-iPS-4), 41 (UtE-iPS-11), 45 (UtE-iPS-7), 41 (AM-iPS-6), 50 (AM-iPS-8), 39 (UtE-iPS-6), 41 (MRC5-iPS-25), and 90 (Edom-iPS-2). Human foreskin fibroblasts (#PCS-201-010, lot# 58490326; ATCC) were cultured in fibroblast basal medium (ATCC) plus fibroblast growth kit-low serum (ATCC). KH2 ESCs expressing the reverse tetracycline-controlled transactivator (rtTA) (Hochedlinger et al., 2005) were maintained in standard mouse ESC (mESC) culture conditions. KH2 lines expressing *Jarid2* and *Jarid2*$_{\Delta RBR}$ were generated by transfection of the relevant pINTA-N3 construct and selecting with 50 μg/ml Zeocin (Invitrogen). Transgene expression was induced with doxycycline for 24 hr. E14Tg2A.4 mESC lines (E14 mESC) and HEK293 and HEK293T cells were cultured as described previously (Kaneko et al., 2010). *Jarid2* knockdown mESCs were described previously (Li et al., 2010).

In Vitro Binding Assays

Protein purification, synthesis of HOTAIR lncRNA (1–333), and biotinylated RNA pull-down assays were described previously (Kaneko et al., 2010). All

RNA fragments were generated by in vitro transcription; see the Supplemental Information for details.

LncRNA-mediated stimulation of EZH2–JARID2 interactions was assayed by incubating FLAG- and 6xHis-tagged EZH2 (20 pmol) with increasing amounts (0–12 pmol) of lncRNAs in 100 μl of binding buffer (50 mM Tris-HCl, pH 7.9; 100 mM KCl; 0.1% NP-40) for 30 min at 25°C. 6xHis-tagged truncated JARID2 (40 pmol) was added to the reaction and incubated for 30 min at 25°C. Complexes were purified using FLAG-M2 affinity gel (Sigma), after washing with binding buffer.

For JARID2–Meg3 in vitro binding assays, 4 μg of GST-JARID2$_{119–450}$ were incubated with a series of Meg3 fragments (1 μg, 400 nt each; see Table S7). Bound RNAs were purified with glutathione beads, resolved with 7M urea gels, stained with SYBR-gold (Invitrogen), and quantified with an ImageQuant LAS4000 (GE Healthcare Life Sciences).

Knockdowns

For conditional knockdown of *Ezh2* in E14 mESCs, we generated stable clones with an integration of pTRIPZ lentiviral inducible shRNAmir targeting human and mouse *Ezh2* (#RHS4696-99635303; Open Biosystems). Selection was done by puromycin (1 μg/ml) and clones were screened by red fluorescent protein (RFP) expression after 3–4 days of doxycycline (1 μg/ml) induction.

For transient knockdown of *Meg3*, we tested four siRNAs from QIAGEN (SI05169486, SI05169710, SI01060129, and SI01060136) and used the siRNA resulting in the most efficient knockdown (SI05169486) (see Supplemental Information).

ChIP

ChIP from hiPSCs, foreskin fibroblast, and mESCs was performed as described (Kaneko et al., 2007), with minor modification, and libraries were constructed as described (Gao et al., 2012). Briefly, cells were crosslinked with 1% formaldehyde for 10 min and sonicated in ChIP buffer (50 mM Tris-HCl [pH 7.9], 150 mM NaCl, 1% Triton X-100, 0.5% NP-40, 5 mM EDTA [pH 8.0], 1 mM phenylmethanesulfonylfluoride, and protease inhibitors) with a Diagenode Bioruptor. Incubations with antibodies were carried out in an ultrasonic water bath for 30 min at 4°C. Samples were decrosslinked at 65°C for ~16 hr for library construction or 95°C 10 min for ChIP-qPCR. See the Supplemental Information for more details.

RNA Immunoprecipitation

For Figures 3G and 3H, nuclear extracts were obtained using an established protocol (Dignam et al., 1983) with minor modifications to minimize RNase activity, lysates were diluted in RIP buffer (20 mM Tris [pH 7.9]$_{4°C}$, 200 mM KCl, 0.05% IGEPAL CA-630, 10 mM EDTA), cleared by centrifugation at 20,000 × *g* for 10 min, and incubated with depleting amounts of antibody for 3 hr at 4°C. Immunocomplexes were recovered with protein G-coupled dynabeads (Invitrogen) for 1 hr at 4°C. Beads were washed in RIP-W buffer (20 mM Tris [pH 7.9]$_{4°C}$, 200 mM KCl, 0.05% IGEPAL CA-630, 1 mM $MgCl_2$) twice and incubated with 2 U TURBO DNase (Ambion) in 20 μl RIP-W buffer for 10 min at room temperature to avoid DNA bridging artifacts. After two additional washes, RNA was eluted and purified with TRIzol (Invitrogen).

For Figures 2C–2F, RIPs were performed on whole-cell lysates, as described before (Kaneko et al., 2010) (see also Supplemental Information).

ChIP-Seq Analysis

Sequenced reads from ChIP-seq experiments were mapped with BOWTIE using parameters -v2 -m4–best (Langmead et al., 2009). Normalized genome-wide read densities were computed and visualized on the UCSC genome browser. Bound regions (ERs) were identified using MACS 2.09 (Zhang et al., 2008) and default parameters. ERs were associated to gene targets using ChIPpeakAnno and ENSEMBL annotation 67.

PAR-CLIP

ESCs were pulsed with 100 μM 4-SU (Sigma) for 16–24 hr and crosslinked with 400 mJ/cm^2 UVA (365 nm) using a Stratalinker UV crosslinker (Stratagene). Cells were lysed for 10 min at 37°C in CLIP buffer (20 mM HEPES [pH 7.4], 5 mM EDTA, 150 mM NaCl, 2% lauryldimethylbetaine) with protease inhibitors, 20 U/ml Turbo DNase (Life Technologies), and 200 U/ml murine RNase inhibitor (New England Biolabs). IPs were carried out in CLIP buffer for 1 hr at 4°C. When necessary, extracts were treated with RNase A + T1 cocktail (Ambion) for 5′ at 37°C. Immunocomplexes were recovered with protein G-coupled dynabeads for 45 min at 4°C. DNA was removed with Turbo DNase (2 U in 20 μl). Crosslinked RNA was labeled by incubations with 5U Antarctic phosphatase and 5U T4 PNK (both from New England Biolabs) in presence of 10 μCi [γ-^{32}P] ATP (PerkinElmer, MA). Labeled material was resolved on 8% bis-tris gels, transferred to nitrocellulose, and exposed to autoradiography films for 1–24 hr.

For PAR-CLIP-seq, 100 pmol of a 3′-blocked DNA adaptor was ligated to the RNA after dephosphorylation and before 5′ labeling by incubating the beads with T4 RNA ligase 1 (New England Biolabs) for 1 hr at 25°C. After autoradiography, bands were excised and the RNA eluted with proteinase K for 30 minutes at 37°C and proteinase K in 3.5M urea for 30 minutes at 55°C. Custom 5′ adapters were ligated, and the products were size-selected on polyacrylamide or agarose gels, amplified, and sequenced on an Illumina HiSeq 2000.

For the analysis, adapter sequences were removed and reads < 17 nt discarded. The remaining reads were mapped to the mm9 genome using BOWTIE (Langmead et al., 2009), allowing two mismatches and removing duplicates. RCSs were identified with PARalyzer (Corcoran et al., 2011) requiring at least two T→C conversions per RCS. For Figure 2F, RCSs were assigned to lncRNAs in ENSEMBL 67 when they overlapped anywhere within the gene body to account for imprecisions in the annotation of lncRNAs.

RNA Structural Predictions

Structural predictions and minimum free-energy calculations shown in Figure S3 were performed with the Vienna RNA Websuite using default settings (Gruber et al., 2008).

ACCESSION NUMBERS

The ChIP and CLIP sequences reported in this paper have been deposited to the Gene Expression Omnibus (GEO) with the accession number GSE48518. EZH2 PAR-CLIP-seq data used for comparative analyses were taken from GEO accession number GSE49433 (Kaneko et al., 2013).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2013.11.012.

ACKNOWLEDGMENTS

We thank the Genome Technology Center at NYU for help with sequencing, John Rinn and Matthias Stadtfeld for comments on the manuscript, and Varun Narendra for bioinformatic analyses. This work was supported by grants from the National Institutes of Health (GM-64844 and R37-37120) and the Howard Hughes Medical Institute (to D.R.). R.B. was supported by a Helen Hay Whitney Foundation postdoctoral fellowship and by the Helen L. and Martin S. Kimmel Center for Stem Cell Biology postdoctoral fellow award. R.S.M. was supported by a Ph.D. and an international research stay fellowship from CONACyT (213029).

Received: July 10, 2013
Revised: October 2, 2013
Accepted: November 21, 2013
Published: December 26, 2013

REFERENCES

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2004). GenBank: update. Nucleic Acids Res. *32* (Database issue), D23–D26.

Bonasio, R., Tu, S., and Reinberg, D. (2010). Molecular signals of epigenetic states. Science *330*, 612–616.

Boulay, G., Dubuissez, M., Van Rechem, C., Forget, A., Helin, K., Ayrault, O., and Leprince, D. (2012). Hypermethylated in cancer 1 (HIC1) recruits polycomb repressive complex 2 (PRC2) to a subset of its target genes through interaction with human polycomb-like (hPCL) proteins. J. Biol. Chem. *287*, 10509–10524.

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell *71*, 515–526.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafrenière, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*, 527–542.

Corcoran, D.L., Georgiev, S., Mukherjee, N., Gottwein, E., Skalsky, R.L., Keene, J.D., and Ohler, U. (2011). PARalyzer: definition of RNA binding sites from PAR-CLIP short-read sequence data. Genome Biol. *12*, R79.

da Rocha, S.T., Edwards, C.A., Ito, M., Ogata, T., and Ferguson-Smith, A.C. (2008). Genomic imprinting at the mammalian Dlk1-Dio3 domain. Trends Genet. *24*, 306–316.

Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. Nat. Struct. Mol. Biol. *20*, 1250–1257.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. *11*, 1475–1489.

Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol. Cell *45*, 344–356.

Ginis, I., Luo, Y., Miura, T., Thies, S., Brandenberger, R., Gerecht-Nir, S., Amit, M., Hoke, A., Carpenter, M.K., Itskovitz-Eldor, J., and Rao, M.S. (2004). Differences between human and mouse embryonic stem cells. Dev. Biol. *269*, 360–380.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008). The Vienna RNA websuite. Nucleic Acids Res. *36* (Web Server issue), W70–W74.

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.

Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jr., Jungkamp, A.C., Munschauer, M., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell *141*, 129–141.

Hochedlinger, K., Yamada, Y., Beard, C., and Jaenisch, R. (2005). Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. Cell *121*, 465–477.

John, R.M., and Surani, M.A. (1996). Imprinted genes and regulation of gene expression by epigenetic inheritance. Curr. Opin. Cell Biol. *8*, 348–353.

Kagami, M., Sekita, Y., Nishimura, G., Irie, M., Kato, F., Okada, M., Yamamori, S., Kishimoto, H., Nakayama, M., Tanaka, Y., et al. (2008). Deletions and epimutations affecting the human 14q32.2 imprinted region in individuals with paternal and maternal upd(14)-like phenotypes. Nat. Genet. *40*, 237–242.

Kaneko, S., Rozenblatt-Rosen, O., Meyerson, M., and Manley, J.L. (2007). The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3′ processing and transcription termination. Genes Dev. *21*, 1779–1789.

Kaneko, S., Li, G., Son, J., Xu, C.F., Margueron, R., Neubert, T.A., and Reinberg, D. (2010). Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA. Genes Dev. *24*, 2615–2620.

Kaneko, S., Son, J., Shen, S.S., Reinberg, D., and Bonasio, R. (2013). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1258–1264.

Kanhere, A., Viiri, K., Araújo, C.C., Rasaiyaah, J., Bouwman, R.D., Whyte, W.A., Pereira, C.F., Brookes, E., Walker, K., Bell, G.W., et al. (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. Mol. Cell *38*, 675–688.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., 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.

Kim, T.G., Kraus, J.C., Chen, J., and Lee, Y. (2003). JUMONJI, a critical factor for cardiac development, functions as a transcriptional repressor. J. Biol. Chem. *278*, 42247–42255.

Kim, H., Kang, K., and Kim, J. (2009). AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. Nucleic Acids Res. *37*, 2940–2950.

Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jørgensen, H.F., Pereira, C.F., Leleu, M., Piccolo, F.M., Spivakov, M., et al. (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. Nat. Cell Biol. *12*, 618–624.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lanzuolo, C., and Orlando, V. (2012). Memories from the polycomb group proteins. Annu. Rev. Genet. *46*, 561–589.

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. Genes Dev. *24*, 368–380.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762–767.

Nishino, K., Toyoda, M., Yamazaki-Inoue, M., Fukawatase, Y., Chikazawa, E., Sakaguchi, H., Akutsu, H., and Umezawa, A. (2011). DNA methylation dynamics in human induced pluripotent stem cells over time. PLoS Genet. *7*, e1002085.

Pasini, D., Cloos, P.A., Walfridsson, J., Olsson, L., Bukowski, J.P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature *464*, 306–310.

Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell *139*, 1290–1302.

Puda, A., Milosevic, J.D., Berg, T., Klampfl, T., Harutyunyan, A.S., Gisslinger, B., Rumi, E., Pietra, D., Malcovati, L., Elena, C., et al. (2012). Frequent deletions of JARID2 in leukemic transformation of chronic myeloid malignancies. Am. J. Hematol. *87*, 245–250.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

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

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat. Rev. Genet. *8*, 9–22.

Shen, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.-C., Lee, Y., and Orkin, S.H. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell *139*, 1303–1314.

Son, J., Shen, S.S., Margueron, R., and Reinberg, D. (2013). Nucleosome binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. Genes Dev., in press. Published online December 15, 2013. gad.225888.113.

Stadtfeld, M., Apostolou, E., Akutsu, H., Fukuda, A., Follett, P., Natesan, S., Kono, T., Shioda, T., and Hochedlinger, K. (2010). Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature *465*, 175–181.

Takahashi, N., Okamoto, A., Kobayashi, R., Shirai, M., Obata, Y., Ogawa, H., Sotomaru, Y., and Kono, T. (2009). Deletion of Gtl2, imprinted non-coding RNA, with its differentially methylated region induces lethal parent-origin-dependent defects in mice. Hum. Mol. Genet. *18*, 1879–1888.

Takeuchi, T., Yamazaki, Y., Katoh-Fukui, Y., Tsuchiya, R., Kondo, S., Motoyama, J., and Higashinakagawa, T. (1995). Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. Genes Dev. *9*, 1211–1222.

Trojer, P., and Reinberg, D. (2007). Facultative heterochromatin: is there a distinctive molecular signature? Mol. Cell *28*, 1–13.

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

Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell *147*, 1537–1550.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120–124.

Yap, K.L., Li, S., Muñoz-Cabello, A.M., Raguz, S., Zeng, L., Mujtaba, S., Gil, J., Walsh, M.J., and Zhou, M.M. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol. Cell *38*, 662–674.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhao, J., Ohsumi, T.K., Kung, J.T., Ogawa, Y., Grau, D.J., Sarma, K., Song, J.J., Kingston, R.E., Borowsky, M., and Lee, J.T. (2010). Genome-wide identification of polycomb-associated RNAs by RIP-seq. Mol. Cell *40*, 939–953.